Optimization on Spaces of Curves

Møller-Andersen, Jakob; Gravesen, Jens

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Optimization on Spaces of Curves

Jakob Møller-Andersen

Kongens Lyngby 2016

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Richard Petersens Plads, building 324,
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk

Summary (English)

This thesis is concerned with computational and theoretical aspects of Riemannian metrics on spaces of regular curves, and their applications. It was recently proved that second order constant coefficient Sobolev metrics on curves are geodesically complete. We extend this result to the case of Sobolev metrics with coefficient functions depending on the length of the curve. We show how to apply this result to analyse a wide range of metrics on the submanifold of unit and constant speed curves.

We present a numerical discretization of second order Sobolev metrics on the space of regular curves in $\mathbb{R}^d$, and methods to solve the initial and boundary value problem for geodesics allowing us to compute the Karcher mean and principal components analysis of data of curves. We apply the methods to study shape variation in synthetic data in the Kimia shape database, in HeLa cell nuclei and cycles of cardiac deformations.

Finally we investigate a new application of Riemannian shape analysis in shape optimization. We setup a simple elliptic model problem, and describe how to apply shape calculus to obtain directional derivatives in the manifold of planar curves. We present an implementation based on parametrization of immersions by B-splines, which ties in naturally with Isogeometric Analysis to solve the PDE. We give numerical examples of solutions, and compare the Riemannian optimization algorithms with different choices of metrics to a naive unregularized discretize-first approach.

Summary (Danish)

Denne afhandling omhandler teoretiske og beregningsmæssige aspekter af Riemannske metrikker på rum af regulære kurver og deres anvendelser. Det blev for nyligt vist at anden ordens Sobolev metrikker med konstante koefficienter er geodætisk fuldstændige. Vi udvider dette resultat til Sobolev metrikker med koefficienter der kan afhænge af længden af kurven. Vi viser desuden hvordan dette kan bruges til at analysere en række metrikker på delmangfoldigheden af kurver med enheds- og konstant fart.

Vi præsenter en numerisk diskretisation af anden ordens Sobolev metrikker på rummet af regulære kurver i $\mathbb{R}^d$, og metoder til at løse de geodæsiske begyndelses- og randværdi problemer. Dette giver mulighed for at beregne Karcher gennemsnit og principal komponent analyse (PCA) af kurvedata. Vi anvender methoderne til at studere formvariation i den syntetiske Kimia database, in Hela cellekerner og cykluser af hjertedeformationer.

Til sidst udforsker vi en ny anvendelse af Riemannsk formanalyse i formoptimering. Vi definerer et simpelt elliptisk model problem, og beskriver hvordan vi kan bruge form kalkyle til at beregne retningsafledte i mangfoldigheden af plane kurver. Vi giver en implementering baseret på Isogeometrisk Analyse til at løse den underliggende partielle differentialligning. Vi viser numeriske eksempler på løsninger, og sammenlinger de Riemannske optimeringsalgoritmer ved forskellige valg af metrikker med en naiv ikke-regulariseret *diskretiser-først* strategi.

Preface

This thesis is submitted in partial fulfilment of the requirements for obtaining the Ph.D. degree. The work has been carried out at the Department of Applied Mathematics and Compute Science (DTU Compute) at the Technical University of Denmark from September 2013 to August 2016 under the supervision of Jens Gravesen.

Prerequisites for reading this thesis are acquittance with Riemannian geometry, standard elliptic PDE theory on Sobolev spaces, B-splines and Galerkin methods.

Acknowledgements

First I want to thank my supervisor Jens Gravesen for the support and advice throughout the project. I am grateful for his encouragement to pursue my own ideas and interests, and for our many long discussions during my project.

I sincerely thank my coauthors Martin Bauer, Martins Bruveris and Philipp Harms for their collaboration on papers, discussions, guidance and for the hospitality during my research visits in London and Vienna.

Parts of this work was supported by the Erwin Schrödinger Institute programme: Infinite-Dimensional Riemannian Geometry with Applications to Image Matching and Shape Analysis. I would like to thank the organizers and participants of this two months programme for the inspiration and discussions, from which I

gained invaluable knowledge and inspiration. I especially want to thank Stephen Preston for very helpful discussions that guided me in my project.

Finally I want to thank my friends and family for supporting me throughout the three years. For listening to my ramblings about mathematics, and for providing so much fun in the time in between research.

Papers submitted

Part of the work in this thesis has so far resulted in the following papers in two peer-reviewed conference procedings and one journal.

- M. Bauer, M. Bruveris, P. Harms, and J. Møller-Andersen. Curve matching with applications in medical imaging. In *5th MICCAI workshop on Mathematical Foundations of Computational Anatomy*, 2015.
- M. Bauer, M. Bruveris, P. Harms, and J. Møller-Andersen. Second order elastic metrics on the shape space of curves. In *1st Workshp on Differential Geometry in Computer Vision for Analysis of Shapes, Images and Trajectories*, 2015.
- M. Bauer, M. Bruveris, P. Harms, and J. Møller-Andersen. A Numerical Framework for Sobolev Metrics on the Space of Curves. *SIAM Journal of Imaging*, to appear, 2016.

Kongens Lyngby, September 7th, 2016

Jakob Møller-Andersen

Contents

Chapter 1

Introduction

A curve or surface is defined in classical differential geometry as a subset of $\mathbb{R}^n$ which can locally be parametrized by $\mathbb{R}^m$ with $m = 1$ or $m = 2$. In this thesis we restrict our selves to the case of closed curves, which can be represented by a mapping $c : S^1 \to \mathbb{R}^d$, but all the questions are constructions can naturally be done for surfaces and submanifolds of codimension greater than one. In many situations one is only interested in the geometric object itself, and not the way it is represented by a parametrization, as there are many parametrizations of the same curve. Any two parametrizations c, d that are related by $c = d \circ \varphi$ of a diffeomorphism φ of S^1, parametrizes the same curve. This notion of equivalence between parametrizations give rise to the notion of a *shape space*, where each element represent an *unparametrized* curve.

From a mathematical point of view, we can view this space as a geometrical object itself; the shape space and the closely related space of all regular parametrizations carry the structure of an infinite dimensional manifold. These spaces can also be equipped with Riemannian metrics, and is a great source of examples for infinite dimensional Riemannian Geometry. This is the theme of Chapter 2 of this dissertation, where we give an introduction to the infinite dimensional geometry of spaces of curves and shape spaces. The central result is an extension of the recently proved theorem, that constant coefficient Sobolev metrics of order two and higher are geodesically complete, to the case where the coefficients are allowed to depends on the length of the curve.

The Riemannian metric allow us to define a distance between shapes by measuring the length of deformations between them. This can be interpreted as a measure of similarity between shapes. Moreover the exponential map locally linearize the space, and in this way allow us to do statistics that respects the non-linear structure of the space. For metrics on closed curves several applications have been considered in medical imaging [84, 86], object tracking [80, 81], computer animation [13, 31], speech recognition [79], biology [45, 78], and many other fields [11, 44]. To apply this method in practice to real data of shapes, it is crucial to develop efficient numerical methods to compute geodesic deformations, for both the initial and boundary value problem. So far numerical methods have largely been restricted to a narrow class of metrics that allow transformations that simplify problem. In Chapter 3 we present a general numerical method that approximates geodesics for higher order metrics with any choice of constants in metric, thereby giving more flexibility in the shape analysis. The contents of Chapter 3 is a copy of soon to be published paper.

Shape optimization has many applications. It can roughly be defined as any optimization problem where the feasible space is a set of shapes. As such it is not defined in a linear vector space, and many different ad-hoc methods have been used to overcome this problem. Typically the problems arise from physical considerations, and the problem is constrained by the solution of a Partial Differential Equation (PDE) which dictates the physics. A classical example is the optimization of the profile of a wing which minimizes drag of the air moving past it. Here the PDE in question is the Navier-Stokes equation which describes the flow of air around the wing. There are many examples coming from structural mechanics, electromagnetics, acoustics, etc. Usually the feasible shapes are represented by parametrizations of the boundary of the domain at hand, which are required to be regular, and the problem doesn't depend on the specific choice of parametrization of a shape. In this way, it is natural to consider shape functionals defined on planar objects, as a function on the manifold of regular curves or shape space. This point of view have been considered very recently in a series of papers, starting with [69], but still leaving many open questions. In this way the non-linear structure of the feasible shapes are captured, and we can use the manifold structure to influence the optimization procedure. The choice of a Riemannian metric on immersions allow us to use generalized versions of classical optimization algorithms like steepest-descent and BFGS, in a way that naturally turns the regularity constraint into an unconstrained problem. Chapter 4 is devoted to a presentation of the tools necessary to implement a version of Riemannian shape optimization, which is then applied to a concrete example where the solution is known.

Chapter 2

Riemannian Geometry of Spaces of Curves

This chapter is devoted to an introduction to infinite dimensional Riemannian manifolds, and spaces of regular curves. The goal will be to define and analyse so-called *shape spaces* of curves. In this context we will consider a planar shape as the outline of a domain, represented by a curve. A shape space contains a unique representative of the image of a curve, without considering specific parametrisations. The presentation will be as follows, first we give a general definition and comment on the differences to the finite dimensional case, and then we will give an example in the setting of immersions. The topics are: the definition of an infinite dimensional manifold, Riemannian metrics, length of curves and Riemannian distances, the geodesic equation and finally metric and geodesic completeness.

2.1 Space of immersions and shape space

First let us present the definition of a (possibly) infinite dimensional manifold, it carries over almost verbatim from the finite dimensional case:

Definition 2.1 A *smooth manifold* modelled on the vector space E is a

Hausdorff topological space M together with a family of *charts* $(u_\alpha, U_\alpha)_{\alpha\in A}$ such that

1. $(U_\alpha)_{\alpha\in A}$ is a covering of M by open sets.
2. $u_\alpha : U_\alpha \subseteq M \to E$ is a homeomorphism onto its image.
3. The transitions mappings $u_\beta \circ u_\alpha^{-1}$: $u_\alpha(U_\alpha \cap U_\beta) \to u_\beta(U_\alpha \cap U_\beta)$ are C^∞-smooth.

The only change is the specification of the modelling space E, which can now be infinite dimensional. If E is finite dimensionsal there is only one choice: $E = \mathbb{R}^n$, for some $n \in \mathbb{N}$. If $\dim E = \infty$ then we have a range of choices of different types of vector spaces. The most familiar cases are Hilbert and Banach spaces. They comes with an inner product or norm, respectively. More generally we can choose E to be a Fréchet space, or a convenient vector space. The latter is a very general type of vector space that we shall not need in this dissertation. Frechét spaces might not be known to the reader, and we give a definition

Definition 2.2 A Fréchet space X is a Hausdorff topological vector space, which satisfy

- The topology may be induced by a countable family of semi-norms $\|\cdot\|_k$: a set U is open iff forall $u \in U$ there exists $K \geq 0$ and $\epsilon > 0$ such that $\{v : \|v - u\|_k < \epsilon, k \leq K\} \subseteq U$.
- It is complete w.r.t the family of semi-norms: Every sequence (x_n) which is Cauchy w.r.t all $\|\cdot\|_k$ converges to a fixed element x in each seminorm.

Alternatively it is a vector space with a topology that can be induced by a translation invariant metric d. Our most important example of a Frechet space is the space $C^\infty([a,b])$ of smooth functions on a compact interval, with the family of semi-norms given by $\|f\|_n = \sup_x\{|f^{(k)}(x)| : x \in [a,b],\, k \leq n\}$. In Hilbert or Banach spaces we have many of the same technical tools as for analysis in finite dimensions, and many results about finite dimensional manifolds can be carried over, but not all. For Fréchet spaces we have less tools available and these spaces become much more difficult to work with. We shall later exemplify these problems, when we also equip our manifolds with Riemannian metrics. Here the differences begin to appear immediately. For an introduction to Hilbert and Banach manifolds, see [46, 41]. In [36] an introduction to Fréchet manifolds is given, and the famous Nash-Moser theorem. For the general case of convenient vector spaces see [50].

The space of immersions

Now we will define the spaces that we shall be using throughout this dissertation, and show how these are naturally smooth infinite dimensional manifolds. The space of smooth, regular, closed curves with values in $\mathbb{R}^d$ is

$$\operatorname{Imm}(S^1,\mathbb{R}^d) = \left\{c \in C^\infty(S^1,\mathbb{R}^d) \colon \forall \theta \in S^1, c'(\theta) \neq 0\right\}, \tag{2.1}$$

where Imm stands for *immersions*. Throughout the text, differentiation is denoted using subscripts as in $c_\theta = \partial_\theta c = c'$. We call such curves parametrized to distinguish them from unparametrized curves defined later in Sect. 3.2.3. The space $\operatorname{Imm}(S^1,\mathbb{R}^d)$ is an open subset of the Fréchet space $C^\infty(S^1,\mathbb{R}^d)$ and therefore it can be considered as a Fréchet manifold. Here $C^\infty(S^1,\mathbb{R}^d)$ has the topology of uniform convergence, i.e. we have $f_n \to f$ if

$$\lim_{n\to\infty} \|f_n^{(k)} - f^{(k)}\|_\infty = 0,\ \forall k \in \mathbb{N},$$

where $\|f\|_\infty = \sup_{\theta\in S^1} |f(\theta)|$. Its tangent space $T_c \operatorname{Imm}(S^1,\mathbb{R}^d)$ at any curve c is the vector space of all vector fields along c, i.e. $C^\infty(S^1,\mathbb{R}^d)$ itself.

In many cases, one is not interested in the specific parametrization of an object, but rather only its properties as a 'shape'. That is, two different parametrizations, in our case curves, can represent the same object. This is visualized in figure 2.1. The idea now is to construct a suitable quotient space of immersions, where each equivalence class is exactly a unique shape. To this end we define the space of smooth diffeomorphisms of S^1,

$$\operatorname{Diff}(S^1) = \{\varphi \in C^\infty(S^1,S^1) : \varphi' > 0\}. \tag{2.2}$$

This is an open set in $C^\infty(S^1,S^1)$ so it is an infinite dimensional manifold, and also a smooth Fréchet Lie group, with the group action given by composition, see [50]. It acts smoothly on $\operatorname{Imm}(S^1,\mathbb{R}^d)$ by compositions on the right. We shall say that two curves c_1 and c_0 represent the same *shape* if they differ by a diffeomorphism of S^1, i.e. there exists $\varphi \in \operatorname{Diff}(S^1)$ such that $c_1 = c_0 \circ \varphi$, i.e. c_1 is a *reparametrization* of c_0. The quotient space

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

consists of the orbits of the action of the diffeomorhpism group, and will be denoted by the space of *unparametrized* curves, since we can think of the elements as curves without a specific parametrization. The resulting quotient space is not a manifold: the action of $\operatorname{Diff}(S^1)$ is not free, and the resulting quotient space is an orbifold with singularities, see [27]. However one can the bypass this difficulty by considering the slightly smaller space of free immersions, those for which $\operatorname{Diff}(S^1)$ acts freely, i.e. the isotropy group is trivial. Let $\operatorname{Imm}_f(S^1,\mathbb{R}^d)$

denote the set of *free* immersions, This is a dense subspace of $\mathrm{Imm}(S^1, \mathbb{R}^d)$. An example of a non-free immersion is the circle traversed two times. The space

$$B_{i,f}(S^1, \mathbb{R}^d) = \mathrm{Imm}_f(S^1, \mathbb{R}^d)/\operatorname{Diff}(S^1), \tag{2.4}$$

is now the space we shall work with, we can summarize its structure as follows.

Theorem 2.3 *The space $B_{i,f}(S^1, \mathbb{R}^d)$ is a Fréchet manifold, and the base space of the principal fibre bundle*

$$\pi : \mathrm{Imm}_f(S^1, \mathbb{R}^d) \to B_{i,f}(S^1, \mathbb{R}^d), \quad c \to c \circ \operatorname{Diff}(S^1).$$

with structure group $\operatorname{Diff}(S^1)$.

This follows from a careful contruction of appropriate charts, for a complete proof of the theorem, see [27, 51]. Now we will just describe the charts. Define the map

$$\psi_c : C^\infty(S^1, (-\epsilon, \epsilon)) \to \mathrm{Imm}_f(S^1, \mathbb{R}^d), \quad \psi_c(a) = c + an_c,$$

With n_c the normal vector of c. This is simply moving $c(\theta)$ point-wise a length $a(\theta)$ in the direction along its unit normal tangent vector field. Then

$$\pi \circ \psi_c : C^\infty(S^1, (-\epsilon, \epsilon)) \to B_{i,f},$$

with ϵ sufficiently small, gives a chart for $B_{i,f}$ around $\pi(c)$. We can *identify* the tangent space $T_{[c]}B_{i,f}$ with the set of all normal vector fields. Note that this does not mean that the set of normal vector fields *is* the tangent space, only that $d\pi_c(an_c)$ is surjective.

In practice for our results in Chapter 2 and 3, distinguishing between B_i and $B_{i,f}$ will not be important, so we will for the remaining part assume that all immersions are free. For more details on non-free immersions see [51, 27].

If we restrict to embeddings, the same manifolds charts works and we can make the exact same construction and obtain a principal bundle.

$$\pi : \mathrm{Emb}(S^1, \mathbb{R}^d) \to B_e(S^1, \mathbb{R}^d),$$

where $\mathrm{Emb}(S^1, \mathbb{R}^d)$ is the space of embeddings, and B_e is its quotient by $\operatorname{Diff}(S^1)$.

In the literature there is a bit of ambiguity when talking about a *shape space*. In [53], the quotient of $B_{i,f}$ by the euclidean group $SE(d) = SO(d) \ltimes \mathbb{R}^d$, $\mathcal{S} = B_{i,f}/SE(d)$ is called shape space, and $B_{i,f}$ is called a *pre-shape space*.

Finally we mention that all these spaces can be generalized to immersions between general manifolds M and N, $\mathrm{Imm}(M, N)$, see the book [50].

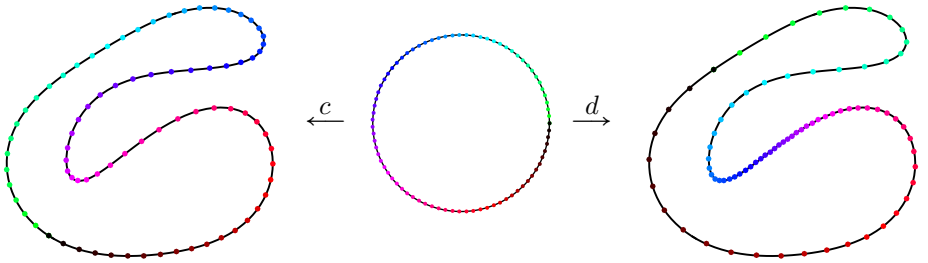


Figure 2.1: Two different parametrizations of the same shape: $c = d \circ \phi$, $\phi \in \mathrm{Diff}(S^1)$

2.1.1 Sobolev immersions

For technical reasons we will later want to consider curves of only Sobolev regularity. We define the immersions of Sobolev order $n \geq 2$,

$$\mathrm{Imm}^n(S^1, \mathbb{R}^d) = \{c \in H^n(S^1, \mathbb{R}^d) : |c'| > 0\}.$$

This is well-defined by the Sobolev Embedding Thm., see Lemma A.1, which gives the continuous inclusion $H^n(S^1) \hookrightarrow C^1(S^1, S^1)$. Likewise we can define Sobolev diffeomorphisms for $n \geq 1$:

$$\mathrm{Diff}^n(S^1) = \{\varphi \in \mathrm{Id} + H^n(S^1, S^1) : \varphi \text{ bijective}, \varphi^{-1} \in \mathrm{Id} + H^n(S^1, S^1)\}$$

Both of these spaces are smooth Hilbert manifolds. In the smooth category we know that the group of diffeomorphisms acts smoothly on curves, this is no longer the case for Sobolev curves:

Lemma 2.4 *Let $n \geq 2$ and $k \geq 1$. Then the composition map*

$$H^{n+k}(S^1, \mathbb{R}^d) \times \mathrm{Diff}^n(S^1) \to H^n(S^1, \mathbb{R}^d) : (f, \varphi) \mapsto f \circ \varphi$$

is a C^k map.

A proof can be found in [39]. We emphasize that the action is strictly C^k and not smoother. This also implies that $\mathrm{Diff}^n(S^1)$ is not a smooth Lie group, but only a topological group: the group action is only continuous. For a Sobolev curve, the normal vector field is in $H^{n-1}(S^1, \mathbb{R}^d)$. This has the consequence that construction of charts for the quotient space

$$\mathcal{B}^n(S^1) = \mathrm{Imm}^n(S^1, \mathbb{R}^d) / \mathrm{Diff}^n(S^1)$$

breaks down, and we cannot prove that $\mathcal{B}^n(S^1)$ is a manifold. One can show that this is a Haussdorf shape, see [23]. One could speculate that instead of using the normal vector field in the construction of a chart, a point-wise transversal H^n vector field could be used instead. The quotient space might then be a Hilbert manifold, but not a smooth principal bundle.

2.2 Riemannian metrics and distance

In the previous section we defined spaces of immersions and how to construct shape spaces as a quotient by the diffeomorphism group. Now we want to equip these with Riemannian metrics, and study the resulting geometry. As before, the generalization of a Riemannian metric to infinite dimension can be stated very simply, expect one has to take special care when specifying the topology. Let us start with the definition

Definition 2.5 Let M be a manifold modelled on E. A *weak* Riemannian metric G is a smooth map

$$G : TM \times_M TM \to \mathbb{R},$$

satisfying

(1) $G_x(\cdot,\cdot)$ is bilinear.

(2) $G_x(h,h) \geq 0$ for all $h \in T_xM$, with equality only for $h = 0$.

A metric G is called a *strong* Riemannian metric, if it in addition satisfies

(3) The topology of the inner product space $(T_xM, G_x(\cdot,\cdot))$ coincides with the inherited manifold topology on T_xM.

If a manifold posses a strong Riemannian metric, it implies that T_xM is a Hilbert space and hence also the modelling space. Hence if the modelling space is a Frechet space, but not a Banach space, we can only have weak metrics. The two books [41, 46] show how much of the theory of finite dimensional manifolds carry over to strong Riemannian manifolds. For general weak Riemannian manifolds, most is handled on a case by case scenario. There exist many pathological examples that show that it is difficult to come up with a general theory about them. We shall encounter some of these problems later. On the other

hand, there are many more examples of weak Riemannian metrics, that have applications in physics, image analysis and more.

A curve in the space of curves is a map $c \in C^\infty(I, C^\infty(S^1, \mathbb{R}^d)) \simeq C^\infty(I \times S^1, \mathbb{R}^d)$ (the equivalence is the exponential for convenient vector spaces), we will refer to this as a *path* in $\operatorname{Imm}(S^1, \mathbb{R}^d)$. Its velocity is denoted by $c_t = \partial_t c = \dot{c}$. Using a Riemannian metric we can define the length of a path connecting two points as

$$L(c) = \int_0^1 \sqrt{G_c(\dot{c}, \dot{c})} dt,$$

where $\cdot = d/dt$ denotes the time derivative of the path c. The geodesic distance between two points can then be defined as usual,

$$\operatorname{dist}_G(c_0, c_1) = \inf\{L(c) \,|\, c(0) = c_0, c(1) = c_1\}.$$

Since $L(c) \geq 0$, we see that it is well-defined and $\operatorname{dist}_G \geq 0$. It is also easy to see that it satisfies the triangle inequality and is symmetric, so it defines a pesudometric, as we shall see in a bit it does not always define a metric on infinite dimensional manifolds. On the quotient space $B_{i,f}$ we can also induce a metric

$$\operatorname{dist}_G^{B_{i,f}}(\pi(c_0), \pi(c_1)) = \inf\{L(c) : c(0) \in [c_0], c(1) \in [c_1]\}$$

In general we will be looking at metrics which makes the projection π into a Riemannian submersion. A necessary condition is that the metric is invariant under the diffeomorphism group, that is for all $\phi \in \operatorname{Diff}(S^1)$,

$$G_{c \circ \phi}(h \circ \phi, k \circ \phi) = G_c(h, k).$$

If the tangent space at c splits into a vertical and horizontal part like

$$T_c \operatorname{Imm}(S^1, \mathbb{R}^d) = \operatorname{Vert}(c) \oplus \operatorname{Hor}(c),$$

where the vertical part is the kernel $d\pi_c$ consisting of tangential vector fields,

$$\operatorname{Vert}(c) = \ker d\pi_c = \{a \cdot c' : a \in C^\infty(S^1)\},$$

and Hor is the orthogonal complement w.r.t the metric

$$\operatorname{Hor}(c) = \{h \in T_c \operatorname{Imm}(S^1, \mathbb{R}^d) : \forall a \in C^\infty(S^1), G_c(h, ac') = 0\}.$$

then a metric on $B_{i,f}$ can be chosen such that the projection π is a Riemannian submersions. This the tool that allows us to analyse the geometry of $B_{i,f}$ in terms of the geometry of the bigger space $\operatorname{Imm}_f(S^1, \mathbb{R}^d)$.

2.2.1 Sobolev metrics on curves

The simplest metric we can put on the space of immersions might be the flat L^2 metric:

$$G_c(h,k) = \int_{S^1} \langle h,k\rangle d\theta,$$

where $\langle\cdot,\cdot\rangle$ is the Euclidean inner product on $\mathbb{R}^d$. This is not invariant under $\mathrm{Diff}(S^1)$, so it does not descent to a meaningful metric on the quotient space, in fact the induced distance is identically zero [51]. Arguably the simplest change we can do to make this invariant, is to change the integration to the arclength measure $ds = |c'|d\theta$. That is, we consider the metric

$$G_c(h,k) = \int_{S^1} \langle h,k\rangle ds. \tag{2.5}$$

This metric is invariant with respect to $\mathrm{Diff}(S^1)$, as we can easily see from a change of variables

$$\begin{aligned} G_{c\circ\phi}(h\circ\phi, k\circ\phi) &= \int_0^{2\pi} \langle h\circ\phi, k\circ\phi\rangle |(c\circ\phi)'|d\theta \\ &= \int_0^{2\pi} \langle h\circ\phi, k\circ\phi\rangle |(c'\circ\phi)|\phi'(\theta)d\theta \\ &= \int_{\phi(0)}^{\phi(2\pi)} \langle h,k\rangle |c'|d\theta \\ &= G_c(h,k). \end{aligned}$$

Now this is no longer a flat metric, as we have a dependence on the footpoint c. The horizontal bundle consists of all normal vectorfield. As noticed above, it is only a weak metric on $\mathrm{Imm}(S^1,\mathbb{R}^d)$ as the topology induced on the tangent space is not equivalent to the Fréchet topology on $C^\infty(S^1,\mathbb{R}^d)$. This metric was first studied in [51], where it was shown that it suffers from a major defect which can only happen for weak metris on infinite dimensional manifolds:

Theorem 2.6 *The geodesic distance induced by* (2.5) *on* $B_{i,f}$ *vanishes everywhere, i.e. for any two shapes* $\pi(c_0)$ *and* $\pi(c_1)$ *we have* $\mathrm{dist}_{L^2}(\pi(c_0),\pi(c_1)) = 0$.

The proof behind this is not that there exist a path between curves which has length zero, but here exists paths of arbitrary small lengths, so when taking the infimum of all paths connecting points in the equivalence classes $[c_0]$ and $[c_1]$. The paths can be explicitly constructed by appropriately sig-sawing between curves. One can extend this result hold for any c_0 and c_1 in $\mathrm{Imm}(S^1,\mathbb{R}^d)$, and not just $[c_0]$ and $[c_1]$. This means that this metric is not able to distinguish between shapes, hence it is not useful if we want to use this in practice later, as we shall explore in Chap. 3. That means we have to look at other, more

complicated, metrics on curves if we want to do any kind of shape analysis of curves. Several different remedies have been proposed. An observation about this behaviour is that large curvature, length or the local stretching of the curves in the short paths is not being taken into account, so the general idea is if we weight the metric by any of these, the length of the problematic paths should increase and give a well-defined distance on the space. The following weighted L^2 type metric is also analyzed in [51]

$$G_c(h,k) = \int_{S^1} (1 + A\kappa^2)\langle h,k\rangle \, ds, \tag{2.6}$$

where κ is the curvature of the curve and $A > 0$. Here it is shown that one can bound the length of path from below by the area swept out by the path, hence it has to separate points in the geodesic distance. This metric suffers from others problems, as we will mention in the next section. A length-weighted L^2 type metric has also been studied in [85].

$$G_c(h,k) = \ell_c \int_{S^1} \langle h,k\rangle \, ds,$$

Which has the point separating property. Another approach to solving the distance problem is to take the derivatives of the vector field in the tangent space into account. Since we want to have a metric which is invariant under the diffeomorpishm group we need to use an invariant differential operator. One choice is the arc-length derivate $D_s = \frac{1}{|c'|}\partial_\theta$ as this satisfy

$$D_s^{c\circ\varphi} h = \frac{1}{|(c\circ\varphi)'|}(h\circ\varphi)' = \frac{1}{|(c'\circ\varphi)\varphi'|}(h'\circ\varphi)\varphi' = \frac{1}{|c'\circ\varphi|}(h'\circ\varphi) = D_s^c h\circ\varphi,$$

where we have used the upper script to emphasize at which curve the operator is evaluated at. Now the invariant metric can written as

$$G_c^1(h,k) = \int_{S^1} \langle h,k\rangle + a_1\langle D_s h, D_s k\rangle \, ds.$$

Here $a_1 > 0$ is a constant to weight the two different contributions. This metric is also point separating. In fact any metric stronger that this H^1-type metric has this property. First we prove a Lemma, which also demonstrates a useful technique in obtaining control of geometrical quantities in the space of immersions.

Lemma 2.7 *If for a weak Riemannian metric G on $\mathrm{Imm}(S^1,\mathbb{R}^d)$ there exists a constant C such that*

$$G_c^1(h,h) \le C G_c(h,h)$$

holds for all $c \in \mathrm{Imm}(S^1,\mathbb{R}^d)$, then the function $\sqrt{\ell}$ is Lipschitz continuous.

PROOF. Let $c : [0,1] \to \mathrm{Imm}(S^1, \mathbb{R}^d)$ be a path in the space of immersions connecting the curves c_1 and c_2, then we can estimate

$$\begin{aligned}\partial_t \ell(t) &= \int_{S^1} \langle D_s c_t, v\rangle \,\mathrm{d}s \leq \left(\int_{S^1} |D_s c_t|^2 \,\mathrm{d}s\right)^{1/2} \left(\int_{S^1} \mathrm{d}s\right)^{1/2} \\ &\leq \frac{1}{a_1}\sqrt{\ell}\sqrt{G_c^1(c_t, c_t)} \leq \sqrt{\ell}\frac{C}{a_1}\sqrt{G_c(c_t, c_t)},\end{aligned}$$

so

$$\partial_t \sqrt{\ell(t)} \frac{\partial_t \ell}{2\sqrt{\ell}} \leq \frac{C}{2a_1}\sqrt{G_c(c_t, c_t)}.$$

Integrating this inequality in t along c results in

$$|\sqrt{\ell_{c_1}} - \sqrt{\ell_{c_2}}| = \int_0^1 |\partial_t \sqrt{\ell(t)}| \,\mathrm{d}t \frac{C}{2a_1} \leq \frac{C}{2a_1}\sqrt{G_c(c_t, c_t)}\,\mathrm{d}t = \frac{C}{2a_1}\mathrm{Len}(c).$$

Where ℓ_{c_1} and ℓ_{c_2} are the lengths of c_1 and c_2 respectively. Taking the infimum of all paths c connecting the curves we get

$$|\sqrt{\ell_{c_1}} - \sqrt{\ell_{c_2}}| \leq \frac{C}{2a_1}\mathrm{dist}(c_1, c_2).$$

Which is exactly the Lipschitz inequality desired. This allows us to prove that such metrics are distance seperating.

PROPOSITION 2.8 *If for a weak Riemannian metric G on $\mathrm{Imm}(S^1, \mathbb{R}^d)$ there exists a constant C such that*

$$G_c^1(h, h) \leq C G_c(h, h)$$

holds for all $c \in \mathrm{Imm}(S^1, \mathbb{R}^d)$, then G satisfies $\mathrm{dist}_G^{B_{i,f}}(x, y) > 0$ for $x \neq y$.

PROOF. Consider a path c connecting any two immersions c_0 and c_1, then we can estimate

$$\begin{aligned}\mathrm{Len}(c) &= \int_0^1 \sqrt{G_c(c_t, c_t)}\,\mathrm{d}t \geq C\int_0^1 \sqrt{\int_{S^1} \langle c_t, c_t\rangle |c'|\,\mathrm{d}\theta}\,\mathrm{d}t \\ &\geq C\int_0^1 \sqrt{\int_{S^1} |\langle c_t, n\rangle|^2 |c'|\,\mathrm{d}\theta}\,\mathrm{d}t \geq C\int_0^1 \left(\int_{S^1} |c'|\,\mathrm{d}\theta\right)^{-1/2} \int_{S^1} |\langle c_t, n\rangle||c'|\,\mathrm{d}\theta\,\mathrm{d}t \\ &\geq C \inf_{t\in[0,1]} \ell^{-1/2} \int_0^1 \int_{S^1} |\langle c_t, n\rangle||c'|\,\mathrm{d}\theta\,\mathrm{d}t = C(\sup_{t\in[0,1]} \ell)^{-1/2}\,\mathrm{Area}(c)\end{aligned}$$

Where $\mathrm{Area}(c)$ is the area swept out by c. This is a lower bound on $\mathrm{Len}(c)$. If $[c_0]$ and $[c_1]$ are different equivalence classes of shapes, the area swept out by

any path connecting them is bounded from below by a positive number. Taking the infimum over all paths connecting c_0 and c_1

$$\operatorname{dist}(c_0, c_1) \sup_{t\in[0,1]} \ell^{1/2} \geq C \inf_c \operatorname{Area}(c) > 0$$

The Lipschitz continuity of $\sqrt{\ell}$ implies that ℓ is bounded when taking infimum, so we get a lower bound for the distance in $B_{i,f}$.

In the next section we will have more to say about this type of metric. This proposition allows us to consider any number of arc-length derivatives

$$G^n_c(h,k) = \int_{S^1} \sum_{j=0}^{n} a_0 \langle D^j_s h, D^j_s k\rangle \, ds. \tag{2.7}$$

We will denote this the *constant coefficient Sobolev metric of order n*. An important observation is that these metrics are also invariant under the action of the euclidean group $SE(d) = SO(d) \ltimes \mathbb{R}^d$, of translations and rotations of $\mathbb{R}^d$, therefore they descend to metrics on the quotient space $\mathcal{S} = B_{i,f}/SE(d)$. If there is no L^2 term, $a_0 = 0$, the metric will have a nullspace of translations, and is only a metric on the quotient space of immersions modulo translations $\operatorname{Imm}(S^1, \mathbb{R}^d)/\mathbb{R}^d$.

2.3 Geodesics and completeness

Geodesics can be defined as usual as critical points of the length functional $L(c)$, or equivalently of the energy functional

$$E(c) = \frac{1}{2} \int_0^1 G_c(\dot{c}, \dot{c}) \, dt.$$

In finite dimensions, one can derive the expression for Euler-Lagrange equations for this equation in local coordinates to be

$$\ddot{c}^k + \Gamma^k_{ij} \dot{c}^i \dot{c}^j = 0,$$

where Γ^k_{ij} are the classical Christoffel symbols. This is a system of ODE's, and if we have a smooth metric we always have short-time solutions to the initial value problem by the Picard-Lindelöf theorem. In the infinite dimensional case things can be much more complicated. We might have the situation that no geodesic equation can be derived - this is refered to as the geodesic equation does not exist, [24]. In order to understand this, let us derive the geodesic equation for

the L^2 metric. Let $D_{c,h}$ denote the derivative at c in the direction h, then at an optimum we must have

$$D_{c,h}E(c) = \int_0^1 \int_{S^1} \langle c_t, h_t \rangle |c'| + \frac{1}{2}\langle c_t, c_t \rangle \langle h_\theta, c_\theta \rangle \frac{1}{|c_\theta|} d\theta\, dt = 0.$$

If we assume that our path of curves and variation is smooth, i.e $c \in \mathrm{Imm}(I \times S^1, \mathbb{R}^d)$ and $h \in C^\infty(I \times S^1, \mathbb{R}^d)$, we can use partial integration in t and θ to obtain

$$\begin{aligned} D_{c,h}E(c) &= \int_0^1 \int_{S^1} -\langle \partial_t(|c_\theta| c_t), h \rangle + \frac{1}{2}\langle c_t, c_t \rangle \langle h_\theta, c_\theta \rangle \frac{1}{|c_\theta|} d\theta\, dt \\ &= \int_0^1 \int_{S^1} -\langle \partial_t(|c_\theta| c_t) + \frac{1}{2}\left(\frac{|c_t|^2}{|c_\theta|} c_\theta \right)_\theta, h \rangle d\theta\, dt. \end{aligned}$$

By a classical variational argument we get a necessary condition for optimality

$$\partial_t(|c_\theta| c_t) + \frac{1}{2}\left(\frac{|c_t|^2}{|c_\theta|} c_\theta \right)_\theta = 0.$$

Which is the final form of the geodesic equation for the L^2 metric. However, the metric is also a weak metric on the space of H^1-type Sobolev curves, but here we cannot do partial integration in θ as the curves are not smooth enough, hence the variational principle does not work on this space: the geodesic equation doesn't exist. As we can see from the above expression, the geodesic equation is usually a PDE or pseudo-differential equation. If a metric admits a geodesic equation, it is also a non-trivial question if the equation has solutions, even for short time. For higher order metrics we can derive the geodesic equation in the same way by partial integration and eliminating h, the final result for the order n metric is

$$\begin{aligned} &\partial_t \left(\sum_{j=0}^{n} (-1) a_j |c'| D_s^{2j} c_t \right) = \\ &-\frac{a_0}{2} |c'| D_s(\langle c_t, c_t \rangle v) + \sum_{k=1}^{n} \sum_{j=1}^{2k-1} (-1)^{k+j} \frac{a_k}{2} |c'| D_s\left(\langle D_s^{2k-j} c_t, D_s^j c_t \rangle v \right) \end{aligned} \tag{2.8}$$

For H^n-immersions some care has to be taken as the above is expression is not well-defined. We will ignore the technicalities of this in this discussion, see [25] for details. For Sobolev type metrics or order 1 and higher the geodesic exists and is locally-well defined,

Theorem 2.9 *Let $n \geq 1$. For each $k \geq 2n+1$ the geodesic equation of G_c^n has a unique local solution in the space of Sobolev H^k-immersions. The solution depends smoothly on the initial conditions $c(0)$ and $c_t(0)$, and the domain of existence is uniform in k, and so the results also hold on* $\mathrm{Imm}(S^1, \mathbb{R}^d)$.

See [52]. For the curvature-weighted metric 2.6, it is unknown if the geodesic equation is even locally well-posed. Long time existence of geodesic is related to a classical question in Riemannian geometry. When is a manifold "complete" in a certain sense? There are usually three types of completeness to consider:

a) Metric completeness: The space (M, dist_G) is metrically complete.

b) Geodesic completeness: Any geodesic can be continued for all time.

c) Geodesic convexity: Any two points in the same connected component can be connected by a length minimizing geodesic.

In finite dimensions the famous Hopf-Rinow theorem asserts that a) and b) are equivalent, and c) follows from either of them. Hence it makes sense to talk about a complete finite dimensional Riemannian manifold, without risk of confusion. In infinite dimensions this is no longer true, and all that can be proven is

Theorem 2.10 *On a strong Riemannian manifold, metric completeness implies geodesic completeness.*

See [46, VIII, Prop. 6.5]. Even for strong metrics, this cannot be refined, as one can construct counterexamples to both implications. In [2] an example is given of a metrically and geodesically complete space, but with two points which cannot be connected by any geodesic, and in [3] an example of a geodesically complete and geodesically convex manifold which is not metrically complete.

For metrics on curves some completeness properties are known. For the L^2 metric, short time existence is an open problem. On the other hand it is easy to see that the L^2 metric is geodesically incomplete, the path

$$c(t,\theta) = \frac{3}{2}t^{2/3}(\cos\theta, \sin\theta, 0, \dots, 0)$$

satisfies the geodesic equation, but its length to zero is finite:

$$L(c) = \int_0^1 \sqrt{\int_{S^1} \frac{3}{2} d\theta} = \sqrt{3\pi}.$$

Some results for H^1 type metrics are easily obtained. To this end, let us consider the map and its inverse

$$R : \mathrm{Imm}(S^1, \mathbb{R}^d) \to C^\infty(S^1, \mathbb{R}^d/\{0\}),$$
$$R(c) = \sqrt{|c_\theta|}v$$
$$R^{-1}(q) = \int_0^\theta |q(\tau)|q(\tau)d\tau.$$

Where we denote the unit tangent of c by $v = \frac{c_\theta}{|c_\theta|}$ and the unit normal by n. This map is usually referred to as the *Square Root Velocity Transform (SRVT)*, and was introduced in [76]. It is easy to compute the differential of R

$$dR_c(h) = \frac{1}{2}|c_\theta|^{1/2}\langle D_s h, v\rangle v + |c_\theta|^{1/2}\langle D_s h, n\rangle n.$$

Where n is the unit normal vectorfield. If we consider the flat L^2 metric, on the image space of smooth curves (not necessarily regular)

$$G_c^{L^2,flat}(h,k) = \int_{S^1}\langle h,k\rangle d\theta,$$

we can compute the pullback metric on $\mathrm{Imm}(S^1,\mathbb{R}^d)$ as

$$\begin{aligned} G_c^*(h,k) &= G_c^{L^2,flat}(dR_c(h), dR_c(k)) \\ &= \int_{S^1}\langle D_s h, n\rangle\langle D_s k, n\rangle + \frac{1}{4}\langle D_s h, v\rangle\langle D_s h, k\rangle ds. \end{aligned}$$

This is a member of the family of $\dot{H}^1$ metric (with no L^2 term), sometimes referred to as elastic metrics

$$\int_{S^1} a^2\langle D_s h, n\rangle\langle D_s k, n\rangle + b^2\langle D_s h, v\rangle\langle D_s h, k\rangle ds.$$

If we for the moment instead of closed immersions, consider the open immersions $\mathrm{Imm}((0,1),\mathbb{R}^d)$ it is easy to conclude that this space is geodesically incomplete since $(C^\infty(S^1,\mathbb{R}^d/\{0\}, G_c^{L^2,flat})$ is flat and incomplete. The condition that a curve is closed is, using R^{-1},

$$\int_0^{2\pi} |q(\tau)|q(\tau)d\tau = 0.$$

This is a codimension 2 submanifold of $C^\infty(S^1,\mathbb{R}^d/\{0\}$. To see that this is also a incomplete space, notice that circles centred at zero map to circles centred at zero through R. Now the path

$$q(t,\theta) = t(\cos(\theta),\sin(\theta),0\ldots,0),$$

is a geodesic in the ambient flat space, and hence also in the submanifold of closed curves, which reaches 0 in finite time, so the space is incomplete. This can be generalized to any value of a, b in the elastic metric, see [10]. This also gives a simple way of numerically computing geodesic for these metrics. For open curves the geodesics are simply straight line in the image of R which can then by pulled back. For closed curves one can use symplectic integrators to follow the submanifold in the flat space.

However for second and higher order metrics on spaces of curves we do get completeness, in [25, 24] the following two statements are proved. First for the case of Sobolev regularity curves.

Theorem 2.11 *For $n \geq 2$ and $a_0, a_n \neq 0$, the space $(\mathrm{Imm}^n(S^1, \mathbb{R}^d), \mathrm{dist}_{G_c})$, where G_c is a constant coefficient Sobolev metric of order n, is a complete metric space, geodesically complete and any two curves c_1, c_2 in the same connected component of $\mathrm{Imm}^n(S^1, \mathbb{R}^d)$ can be connected by a minimizing geodesic.*

In the smooth category we have following result

Theorem 2.12 *For $n \geq 2$ and $a_0, a_n \neq 0$, the space $(\mathrm{Imm}(S^1, \mathbb{R}^d), G_c)$, where G_c is a constant coefficient Sobolev metric of order n, is geodesically complete.*

We will not give a proof of this statement here, since in Sec. 2.4 we shall give a proof of a generalization of this to a family of length weighted metrics which includes the constant coefficient metrics. We will just mention that a crucial ingredient of the proof is the Sobolev embedding theorem, that says that the weak regularity or order two or higher implies that the curves are also C^1.

2.4 Completeness for length-weighted metrics

The contents of this section is joint work with Martins Bruveris.

In [25, 23] it was first proved that the geodesic equation for constant Sobolev metrics on curves of order two and higher is globally well-posed: the solution exists for all time. It was also shown that the space $(\mathrm{Imm}^n(S^1, \mathbb{R}^d), G_c^n)$ is metrically complete. In this case the metric is strong, and the former result follows from the latter. As mentioned in Section 2.2, constant coefficient Sobolev metrics are not invariant under scalings of curves, so they do not descent to metrics on the full shape space $\mathcal{B}$. We can make metric scale-invariant by weighting it by appropriate powers of the length ℓ_c of the curve c :

$$G_c^{n,scale}(h,k) = \int_{S^1} a_0 \frac{1}{\ell_c^3}\langle h,k\rangle + \cdots + a_n \ell_c^{2n-3}\langle D_s^n h, D_s^n k\rangle \,\mathrm{d}s. \tag{2.9}$$

It is now our goal to extend the completeness results to a more general case of length weighted metrics

$$G_c^{n,\ell_c}(h,k) = \sum_{j=0}^{n} \int_{S^1} a_j(\ell_c)\langle D_s^j h, D_s^j k\rangle \,\mathrm{d}s. \tag{2.10}$$

Where we make the following assumptions of the coefficients functions

$$a_j \in C^\infty(\mathbb{R}_{>0}, \mathbb{R}_{\geq 0}),\ a_0, a_n > 0. \tag{2.11}$$

Note that $a_0, a_n > 0$, but is allowed to go towards zero for $\ell_c \to 0$ or $\ell_c \to \infty$. The scale-invariant metric is choice $a_j(\ell_c) = \ell_c^{2j-3}$. We will obtain a necessary and sufficient condition for the coefficient functions $a_j(\ell_c)$ to give a complete metric. The proof will proceed as in [23]. For notation, define the Sobolev norms

$$\|f\|^2_{L^2(d\theta)} \int_{S^1} |f(\theta)|^2 \,\mathrm{d}\theta,$$
$$\|f\|^2_{H^n(d\theta)} = \int_{S^1} |f(\theta)|^2 + |f^{(n)}(\theta)|^2 \,\mathrm{d}\theta,$$

the curve weighted L^2 norm

$$\|f\|^2_{L^2(ds)} = \|f\sqrt{|c_\theta|}\|^2_{L^2(d\theta)} = \int_{S^1} f^2 \,\mathrm{d}s,$$

and the curve weighted H^n norm

$$\|f\|^2_{H^n(ds)} = \int_{S^1} |f(s)|^2 + |D_s^n f(s)|^2 \,\mathrm{d}s.$$

2.4.1 Necessary conditions

We are interested in conditions on the coefficient functions a_k, such that the metric is complete. A necessary condition for completeness is that it is neither possible to shrink a curve to a point in finite time nor to blow it up toward infinity.

Fix $c_0 \in \mathrm{Imm}^n(S^1, \mathbb{R}^d)$ and consider the path $c(t,\theta) = r(t)c_0(\theta)$ with $r(0) = 1$, $r(1) = R$ and $r_t(t) > 0$. The length of this path is

$$\mathrm{Len}(c) = \int_0^1 \sqrt{G_{rc_0}(r_t c_0, r_t c_0)}\,\mathrm{d}t = \int_0^R \sqrt{G_{rc_0}(c_0, c_0)}\,\mathrm{d}r\,.$$

Writing D_c for D_s to emphasize the dependence on the curve,

$$G_{rc_0}(c_0, c_0) = \int_{S^1} \sum_{k=0}^{n} a_k(r\ell(c_0)) \left|D^k_{rc_0} c_0\right|^2 |rc_0'|\,\mathrm{d}\theta\,.$$

Assume w.l.o.g. $\ell_{c_0} = 1$. Then, since $D^k_{rc_0} c_0 = r^{-k} D^k_{c_0} c_0$,

$$G_{rc_0}(c_0, c_0) = \sum_{k=0}^{n} a_k(r) r^{1-2k} \int_{S^1} \left|D^k_{c_0} c_0\right|^2 |c_0'|\,\mathrm{d}\theta\,.$$

Note that all integrals in the above sum are positive. It follows, that

$$\begin{aligned}\lim_{R\to\infty} \mathrm{Len}(c) = \infty &\Leftrightarrow \int_1^\infty \left(\sum_{k=0}^{n} a_k(r) r^{1-2k}\right)^{1/2} \mathrm{d}r = \infty \\ &\Leftrightarrow \sum_{k=0}^{n} \int_1^\infty r^{1/2-k}\sqrt{a_k(r)}\,\mathrm{d}r = \infty \\ &\Leftrightarrow \int_1^\infty r^{1/2-k}\sqrt{a_k(r)}\,\mathrm{d}r = \infty \text{ for some } 0 \le k \le n\,.\end{aligned}$$

Thus a necessary condition for completeness is that at least one of the integrals

$$I_{k,\infty} = \int_1^\infty r^{1/2-k}\sqrt{a_k(r)}\,\mathrm{d}r$$

diverges.

Similarly one can consider the shrinking of curves by setting $r(0) = 1$, $r(1) = R > 0$ and $r_t(t) < 0$. Then

$$\lim_{R\to 0} \mathrm{Len}(c) = \infty \Leftrightarrow \int_0^1 r^{1/2-k}\sqrt{a_k(r)}\,\mathrm{d}r = \infty \text{ for some } 0 \le k \le n\,.$$

Thus, the second necessary condition is the divergence of at least one of the integrals

$$I_{k,0} = \int_0^1 r^{1/2-k}\sqrt{a_k(r)}\,\mathrm{d}r\,.$$

The main result is that these two conditions are also sufficient for the metric to be complete. We define for Sobolev metrics of order n of the form (2.10) the two properties

$$\max_{1\le k\le n} I_{k,0} = \infty \tag{I_0}$$

$$\max_{1\le k\le n} I_{k,\infty} = \infty\,. \tag{I_∞}$$

These are sufficient conditions to prevent finite time shrinkage and blow up of curves along radial paths $c(t,\theta) = r(t)c_0(\theta)$. We will show that they also prevent finite time shrinkage and blow up along arbitrary paths.

Remark 1 *Note that $1 \le k \le n$ in (I_0) and (I_∞). The case when only $I_{0,\infty} = \infty$ or $I_{0,0} = \infty$ is be more subtle, and one can ask if this is also sufficient for completeness. In Sec. 2.4.4 we shall give two families of metrics which satisfies separately only $I_{0,0} = \infty$ and $I_{0,\infty} = \infty$, and which are not metrically complete.*

2.4.2 Controlling length

In this section we prove that the length ℓ_c is bounded on geodesic balls. This result constitute the main ingredient to generalize the completeness for constant coefficient metrics. First we need a set of Poincare type inequalities in the curve weighted norm

Lemma 2.13 *For a curve $c \in \operatorname{Imm}^2(S^1, \mathbb{R}^d)$ and $h \in H^2(S^1, \mathbb{R}^d)$ the following inequalities hold:*

- $\|h\|_\infty^2 \leq \frac{2}{\ell_c}\|h\|_{L^2(ds)}^2 + \frac{\ell_c}{2}\|D_s h\|_{L^2(ds)}^2$
- $\|D_s h\|_\infty^2 \leq \frac{\ell_c}{4}\|D_s^2 h\|_{L^2(ds)}^2$
- $\|D_s h\|_{L^2(ds)}^2 \leq \frac{\ell_c^2}{4}\|D_s^2 h\|_{L^2(ds)}^2$

If $c \in \operatorname{Imm}^n(S^1, \mathbb{R}^d)$ and $h \in \operatorname{Imm}^n(S^1, \mathbb{R}^d)$ then for $0 \leq k \leq n$,

- $\|D_s^k\|_{L^2(ds)}^2 \leq \|h\|_{L^2(ds)}^2 + \|D_s^n h\|_{L^2(ds)}^2$

A proof can be found in [25]. This allows us to prove that the length ℓ_c is locally bounded on geodesic balls.

Lemma 2.14 *Let G be a length-weighted Sobolev metric of order $n \geq 2$ satisfying (I_0) and (I_∞). Then, given $c_0 \in \operatorname{Imm}^n(S^1, \mathbb{R}^d)$ and $R > 0$, there exists $C = C(c_0, R) > 0$ such that*

$$C^{-1} \leq \ell_c \leq C\,,$$

holds for all $c \in \operatorname{Imm}^n(S^1, \mathbb{R}^d)$ with $\operatorname{dist}(c_0, c) < R$.

Especially the last part shows that if we add derivatives of order between 0 and n the norm doesn't change. Proof. Let $c_1 \in \operatorname{Imm}^n(S^1, \mathbb{R}^d)$ with $\operatorname{dist}(c_0, c_1) < R$ and let $c(t, \theta)$ be a path connecting c_0 to c_1. Computing

$$\partial_t \ell_c = \int_{S^1} \langle D_s c_t, v\rangle \,\mathrm{d}s\,,$$

and we can estimate using Cauchy–Schwartz and Lemma 2.13

$$\begin{aligned}|\partial_t \ell_c| &\leq \int_{S^1} |\langle D_s c_t, v\rangle|\,|c'|\,\mathrm{d}\theta \leq \sqrt{\int_{S^1} |c'|\,\mathrm{d}\theta}\sqrt{\int_{S^1} |\langle D_s c_t, v\rangle|^2\,|c'|\,\mathrm{d}\theta} \\ &\leq \ell_c^{1/2}\|D_s c_t\|_{L^2(ds)} \leq \ell_c^{k-1/2}\|D_s^k c_t\|_{L^2(ds)}\,.\end{aligned}$$

Now define the function

$$W(r) = \sum_{k=1}^{n} \int_1^r \varrho^{1/2-k} \sqrt{a_k(\varrho)}\, \mathrm{d}\varrho,$$

the assumption on G ensures that $W'(r) > 0$, while (I_0) and (I_∞) implies that $\lim_{t\to 0} W(t) = -\infty$ and $\lim_{t\to\infty} W(t) = \infty$, respectively. Hence $W : (0,\infty) \to \mathbb{R}$ is a diffeomorphism. Taking the time derivative and combining with the above inequality yields

$$|\partial_t W(\ell_c)| \le \sum_{k=1}^{n} \ell_c^{1/2-k} \sqrt{a_k(\ell_c)} |\partial_t \ell_c| \le \sum_{k=1}^{n} \sqrt{a_k(\ell_c)} \|D_s^k c_t\|_{L^2(ds)} \le \sqrt{G_c(c_t, c_t)}$$

Integrating along c leads to

$$|W(\ell_{c_1}) - W(\ell_{c_0})| \le \int_0^1 |\partial_t W(\ell_c)|\, \mathrm{d}t \le \int_0^1 \sqrt{G_c(c_t, c_t)}\, \mathrm{d}t = \mathrm{Len}(c)\,,$$

By taking the infimum over all paths we arrive at

$$|W(\ell_{c_1}) - W(\ell_{c_0})| \le \mathrm{dist}(c_0, c_1).$$

That is, $W(\ell_c)$ is Lipschitz continuous w.r.t dist, and so is bounded on any metric ball $B(c_0, R)$. Since W is a diffeomorphism, also ℓ_c must be bounded from above and away from 0. □

We will also need that $\log|c'|$ is locally Lipschitz continuous w.r.t to the geodesic distance.

Lemma 2.15 *Let $n \ge 2$. The following function is locally Lipschitz continuous,*

$$\log|c'| : (\mathrm{Imm}^n(S^1, \mathbb{R}^2), \mathrm{dist}) \to L^\infty(S^1, \mathbb{R})$$

Equivalently, there exists a constant $D = D(c_0, R)$ such that for all c with $\mathrm{dist}(c_0, c) < R$ we have the bound

$$D^{-1} \le |c'(\theta)| \le D$$

Proof. Let $c_1, c_2 \in \mathrm{Imm}^n(S^1, \mathbb{R}^2)$ with $\mathrm{dist}(c_0, c_1) < R$, and $c(t, \theta)$ be a path connecting them. Then we have

$$\partial_t(\log|c'(\theta)|) = \langle D_s c_t, v\rangle.$$

Lemma 2.14 implies that ℓ_c is bounded on the ball $B(c_0, R)$, and since a_n is smooth, $a_n(\ell_c)$ is bounded as well. Using this we can find a constant $A = A(c_0, R)$ such that by Lemma 2.13, we get

$$\|D_s c_t\|_{L^\infty} \le \frac{\ell_c^{1/2}}{2} \|D_s^2 c_t\|_{L^2(ds)} \le \frac{\ell_c^{n-3/2}}{2^{n-1}} \|D_s^n c_t\|_{L^2(ds)} \le A\sqrt{G_c(c_t, c_t)}.$$

Repeating the same type of argument as in Lemma 2.14 by integrating, using the estimate and taking the infimum over paths c, we get

$$\|\log|c_1'| - \log|c_2'|\|_{L^\infty} \leq A\,\mathrm{dist}(c).$$

The local boundedness of $|c'(\theta)|$ is then immediate. □

In [25, 23] the following assumption on a metric is central to the remaining results

> Given a metric ball $B(c_0, r)$ in $\mathrm{Imm}(S^1, \mathbb{R}^d)$, there exists a constant C, such that
> $$\|h\|_{H^n(ds)} = \int_{S^1} |h|^2 + |D_s h|^2 \,\mathrm{d}s \leq C G_c(h,h) \qquad (H_n)$$
> holds for all $c \in B(c_0, r)$.

The constant coefficient metric with $a_0, a_n > 0$ satisfies (H_n) by the last part of Lemma (2.13). We now show that the length weighted-metric has this property.

Lemma 2.16 *The length-weighted metric $G_c^{n,\ell}$ on $\mathrm{Imm}(S^1, \mathbb{R}^d)$ satisfying (I_0) and (I_∞), satisfies (H_n).*

Proof. Let $B(c_0, r)$ be a metric ball w.r.t. G_c^{n,ℓ_c}. By Lemma 2.14 ℓ_c is bounded on $B(c_0, r)$ by the constant C. $a_j(\ell_c)$ is smooth so is also bounded, then we have

$$\int_{S^1} \sum_{j=0}^{n} \min_{C^{-1} \leq \ell_c \leq C} a_j(\ell) |D_s^j h|^2 \,\mathrm{d}s \leq G_c^{n,\ell}(h,h)$$

As $a_0(\ell_c), a_n(\ell_c) > 0$ the left side defines a constant coefficient metric which satisfies (H_n), so

$$\|h\|_{H^n(ds)} \lesssim G_c^{n,\ell_c}(h,h)$$

holds for all $c \in B(c_0, r)$. □

In [25] Lemma 2.14 was shown to hold for metrics satisfying (H_n), but here we needed to show the reverse implication first.

2.4.3 The main result

Now we have all the tools to finish the completeness of the metric. The remaining part proceeds as in [25, 23] with the Lemmas replaced by their length-weighted counterparts. First we prove that the metric is uniformly equivalent to the flat Sobolev metric on any metric balls. For the standard $H^n(ds)$-norm it was proven that the Riemannian metric is uniformly equivalent to the flat $H^n(d\theta)$-norm. From this we can obtain the same result for the length-weighted metric

Proposition 2.17 *Let $n \geq 2$ and G a weak Riemannian metric on $\mathrm{Imm}(S^1, \mathbb{R}^d)$ satisfying (H_n). Then, given a metric ball $B(c_0, r)$ in $\mathrm{Imm}(S^1, \mathbb{R}^d)$ there exists a constant C such that*

$$C^{-1}\|h\|_{H^n(d\theta)} \leq \|h\|_{H^n(ds)} \leq C\|h\|_{H^n(d\theta)}$$

holds for all $c \in B(c_0, r)$ and all $h \in H^n(S^1, \mathbb{R}^d)$

Corollary 2.18 *Let $n \geq 2$ and G be a length-weighted Sobolev metric of order n, satisfying (I_0) and (I_∞). Then, given a metric ball $B(c_0, r)$ in $\mathrm{Imm}(S^1, \mathbb{R}^d)$ there exists a constant C such that*

$$C^{-1}\|h\|_{H^n(d\theta)} \leq \sqrt{G_c(h,h)} \leq C\|h\|_{H^n(d\theta)}$$

holds for all $c \in B(c_0, r)$ and all $h \in H^n(S^1, \mathbb{R}^d)$

Proof. Proceeding as in the proof of Lemma 2.16, replacing min by max and using the last part of Lemma 2.13 we get

$$\sqrt{G_c(h,h)} \leq C\|h\|_{H^n(ds)}.$$

Along with (H_n) we get the equivalence witht the $H^n(ds)$ norm

$$C^{-1}\|h\|_{H^n(ds)} \leq \sqrt{G_c(h,h)} \leq C\|h\|_{H^n(ds)}.$$

Combining with the result of the proposition gives the result □ .

Remark 1 *Note that all the results so far can be extended to hold on $\mathrm{Imm}^n(S^1, \mathbb{R}^d)$ by a limiting argument.*

Lemma 2.19 *Let $n \geq 2$ and G be a length-weighted Sobolev metric of order n, satisfying (I_0) and (I_∞). Then*

1. *Given a metric ball $B(c_0, r)$ in $\operatorname{Imm}^n(S^1, \mathbb{R}^d)$ there exists C such that*

$$\|c_1 - c_2\|_{H^n(d\theta)} \leq C \operatorname{dist}(c_1, c_2)$$

   holds for all $c_1, c_2 \in B(c_0, r)$
2. *Given $c_0 \in \operatorname{Imm}^n(S^1, \mathbb{R}^d)$, there exists $r > 0$ and C such that*

$$\operatorname{dist}(c_1, c_2) \leq C\|c_1 - c_2\|_{H^n(d\theta)}$$

   holds for all c_1, c_2 in $B(c_0, r)$.

PROOF. Let $c_1, c_2 \in B(c_0, r)$ and c be a path of length $L(c) < r$ connecting them in $B(c_0, r)$, then by Cor. 2.18

$$\|c_1 - c_2\|_{H^n(d\theta)} \leq \int_0^1 \|c_t(t)\|_{H^n(d\theta)} \,\mathrm{d}t \leq C \int_0^1 \sqrt{G_c(c_t, c_t)} \,\mathrm{d}t \leq CL(c),$$

and the constant depends only on c_0 and r. Taking the infimum over all paths, and note that all shortest paths have to be contained in $B(c_0, r)$ we get the first part.
Let $c_0 \in \operatorname{Imm}^n(S^1, \mathbb{R}^d)$, and U be an open convex neighbourhood in the flat ambient space $\operatorname{Imm}^n(S^1, \mathbb{R}^d)$. The metric G is a smooth strong Riemannian metric, so by [46, VII Prop. 6.1] the metric distance induces the manifold topology, there exists a ball $B(c_0, r$ w.r.t to dist_G contained in U. For $c_1, c_2 \in B(c_0, r)$ let $c(t) = c_1 + t(c_2 - c_1)$ be the linear path connecting them. By Cor. 2.18 we get

$$\operatorname{dist}_G(c_1, c_2) \leq L(c) = \int_0^1 \sqrt{G_c(c_2 - c_1, c_2 - c_1)} \,\mathrm{d}t \leq C\|c_2 - c_1\|_{H^n(d\theta)}.$$

This is the second part. □

Now we can prove the main result about the completeness of $G_c^{n,\ell}$.

THEOREM 2.20 *Let $n \geq 2$ and G be a length-weighted Sobolev metric of order n, satisfying (I_0) and (I_∞). Then*

1. $(\operatorname{Imm}^n(S^1, \mathbb{R}^d), \operatorname{dist})$ *is a complete metric space.*
2. $(\operatorname{Imm}^n(S^1, \mathbb{R}^d), G)$ *is geodesically complete.*

PROOF. Let $(c^j)_{j\in\mathbb{N}}$ be a Cauchy sequence w.r.t the geodesic distance. The sequence is contained in a ball $B(c_0, r)$, so by the first part of Lemma 2.19 it is also a Cauchy sequence w.r.t. the $H^n(d\theta)$-norm. Hence there exists a limit

$c^* \in H^n(S^1, \mathbb{R}^d)$ such that $\|c^j - c^*\|_{H^n(d\theta)} \to 0$. By Lemma 2.15 there exists a constant C such that we have the point-wise lower bound $\|c^j_\theta(\theta)\| \geq C > 0$. This must also hold for the limit, so it is still an immersion: $c^* \in \mathrm{Imm}^n(S^1, \mathbb{R}^d)$. The second part of Lemma 2.19 implies that $\mathrm{dist}(c^j, c^*) \to 0$, hence the first part of the statement is shown. For a smooth strong Riemannian metric, a part of the Hopf-Rinow theorem holds, that metric completeness implies geodesic completeness, see [46, VIII Prop 6.5], this is the second part of the statement.

2.4.4 Counterexamples

The main completeness results was proven under the assumption that (I_0) and (I_∞) holds, i.e. we can control the behaviour of the first or higher order coefficients when the length ℓ_c of a curve goes to zero or infinity. When taking a time derivative $\partial_t \ell_c$ we can use the first order term in the metric, or a Poincare inequality and the higher order terms, to estimate this quantity, and use this control to obtain completeness. This technique and the L^2 terms does not allow us to control $\partial_t \ell_c$, so one may ask if completeness still holds when (I_0) or (I_∞) only for $k = 0$, and where radial paths are still have infinite energy. We will now present two families of metrics which satisfy exactly these conditions only for $k = 0$, but are not metrically complete. Whether they are geodesically complete is unknown.

We consider the metric

$$G_c(h,k) = \int_{S^1} a_0(\ell_c)\langle h, k\rangle + a_2(\ell_c)\langle D_s^2 h, D_s^2 k\rangle \,\mathrm{d}s\,.$$

If we let $a_0(\ell_c) = \ell_c^{-3}$ for $\ell_c \to \infty$, then

$$I_{0,\infty} = \int_1^\infty r^{1/2}.r^{-3/2}\,\mathrm{d}r = \log r\big|_{r=1}^\infty = \infty\,.$$

Similarly, if $a_0(\ell_c) = \ell_c^{-3}$ for $\ell_c \to 0$, then $I_{0,0} = \infty$.

We shall consider the following two cases:

1. $a_0(\ell_c) = \ell_c^{-3}$ and $a_2(\ell_c) = \ell_c^p$, $p < 1$ for $\ell_c \to \infty$. Then $I_{0,\infty} = \infty$ and $I_{2,\infty} < \infty$, but we can find a Cauchy sequence $(c_n)_{n\in\mathbb{N}}$ with $\ell_{c_n} \to \infty$; hence the Riemannian metric cannot be metrically complete.
2. $a_0(\ell_c) = \ell_c^{-3}$ and $a_2(\ell_c) = \ell_c^p$, $p > 1$ for $\ell_c \to 0$. Then $I_{0,0} = \infty$ and $I_{2,0} < \infty$, but we can find a Cauchy sequence $(c_n)_{n\in\mathbb{N}}$ with $\ell_{c_n} \to 0$; hence the Riemannian metric cannot be metrically complete.

Observe that ℓ_c^{-3} is the only choice of polynomial which give $I_{0,\infty} = I_{0,0} = \infty$ simultaneously. For the H^2 condition $p \geq 1$ ensures $I_{2,\infty} < \infty$ while $I_{2,0} = \infty$, and $p \leq 1$ gives $I_{2,0} < \infty$ and $I_{2,\infty} = \infty$. The borderline case $p = 1$ has $I_{2,0} = I_{2,\infty} = \infty$ so it is complete.

Consider two sequences $(r_n)_{n\in\mathbb{N}}$, $(\lambda_n)_{n\in\mathbb{N}}$ with $r_{n+1} = ar_n$, $\lambda_{n+1} = b\lambda_n$ where $0 < a < 1 < b$ and $0 < r_0 < 1$, $2 < \lambda_0$. Then $r_n \searrow 0$, $\lambda_n \nearrow \infty$ and the sequence $(r_n)_{n\in\mathbb{N}}$ is decreasing while $(\lambda_n)_{n\in\mathbb{N}}$ is increasing; furthermore $\lambda_n \geq 2$. Define the sequence of curves

$$c_n(\theta) = r_n \left(1 + \varepsilon \sin(\lambda_n \theta)\right) \vec{n}\,,$$

with $\vec{n} = (\cos\theta, \sin\theta)$ and $0 < \varepsilon < \frac{1}{3}$. Set $\vec{v} = (-\sin\theta, \cos\theta)$; then $\vec{n}' = \vec{v}$ and $\vec{v}' = -\vec{n}$. This is a circle of radius r_n with λ_n bumps of amplitude ϵ.

We want to estimate the geodesic distance $\operatorname{dist}(c_n, c_{n+1})$. To do so, we define the intermediate curve

$$\tilde{c}_n(\theta) = r_{n+1} \left(1 + \varepsilon \sin(\lambda_n \theta)\right) \vec{n}\,.$$

We will estimate $\operatorname{dist}(c_n, \tilde{c}_n)$ and $\operatorname{dist}(\tilde{c}_n, c_{n+1})$ separately using the linear path between the curves. The derivatives of c_n are

$$\begin{aligned}
c_n'(\theta) &= r_n \left(1 + \varepsilon \sin(\lambda_n \theta)\right) \vec{v} + \varepsilon r_n \lambda_n \cos(\lambda_n \theta) \vec{n} \\
c_n''(\theta) &= 2\varepsilon r_n \lambda_n \cos(\lambda_n \theta) \vec{v} - r_n \left(1 + \varepsilon \left(1 + r_n \lambda_n^2\right) \sin(\lambda_n \theta)\right) \vec{n}\,.
\end{aligned}$$

We have the following pointwise estimates,

$$\begin{aligned}
|c_n(\theta)| &\leq r_n(1+\varepsilon) \lesssim r_n \\
|c_n'(\theta)| &\leq r_n \left(1 + \varepsilon + \varepsilon\lambda_n\right) \leq r_n \left(2 + \lambda_n\right) \lesssim r_n \lambda_n \\
|c_n''(\theta)| &\leq 2r_n \left(1 + \varepsilon + 2\varepsilon\lambda_n + \varepsilon\lambda_n^2\right) \leq 2r_n (2 + 2\lambda_n + \lambda_n^2) \lesssim r_n \lambda_n^2\,.
\end{aligned}$$

For $\tilde{c}_n$ we have the same estimates

$$|\tilde{c}_n(\theta)| \lesssim r_n \qquad |\tilde{c}_n'(\theta)| \lesssim r_n \lambda_n \qquad |\tilde{c}_n''(\theta)| \lesssim r_n \lambda_n^2\,,$$

because $r_{n+1} \lesssim r_n$.

To estimate $\operatorname{dist}(c_n, \tilde{c}_n)$ we define the path

$$\begin{aligned}
c(t,\theta) &= (1-t) c_n(\theta) + t\tilde{c}_n(\theta) \\
&= (1-t) r_n \left(1 + \varepsilon \sin(\lambda_n \theta)\right) \vec{n} + t r_{n+1} \left(1 + \varepsilon \sin(\lambda_n \theta)\right) \vec{n} \\
&= \left(r_n + t(r_{n+1} - r_n)\right) \left(1 + \varepsilon \sin(\lambda_n \theta)\right) \vec{n}
\end{aligned}$$

Then

$$c'(t,\theta) = \left(r_n + t(r_{n+1} - r_n)\right) \left[\varepsilon \lambda_n \cos(\lambda_n \theta) \vec{v} + \left(1 + \varepsilon \sin(\lambda_n \theta)\right) \vec{n}\right]$$

and because $\langle \vec{n}, \vec{v} \rangle = 0$, we have the lower bound

$$|c'| \geq (r_n + t(r_{n+1} - r_n))\,(1 + \varepsilon \sin(\lambda_n \theta)) \geq (1 - \varepsilon) r_{n+1} \gtrsim r_n \,.$$

We will also need a slightly sharper lower bound for the length ℓ_c. Starting from

$$|c'(t,\theta)| \geq (r_n + t(r_{n+1} - r_n)) \varepsilon \lambda_n \,|\cos(\lambda_n \theta)| \gtrsim r_n \lambda_n \,|\cos(\lambda_n \theta)| \,,$$

we obtain by integration, since $\int_0^{2\pi} |\cos \lambda_n \theta| \,\mathrm{d}\theta = 4$ for $\lambda_n \in \mathbb{N}$, the estimate $\ell_c \gtrsim r_n \lambda_n$. Thus we have

$$r_n \lambda_n \lesssim \ell_c \lesssim r_n \lambda_n \,.$$

Next we need to estimate the velocity of the path

$$\partial_t c(t,\theta) = \tilde{c}_n(\theta) - c_n(\theta) \,.$$

The simple estimate is

$$|\partial_t c| \lesssim r_n \,,$$

and therefore

$$\int_{S^1} |\partial_t c|^2 \,|c'| \,\mathrm{d}\theta \lesssim r_n^3 \lambda_n \,.$$

We also need to estimate $D_s^2(\partial_t c)$. For this we use the formula

$$D_s^2 h = \frac{1}{|c'|} \left(\frac{1}{|c'|} h' \right)' = \frac{1}{|c'|^2} h'' - \frac{1}{|c'|^4} \langle c', c'' \rangle h' \,.$$

Up to constants we obtain

$$\begin{aligned} \left| D_s^2(\partial_t c) \right| &\lesssim r_n^{-2} \cdot r_n \lambda_n^2 + r_n^{-4} \cdot r_n \lambda_n \cdot r_n \lambda_n^2 \cdot r_n \lambda_n \\ &\lesssim r_n^{-1} \lambda_n^2 + r_n^{-1} \lambda_n^4 \lesssim r_n^{-1} \lambda_n^4 \,. \end{aligned}$$

Thus

$$\int_{S^1} \left| D_s^2(\partial_t c) \right|^2 |c'| \,\mathrm{d}\theta \lesssim r_n^{-2} \lambda_n^8 \cdot r_n \lambda_n \lesssim r_n^{-1} \lambda_n^9 \,.$$

We will obtain similar estimates for $\operatorname{dist}(\tilde{c}_n, c_{n+1})$. Define the path

$$\begin{aligned} c(t,\theta) &= (1-t)\tilde{c}_n(\theta) + t c_{n+1}(\theta) \\ &= \left[r_{n+1} + \varepsilon r_{n+1} \left((1-t) \sin(\lambda_n \theta) + t \sin(\lambda_{n+1} \theta) \right) \right] \vec{n} \,. \end{aligned}$$

Then

$$\begin{aligned} c'(t,\theta) = \Big[r_{n+1} &+ \varepsilon r_{n+1} \left((1-t) \sin(\lambda_n \theta) + t \sin(\lambda_{n+1} \theta) \right) \Big] \vec{v} \\ &+ \varepsilon r_{n+1} \left((1-t) \lambda_n \cos(\lambda_n \theta) + t \lambda_{n+1} \cos(\lambda_{n+1} \theta) \right) \vec{n} \,, \end{aligned}$$

and

$$|c'| \geq \left| r_{n+1} + \varepsilon r_{n+1} \left((1-t)\sin(\lambda_n \theta) + t \sin(\lambda_{n+1}\theta) \right) \right| \geq (1-2\varepsilon) r_{n+1} \gtrsim r_n \,.$$

We also have the estimate

$$|c'(t,\theta)| \geq \varepsilon r_{n+1} \left| (1-t)\lambda_n \cos(\lambda_n \theta) + t\lambda_{n+1} \cos(\lambda_{n+1}\theta) \right| ,$$

which allows us to find a lower bound for the length

$$\begin{aligned} \ell_c &\gtrsim r_n \lambda_n \int_0^{2\pi} \left| (1-t)\cos(\lambda_n \theta) + tb\cos(b\lambda_n \theta) \right| \mathrm{d}\theta \\ &= r_n \lambda_n \int_0^{2\pi} \left| (1-t)\cos(\theta) + tb\cos(b\theta) \right| \mathrm{d}\theta \\ &\gtrsim r_n \lambda_n \,. \end{aligned}$$

The last inequality is independent of t, because the path $t \mapsto (1-t)\cos\theta + tb\cos(b\theta)$ into L^1 is continuous and does not pass through the zero function. Thus we have again the upper and lower bounds

$$r_n \lambda_n \lesssim \ell_c \lesssim r_n \lambda_n \,,$$

and we can derive the estimates

$$\int_{S^1} |\partial_t c|^2 \, |c'| \, \mathrm{d}\theta \lesssim r_n^3 \lambda_n \qquad\qquad \int_{S^1} \left| D_s^2 \partial_t c \right|^2 |c'| \, \mathrm{d}\theta \lesssim r_n^{-1} \lambda_n^9 \,.$$

as before.

Case (1). Note that by the bounds on ℓ_c, we have

$$a_0(\ell_c) \lesssim r_n^{-3} \lambda_n^{-3} \,, \qquad\qquad a_2(\ell_c) \lesssim r_n^p \lambda_n^p \,.$$

With $p < 1$. Therefore

$$\mathrm{dist}(c_n, \tilde{c}_n)^2 \lesssim r_n^{-3} \lambda_n^{-3} \cdot r_n^3 \lambda_n + r_n^p \lambda_n^p \cdot r_n^{-1} \lambda_n^9 \lesssim \lambda_n^{-2} + r_n^{p-1} \lambda_n^{p+9} \,.$$

We choose $r_n = \lambda_n^q$, for some q. We get the estimate

$$\mathrm{dist}(c_n, \tilde{c}_n)^2 \lesssim \lambda_n^{-2} + \lambda_n^{q(p-1)+p+9},$$

Choosing q to satisfy

$$q > -\frac{p+9}{p-1} > -1$$

where the last inequality holds for any $p < 1$, this gives the estimate

$$\mathrm{dist}(c_n, \tilde{c}_n)^2 \lesssim \lambda_n^r$$

with $r = q(p-1) + p + 9 < 0$, and the same estimates hold for $\operatorname{dist}(\tilde{c}_n, c_{n+1})$. Therefore

$$\operatorname{dist}(c_n, c_{n+1}) \lesssim \lambda_n^r \,,$$

and since $\sum_n \lambda_n^r < \infty$, it follows that $(c_n)_{n\in\mathbb{N}}$ is a Cauchy sequence and

$$\ell_{c_n} \gtrsim r_n \lambda_n = \lambda_n^{1+q} \to \infty \,.$$

Case (2). Now using the same bounds on ℓ_c, we have the estimates

$$a_0(\ell_c) \lesssim r_n^{-3} \lambda_n^{-3} \,, \qquad\qquad a_2(\ell_c) \lesssim r_n^p \lambda_n^p \,.$$

With $p > 1$. We choose $r_n = \lambda_n^q$, and get the same estimate on the geodesic distance as before

$$\operatorname{dist}(c_n, \tilde{c}_n)^2 \lesssim \lambda_n^{-2} + \lambda_n^{q(p-1)+p+9},$$

Choosing q to now satisfy

$$q < -\frac{p+9}{p-1} < -1$$

with the last inequality holding for all p, gives

$$\operatorname{dist}(c_n, \tilde{c}_n)^2 \lesssim \lambda_n^r$$

with $r < 0$, and the same estimate holds for $\operatorname{dist}(\tilde{c}_n, c_{n+1})$. Therefore

$$\operatorname{dist}(c_n, c_{n+1}) \lesssim \lambda_n^r \,,$$

and as before it follows that $(c_n)_{n\in\mathbb{N}}$ is a Cauchy sequence and

$$\ell_{c_n} \lesssim \lambda_n^{1+q} \to 0 \,.$$

2.5 Sobolev metrics on constant speed curves

A simple fact is that any C^1 curve can be reparametrized to unit speed, or without changing the domain of definition $I = [0, 2\pi]$, reparametrized to constant speed such that $|c'| = \ell/2\pi$. Hence as an alternative to shape space of immersions modulo reparametrizations, we can study the space of closed constant speed curves of Sobolev regularity, as a submanifold of the space of immersions presented earlier On this space we shall consider a class of Sobolev metrics of order two and higher. We show that the space is a submanifold, derive an expression for the orthogonal projection on the tangent space and equivalently the geodesic equation. Utilizing the result of the previous section we can establish some easy results on existence of geodesics.

Some words about earlier work: The L^2 metric on unit-speed curves has been studied by Preston in [63] and [62] for a whip-boundary condition: one end fixed and one end free, and periodic curves. Here local existence of geodesics was obtained through energy methods. Constant speed curves are mentioned, but local existence is not established. In [64] the H^1 metric on C^1 unit-speed curves was shown to have local existence of geodesics, for both whips and closed curves. Global existence was also shown for whips, and its indicated that this fails for closed curves. In [16] the general case of smooth volume preserving immersions was shown to be a Frechet submanifold of all immersions, this generalizes unit-speed curves.

2.5.1 Manifold structure

We will consider the space of unit- and constant-speed curves, which we define for $n \geq 1$

$$\mathcal{A}_1 = \{c \in H^n(S^1, \mathbb{R}^d) : |c'| = 1\},$$
$$\mathcal{A}_\ell = \{c \in H^n(S^1, \mathbb{R}^d) : |c'| = \ell, \ell \in \mathbb{R}_{>0}\}.$$

For $n \geq 2$, the Sobolev Embedding Lemma A.1 gives the continuous inclusion of $H^n(S^1, \mathbb{R}^d) \hookrightarrow C^1(S^1, \mathbb{R}^d)$, so the condition on the tangent vector can be made point wise. For $n = 1$ we have to interpret equality in the sense of L^2 functions. Now let $H^n(S^1)/\mathbb{R}$ denote the space of Sobolev functions modulo constant functions, i.e. equivalent classes of $[f(x)] = [f(x) + t]$. It can be represented as the space of all functions with mean zero $\int_{S^1} f(x)\,\mathrm{d}x = 0$, which is a closed subspace of $H^n(S^1)$. Recall that we denote the unit tangent by v and unit normal by n. We show that the spaces are submanifolds,

Theorem 2.21 *For $n \geq 2$ the spaces $\mathcal{A}_1$ and $\mathcal{A}_\ell$ are smooth Hilbert submanifolds of $H^n(S^1, \mathbb{R}^d)$, and the tangent spaces are given as*

$$T_c\mathcal{A}_1 = \{h \in H^n(S^1, \mathbb{R}^d) : \langle v, h' \rangle = 0\}, \tag{2.12}$$
$$T_c\mathcal{A}_\ell = \{h \in H^n(S^1, \mathbb{R}^d) : \langle v, h' \rangle = t \in \mathbb{R}\} \tag{2.13}$$

Proof. This is an application of the inverse function theorem in Banach spaces, see [30, Thm. 10.2.1]. Since $n \geq 2$, we know that $H^{n-1}(S^1)$ is an stable under multiplication, so $|c'|^2 \in H^{n-1}$. We define the map

$$L : H^n(S^1, \mathbb{R}^d) \to H^{n-1}(S^1),$$
$$L(c) = |c'|^2.$$

Observe that $\mathcal{A}_1 = L^{-1}(1)$ and $\mathcal{A}_\ell = L^{-1}([0])$. The first and second derivatives of L are

$$dL_c(h) = 2\langle c', h'\rangle, \quad d^2L_c(h,k) = 2\langle h', k'\rangle.$$

All higher order derivatives are zero, so L is smooth. We need to construct closed subspaces s.t. $H^n(S^1,\mathbb{R}^d) = E \times F$ and the partial derivative $d_F L_c$ is bijective bounded map. Let $E = \ker dL_c$ and $F = E^\perp$ its orthogonal complement. If dL_c is bounded and surjective then so is $d_F L_c$. A simple computation shows that dL_c is bounded

$$\begin{aligned}\|dL_c(h)\|^2_{H^{n-1}} &= 4\int_{S^1} \langle c', h'\rangle^2 + |\partial_\theta^{n-1}(\langle c', h'\rangle)|^2 \,\mathrm{d}\theta \\ &\le C\int_{S^1} |h'|^2 + \sum_{k=1}^{n} \binom{n}{k} |\langle c^{(k)}, h^{(n-k+1)}\rangle|^2 \,\mathrm{d}\theta \\ &\le C\int_{S^1} |h|^2 + |h^{(n)}|^2 \,\mathrm{d}\theta.\end{aligned}$$

Where we have used the triangle inequality for the first inequality. For the second inequality, we have estimated the terms from the Leibniz rule as follows. All the terms with $k \le n-1$ we can bound $c^{(k)}(\theta)$ point wise by $\|c^{(k)}\|_{L^\infty}$, which leaves the $\|h^{(n-k)}\|_{L^2(\theta)}(\theta)$ term. The highest order term $|\langle c^{(n)}, h'\rangle|^2$ we can bound using the continuous embedding into C^1

$$\int_{S^1} |\langle c^{(n)}, h'\rangle|^2 \,\mathrm{d}\theta \le \|h'\|^2_{L^\infty} \int_{S^1} |c^{(n)}(\theta)|^2 \,\mathrm{d}\theta \le C\|c^{(n)}\|^2_{L^2(d\theta)} \|h\|^2_{H^2(d\theta)},$$

and Lemma 2.13 allows us to estimate all the middle order terms by the highest and lowest. Now for surjectivity it is easy to check that the element

$$J_c(f) = \int_0^s \frac{1}{2} f(\theta) v(\theta) + g n(\theta) \,\mathrm{d}\theta.$$

for some $g \in H^{n-1}(S^1)$ is in $H^n(S^1,\mathbb{R}^d)$ if $f \in H^{n-1}(S^1)$ and satisfies $dL_c(J_c(f)) = f$. We need to choose g such that $J_c(f)$ is periodic, i.e. $\int_{S^1} J_c(f)\,\mathrm{d}\theta = 0$. If we let $g = \langle (a,b), n\rangle$, then

$$\int_{S^1} n^T n \,\mathrm{d}x \begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{2}\int_{S^1} f v \,\mathrm{d}\theta$$

Let $n = (-\sin(\alpha), \cos(\alpha))$, then the determinant of the system is zero iff

$$\int_{S^1} \sin(\alpha)^2 \,\mathrm{d}\theta \int_{S^1} \cos(\alpha)^2 \,\mathrm{d}\theta = \left(\int_{S^1} \sin(\alpha)\cos(\alpha)\,\mathrm{d}\theta\right)^2,$$

which is true iff α is constant, i.e. c has zero curvature κ. Since $c \in C^1(S^1,\mathbb{R}^d)$ and is periodic we can never have $\kappa \equiv 0$ (c cannot be C^1, a line and closed). Hence we can solve for (a,b), so dL_c is surjective. By the bounded inverse

theorem dL_c is a homeomorphism, and we get the existence of a smooth chart $g_c : U \subset E \to F$. Chart changes between two charts centered at c_1 and c_2 are simply compositions of g_{c_1} and $g_{c_2}^{-1}$, so they are smooth. The tangent space is given by $\ker dL_c$, which is equivalent to .

For $\mathcal{A}_\ell$ we consider the map $L \circ \pi$ with the projcetion $\pi : H^{n-1}(S^1) \to H^{n-1}(S^1)/\mathbb{R}$. Boundedness of $d(L \circ \pi)$ is clear, and surjectivity is the same. Only the kernel changes, as is reflected in the statement. □

2.5.2 Sobolev metrics

The metrics we will consider are again Sobolev type metrics. For constant speed curves we will consider metrics of the form

$$\begin{aligned} G_c(h,k) &= \int_{S^1} \sum_{j=0}^{2n} a_j(\ell_c) \langle D_s^j h(s), D_s^j k(s) \rangle \, \mathrm{d}s \\ &= \langle \Lambda h, k \rangle_{H^{-n}, H^n}. \end{aligned} \tag{2.14}$$

Where $a_j \in C^\infty(\mathbb{R}_{>0}, \mathbb{R}_{\geq 0})$ and $a_0, a_n > 0$. $\Lambda : H^p(S^1) \to H^{p-2n}(S_1)$ is an self-adjoint elliptic operator or order $2n$ with coefficients depending on ℓ_c, and $\langle \cdot, \cdot \rangle_{H^{-n}, H^n}$ is the dual pairing. For constant speed curves $D_s = \frac{1}{|c'|}\partial_\theta$ is just a constant multiple of ∂_θ. For each c this defines a metric, which by the conditions of a_0, a_n gives a norm equivalent to the H^n-norm. By Thm. A.4 surjectivity of Λ is equivalent to injectivity, which is guaranteed by the positive definiteness of the metric. Self-adjoint means there are no odd order terms in Λ, which is of the form

$$\Lambda(h) = \sum_{j=0}^{n} a_j(\ell_c) D_s^{2j} h$$

Note that these metrics are exactly the restriction of length-weighted Sobolev metrics (2.10) on $\mathrm{Imm}^n(S^1, \mathbb{R}^d)$. On $\mathcal{A}_1$ it is simply the restriction of a constant coefficient Sobolev metric. As the metrics on Sobolev immersions are strong, and $\mathcal{A}_1$ and $\mathcal{A}_\ell$ are smooth Hilbert submanifolds, the induced metrics are also strong, see [46].

2.5.3 The normal space

To do any kind of analysis of the Riemannian submanifolds, we need to compute the projection onto the tangent and normal spaces. We know what the tangent

spaces are, so we start of by finding the normal space. Here we need to use negative power Sobolev space, see Appendix A, to get a good characterization.

Proposition 2.22 *Let $\mathcal{A}_1$ be equipped with the metric* (2.14)*, the normal space at c is given by*

$$N_c\mathcal{A}_1 = \{\Lambda^{-1}\partial_s(fc') : f \in H^{-n-1}(S^1)\}.$$

For $\mathcal{A}_\ell$ the normal space is given by

$$N_c\mathcal{A}_\ell = \{\Lambda^{-1}\partial_s(fc') : f \in H^{-n+1}(S^1), \langle f, t\rangle_{H^{-n+1},H^{n-1}} = 0, t \in \mathbb{R}\}.$$

Remark 2 *Notice that the elements of $N_c\mathcal{A}_\ell$ are indeed in H^n, since $h \in H^{-n+1}$ and $c' \in H^{n-1}$ then $hc' \in H^{-n+1}$ by Lemma A.2, we differentiate and lose one order and Λ^{-1} raises the order by $2n$ to n. The property that f vanishes as a distribution on constant functions, is in terms of the canonical embedding of H^{n-1} into H^{-n+1}, that f has zero mean.*

Proof. We start of by showing that elements of the given form are orthogonal to tangent vectors. Now let $u = \Lambda^{-1}\partial_s(fc')$, and $h \in T_c\mathcal{A}_1$

$$\begin{aligned}\langle \Lambda u, h\rangle_{H^{-n},H^n} &= \langle \partial_s(fc'), h\rangle_{H^{-n},H^n}\\ &= -\langle fc', h'\rangle_{H^{-n+1},H^{n-1}}\\ &= -\langle f, \langle c', h'\rangle\rangle_{H^{-n+1},H^{n-1}}\\ &= 0.\end{aligned}$$

For $\mathcal{A}_\ell$ the last equality follows since $\langle c', h'\rangle$ is constant. Now let

$$\langle \Lambda u, h\rangle_{H^{-n},H^n} = 0, \quad \forall h \in T_c\mathcal{A}_1.$$

Since Λ is an isomorphism of H^n and H^{-n}, we can put $u = \Lambda^{-1}w$ for some $w \in H^{-n}$. Constant vector fields h_0 are in the tangent space so $\langle w, h_0\rangle_{H^{-n},H^n} = 0$, which implies that the zero order coefficient of the Fourier expansion of w is zero, and hence there exists $p \in H^{n-1}$ s.t. $p' = w$. Any element of H^{n-1} can be written uniquely in terms of the unit tangent v and normal n, which induces the same splitting in H^{-n+1}, so $p = fv + gn$ for some $f, g \in H^{-n+1}$.

$$\begin{aligned}\langle \partial_s p, h\rangle_{H^{-n},H^n} &= -\langle fv + gn, h'\rangle_{H^{-n+1},H^{n-1}}\\ &= -\langle f, \langle v, h'\rangle\rangle_{H^{-n+1},H^{n-1}} - \langle g, \langle n, h'\rangle\rangle_{H^{-n+1},H^{n-1}}\\ &= -\langle g, \langle n, h'\rangle\rangle_{H^{-n+1},H^{n-1}}\\ &= 0\end{aligned}$$

Which implies $g = 0$ since $\langle n, h'\rangle$ vary over all functions in H^{n-1}. Hence we have shown that $u = \Lambda^{-1}\partial_s(fc')$. For $\mathcal{A}_\ell$ we follow the same steps and instead end up with

$$\langle f, t\rangle_{H^{-n+1},H^{n-1}} + \langle g, \langle n, v'\rangle\rangle_{H^{-n+1},H^{n-1}} = 0.$$

For all constants t. Picking $t = 0$ gives $g = 0$ again, and so we get the same expression $u = \Lambda^{-1}\partial_s(fc')$ and h satisfies $\langle f, t\rangle_{H^{-n+1},H^{n-1}} = 0$. □

2.5.4 The geodesic equation

On our step to derive an expression for the geodesic equation, we compute the projection onto the normal and tangent space. From normal Hilbert space theory, the projection operators are well-defined since the metric at each curve c gives a constant coefficient Sobolev inner product. Hence we have the splitting of a fixed u according to

$$u = u^{\perp} + u^{\top} = \Lambda^{-1}\partial_s(fc') + w \tag{2.15}$$

We need to show how to find f and w. We will need the following operators

$$M(h) = \langle c', \partial_s\Lambda^{-1}\partial_s(hc')\rangle$$
$$N(h) = \langle n, \partial_s\Lambda^{-1}\partial_s(hc')\rangle$$

Where k and h are unknown. By the same argument as in Remark 2, the left hand side is well defined operator from $M : H^{n-1} \to H^n$. Note that M is not a differential operator, and has Sobolev coefficients (from the presence of c'), so standard elliptic theory is not directly applicable. The following Lemma shows that K is invertible.

Lemma 2.23 *For $n \geq 2$ the operator $M : H^{-n+1} \to H^{n-1}$ defined by*

$$M(f) = \langle c', \partial_s\Lambda^{-1}\partial_s(fc')\rangle$$

is a bounded and invertible operator.

Proof. Boundedness is already established by using the estimates from Lemma A.2. Define the L^2 adjoint of dL_c

$$\langle dL_c(h), k\rangle_0 = \langle\langle h', c'\rangle, k\rangle_0 = \langle h, (kc')'\rangle_{0,d} = \langle h, dL_c^*(k)\rangle_{0,d}.$$

Where $\langle\cdot,\cdot\rangle_{0,d}$ is the L^2 inner product for $\mathbb{R}^d$ valued functions. Then $M = dL_c \circ \Lambda^{-1} \circ dL_c^*$, and $T_c\mathcal{A}_1 = \ker dL_c$ (by definition). From this it is clear that M is L^2-selfadjoint, but notice that this does not imply the closedness of M immediately since we have a different domain. Proposition 2.22, implies that we have the g_c-orthogonal direct sum $H^n = \ker dL_c \oplus \operatorname{im}(L^{-1} \circ dL_c^*)$, especially

$\ker dL_c \cap \operatorname{im}(L^{-1} \circ dL_c^*) = \{0\}$. L^{-1} is an isomorphism, so $\ker M = \ker dL_c^*$. C^∞ is dense in all H^n so let $h \in \ker dL_c^*$ be smooth, then $(hc')' = h'c' + hc'' = 0$. If the curvature κ is not identically zero then c' and c'' are linearly independent so $h = 0$, by density and boundedness this extends to any $h \in H^{-n+1}$. Since $c \in C^1(S^1, \mathbb{R}^d)$ and is periodic we can never have $\kappa \equiv 0$ (c cannot be C^1, a line and closed). So $\ker M = \ker dL_c^* = 0$. We have already showed in the proof of Theorem 2.21 that dL_c is surjective. Applying dL_c to the direct sum shows that $\operatorname{im} dL_c = \operatorname{im} M$, so M is surjective. □

Proposition 2.24 *Let $\pi_1^\top : H^n \to T_c\mathcal{A}_1$ and $\pi_\ell^\top : H^n \to T_c\mathcal{A}_\ell$ denote the orthogonal projections onto the two tangent spaces respectively, then they are determined by the follow relations,*

$$(\pi_1^\top(u))' = (NM^{-1}(\langle u', c'\rangle) - \langle u', n\rangle)n,$$

and

$$(\pi_\ell^\top(u))' = (N(M^{-1}(\langle u, c'\rangle - t)) - \langle u', n\rangle)n + tv, \quad t = \frac{\int_{S^1} K^{-1}(\langle u', c'\rangle)\,\mathrm{d}\theta}{\int_{S^1} K^{-1}(1)\,\mathrm{d}\theta}.$$

Proof. We will consider $\mathcal{A}_\ell$ first. Let $u \in T_c\mathcal{A}_\ell$, which splits as in (2.15). Then $\langle w', c'\rangle = t \in \mathbb{R}$, for some t. After differentiating (2.15) and dotting with c',

$$\langle u', c'\rangle = M(f) + t$$

By Lemma 2.23 we can invert K, and obtain

$$f = M^{-1}(\langle u', c'\rangle - t),$$

where t is unknown for know. To determine t, we use that f has to have zero mean, $\int_{S^1} f\,\mathrm{d}x = 0$.

$$t = \frac{\int_{S^1} M^{-1}(\langle u', c'\rangle)\,\mathrm{d}\theta}{\int_{S^1} M^{-1}(1)\,\mathrm{d}\theta}.$$

From which the statement follows. For $\mathcal{A}_1$, $t = 0$. □ We know that both

A_1 and A_ℓ are smooth strong Hilbert submanifolds of $(\mathrm{Imm}^n(S^1, \mathbb{R}^d), G_c^{n,\ell})$, so the geodesic equations exist and are locally well-posed. Writing up the geodesic equation is another question. The Levi-Cevita connection is given by the usual formula for submanifolds: $\nabla_h^{\mathcal{A}} k = (\nabla_h^{\mathrm{Imm}} k)^\top$, where ∇^{Imm} is the Levi-Cevita connection on $(\mathrm{Imm}(S^1, \mathbb{R}^d)$ w.r.t the metric $G_c^{n,\ell}$, and h and k are vectorfields extended from the submanifold to the ambient space.

We start with the geodesic equation on $\mathcal{A}_1$. Since $\ell = 1$, all coefficients in (2.14) are constant, and covariant derivatives are standard derivates, so all metrics

G_c are induced by a flat metric on the ambient space, hence $\nabla_{c_t} c_t = c_{tt}$. Let $c : I \subset \mathbb{R} \to \mathcal{A}_1 \subset \mathrm{Imm}^n(S^1, \mathbb{R}^d)$ be a unit-speed path. By differentiating the defining equation, we get the following set of identities valid along the path,

$$|c'|^2 = 1, \quad \langle c'_t, c' \rangle = 0, \quad , |c'_t|^2 + \langle c'_{tt}, c' \rangle = 0.$$

Using the last equality and Prop. 2.24 we can write the geodesic equation $\nabla^A_{c_t} c_t = 0$ as

$$\langle c'_{tt}, n \rangle = -NM^{-1}(|c'_t|). \tag{2.16}$$

For the geodesic equation on $\mathcal{A}_\ell$, one can follow the same procedure. However, on $\mathcal{A}_\ell$ the metric is no longer induced by a flat metric on the ambient space, so we need to compute the Levi-Cevita connection of the length-weighted metric on $\mathrm{Imm}(S^1, \mathbb{R}^d)$, which is quite involved. If we assume for simplicity that the coefficients func $a_j(\ell)$ are constant, the Levi-Cevita connection on $\mathrm{Imm}(S^1, \mathbb{R}^d)$ can be extracted from the geodesic equation (2.8), as this is $\nabla_{c_t} c_t = 0$. Using that $\langle D_s c_t, v \rangle$ is constant, one can simplify the equation a bit. We just state the final expression

$$\Lambda(c_{tt}) - \sum_{k=0}^{n} a_k (1-2k) |c'| \left(D_s(\langle D_s^k c_t, D_s^k c_t \rangle v) - \langle D_s c_t, v \rangle D_s^{2k} c_t \right) = 0.$$

Now we would need to apply the orthogonal projection onto $T_c \mathcal{A}_\ell$ to the left hand side. We wont show the final result. The case that $a_j(\ell)$ is not constant is even less desirable to write out in detail.

On the other hand it is easy show by using the results of Section 2.4 that we gain local and global well-posedness of the geodesic equations.

Proposition 2.25 *The geodesic equation on $\mathcal{A}_1$ is locally and globally well-posed. The geodesic equation on $\mathcal{A}_\ell$ is locally well-posed and globally well-posed if the coefficients satisfy I_0 and I_∞.*

Proof. In both cases the induced metric is strong, so local well-posedness follows from say [46, Prop. VII.5.1]. $\mathcal{A}_1$ is a closed subset of Sobolev space with the induced metric, so is metrically complete, hence geodesically complete. $\mathcal{A}_\ell$ is not closed in $H^n(S^1, \mathbb{R}^d)$ but it is closed in $\mathrm{Imm}^n(S^1 \mathbb{R}^d)$. We have already shown that the geodesic distance on $\mathrm{Imm}^n(S^1, \mathbb{R}^d$ induces the same topology as the manifold topology, and is a complete metric space. Hence $\mathcal{A}_\ell$ is metrically complete and geodesically complete. □

The space of constant speed curves is an alternative shape space to quotient out immersions by diffeomorphisms, and one could ask if this is better for application purposes. In anticipation of the results of Chap. 3, we can say that this not the

case. It is more cumbersome in practice to represent curves that are constant speed. In Riemannian shape analysis we also want the geodesic deformations to ideally model some "natural" small deformation between curves, but the rigidity in using a fixed parametrization can make even similar curves have very large minimal deformations between them.

Chapter 3

Paper: A Numerical Framework for Geodesics

In this chapter we present the contents of the papers [8], [6] and [7]. Sec. 3.1-3.4 is a verbatim copy of [8]. Section 3.5 is an addendum which contain the applications from the papers [6] and [7]. This is joint work with Martin Bauer, Martins Bruveris and Philipp Harms.

3.1 Introduction

The comparison and analysis of geometric shapes plays a central role in many applications. A particularly important class of shapes is the space of curves, which is used to model applied problems in medical imaging [84, 86], object tracking [80, 81], computer animation [13, 31], speech recognition [79], biology [45, 78], and many other fields [11, 44].

In this article we consider the space $\mathrm{Imm}(S^1, \mathbb{R}^d)$ of closed, regular (or immersed) curves in $\mathbb{R}^d$ as well as some quotients of this space by reparametrizations and Euclidean motions. These spaces of shapes are inherently nonlinear. To make standard methods of statistical analysis applicable, one can linearize the space locally around each shape. This can be achieved by introducing a

Riemannian structure, which describes both the global nonlinearity of the space as well as its local linearity. Over the past decade Riemannian shape analysis has become an active area of research in pure and applied mathematics. Driven by applications, a variety of different Riemannian metrics has been used.

An important class of metrics are Sobolev metrics. These metrics can be defined initially on the space $\mathrm{Imm}(S^1, \mathbb{R}^d)$ and then induced on quotients of this space by requiring the projections to be Riemannian submersions (see Def. 3.1 and Thm. 3.6). Recently Sobolev metrics of order two were shown to possess much nicer properties than metrics of lower order: the geodesic distance is non-degenerate, the geodesic equation is globally well-posed, any two curves in the same connected component can be connected by a minimizing geodesic, the metric completion consists of all H^2-immersions, and the metric extends to a strong Riemannian metric on the metric completion [23, 26].

Numerical methods for the statistical analysis of shapes under second order metrics are, however, still largely missing. This is in contrast to first order metrics, where isometries to simpler spaces led to explicit formulas for geodesics under many parameter configurations of the metric [9, 40, 77, 87]. For certain H^2-metrics an analogous approach was developed in [12]. Moreover, the geodesic boundary value problem under second order Finsler metrics on the space of BV^2-curves was implemented numerically in [54]. For general second order Sobolev metrics on spaces of unparametrized curves a numerical framework is, however, still lacking. This is the topic of this paper.

We present a numerical implementation of the initial and boundary value problems for geodesics under second order Sobolev metrics.[1] Our implementation is based on a discretization of the Riemannian energy functional using B-splines. The boundary value problem for geodesics is solved by a standard minimization procedure on the set of discretized paths and the initial value problem by discrete geodesic calculus [67]. Our approach is general in that it allows to factor out reparametrizations and rigid transformations. Moreover, it involves no restriction on the parameters of the metric and could be applied to other, higher-order metrics, as well.

In future work our framework could be applied to other spaces of mappings like manifold-valued curves, embedded surfaces, or more general spaces of immersions (see [4, 15] for details and [11] for a general overview).

[1] Our code can be downloaded from `https://github.com/h2metrics/h2metrics.git`.

3.2 Sobolev metrics on spaces of curves

3.2.1 Notation

The space of smooth, regular curves with values in $\mathbb{R}^d$ is

$$\operatorname{Imm}(S^1,\mathbb{R}^d) = \left\{c \in C^\infty(S^1,\mathbb{R}^d) \colon \forall \theta \in S^1, c'(\theta) \neq 0\right\}, \tag{3.1}$$

where Imm stands for *immersions*. We call such curves parametrized to distinguish them from unparametrized curves defined in Sect. 3.2.3. The space $\operatorname{Imm}(S^1,\mathbb{R}^d)$ is an open subset of the Fréchet space $C^\infty(S^1,\mathbb{R}^d)$ and therefore can be considered as a Fréchet manifold. Its tangent space $T_c\operatorname{Imm}(S^1,\mathbb{R}^d)$ at any curve c is the vector space $C^\infty(S^1,\mathbb{R}^d)$ itself. We denote the Euclidean inner product on $\mathbb{R}^d$ by $\langle\cdot,\cdot\rangle$. Differentiation is sometimes denoted using subscripts as in $c_\theta = \partial_\theta c = c'$. Moreover, for any fixed curve c, we denote differentiation and integration with respect to arc length by $D_s = \partial_\theta/|c_\theta|$ and $\mathrm{d}s = |c_\theta|\,\mathrm{d}\theta$, respectively. A path of curves is a mapping $c\colon[0,1] \to \operatorname{Imm}(S^1,\mathbb{R}^d)$; its velocity is denoted by $c_t = \partial_t c = \dot{c}$.

3.2.2 Parametrized curves

In this article we study the following class of weak Riemannian metrics on $\operatorname{Imm}(S^1,\mathbb{R}^d)$.

Definition 3.1 A *second order Sobolev metric with constant coefficients* on $\operatorname{Imm}(S^1,\mathbb{R}^d)$ is a weak Riemannian metric of the form

$$G_c(h,k) = \int_{S^1} a_0\langle h,k\rangle + a_1\langle D_s h, D_s k\rangle + a_2\langle D_s^2 h, D_s^2 k\rangle\,\mathrm{d}s\,, \tag{3.2}$$

where $h,k \in T_c\operatorname{Imm}(S^1,\mathbb{R}^d)$, and $a_j \in \mathbb{R}$ are constants with $a_0, a_2 > 0$ and $a_1 \geq 0$. If $a_2 = 0$ and $a_1 > 0$ it is a first order metric and if $a_1 = a_2 = 0$ it is a zero order or L^2-metric.

Note that the symbols D_s and $\mathrm{d}s$ hide the dependency of the Riemannian metric on the base point c. Expressing derivatives in terms of θ instead of arc length, one has

$$\begin{aligned} G_c(h,k) = \int_0^{2\pi} & a_0|c'|\langle h,k\rangle + \frac{a_1}{|c'|}\langle h',k'\rangle + \frac{a_2}{|c'|^7}\langle c',c''\rangle^2\langle h',k'\rangle \\ & - \frac{a_2}{|c'|^5}\langle c',c''\rangle\big(\langle h',k''\rangle + \langle h'',k'\rangle\big) + \frac{a_2}{|c'|^3}\langle h'',k''\rangle\,\mathrm{d}\theta\,. \end{aligned} \tag{3.3}$$

In the Riemannian setting the length of a path $c\colon [0,1] \to \operatorname{Imm}(S^1, \mathbb{R}^d)$ is defined as

$$L(c) = \int_0^1 \sqrt{G_{c(t)}(c_t(t), c_t(t))}\, \mathrm{d}t\,, \tag{3.4}$$

and the geodesic distance between two curves $c_0, c_1 \in \operatorname{Imm}(S^1, \mathbb{R}^d)$ is the infimum of the lengths of all paths connecting these curves, i.e.,

$$\operatorname{dist}(c_0, c_1) = \inf_c \{L(c)\colon c(0) = c_0,\, c(1) = c_1\}\,.$$

On finite-dimensional manifolds the topology induced by the geodesic distance coincides with the manifold topology by the Hopf–Rinow theorem. On infinite-dimensional manifolds with weak Riemannian metrics this is not true anymore. For example, the geodesic distance induced by the L^2-metric on curves vanishes identically [5, 51]. On the other hand, first and second order metrics overcome this degeneracy, as the following result of [51, 52] shows.

Theorem 3.2 *The geodesic distance of first and second order metrics on* $\operatorname{Imm}(S^1, \mathbb{R}^d)$ *separates points, i.e.,* $\operatorname{dist}(c_0, c_1) > 0$ *holds for all* $c_0 \neq c_1$.

Geodesics are locally distance-minimizing paths. They can be described by a partial differential equation, called the geodesic equation. It is the first order condition for minima of the energy functional

$$E(c) = \frac{1}{2} \int_0^1 G_{c(t)}(c_t(t), c_t(t))\, \mathrm{d}t\,. \tag{3.5}$$

Recently some local and global existence results for geodesics of Sobolev metrics were shown in [26, 23, 52]. We summarize them here since they provide the theoretical underpinnings for the numerical methods presented in this paper.

Theorem 3.3 *The geodesic equation of second order metrics, written in terms of the momentum* $p = |c'|(a_0 c_t - a_1 D_s^2 c_t + a_2 D_s^4 c_t)$, *is given by*

$$\begin{aligned} \partial_t p = &-\frac{a_0}{2} |c_\theta| D_s(\langle c_t, c_t\rangle D_s c) + \frac{a_1}{2} |c_\theta| D_s(\langle D_s c_t, D_s c_t\rangle D_s c) \\ &- \frac{a_2}{2} |c_\theta| D_s(\langle D_s^3 c_t, D_s c_t\rangle D_s c) + \frac{a_2}{2} |c_\theta| D_s(\langle D_s^2 c_t, D_s^2 c_t\rangle D_s c)\,. \end{aligned} \tag{3.6}$$

For any initial condition $(c_0, u_0) \in T \operatorname{Imm}(S^1, \mathbb{R}^d)$ *the geodesic equation has a unique solution, which exists for all time. In contrast, the geodesic equation of first order Sobolev metrics is locally, but not globally, well-posed.*

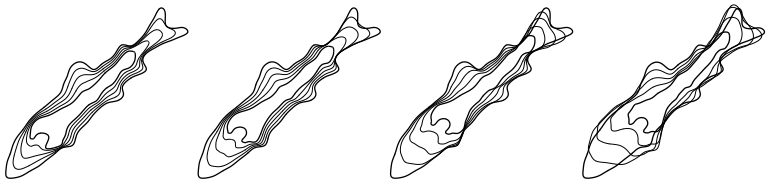


Figure 3.1: Influence of the constants in the metric on geodesics between a fish and a tool in the space of unparametrized curves. The metric parameter a_1 is set to zero, whereas the parameter a_2 is increased by a factor 10 in the second, a factor 100 in the third, and a factor 1000 in the fourth column. The corresponding geodesic distances are 135.65, 162.35, 229.26 and 451.9. Note that since we also optimize over translations and rotations of the target curve, the position in space varies.

Remark 2 The choice of parameters a_0, a_1, and a_2 of the Riemannian metric can have a large influence on the resulting optimal deformations. We illustrate this in Fig. 3.1, where we show the geodesic between a fish-like and a tool-like curve for various choices of parameters.

For second order metrics it is possible to compute the metric completion of the space of smooth immersions. We introduce the Banach manifold of Sobolev immersions

$$\mathcal{I}^2(S^1,\mathbb{R}^d) = \left\{c \in H^2(S^1,\mathbb{R}^d) \colon \forall \theta \in S^1, c'(\theta) \neq 0\right\} . \tag{3.7}$$

By the Sobolev embedding theorem this space is well-defined and an open subset of the space of all C^1-immersions. It has been shown in [14, 23] that $\mathcal{I}^2(S^1,\mathbb{R}^d)$ coincides with the metric completion of the space of smooth immersions:

Theorem 3.4 *The metric completion of the space* $\operatorname{Imm}(S^1,\mathbb{R}^d)$ *endowed with a second order Sobolev metric is* $\mathcal{I}^2(S^1,\mathbb{R}^d)$*. Furthermore, any two two curves* c_0 *and* c_1 *in the same connected component of* $\mathcal{I}^2(S^1,\mathbb{R}^d)$ *can be joined by a minimizing geodesic.*

3.2.3 Unparametrized curves

In many applications curves are considered equal if they differ only by their parametrization, i.e., we identify the curves c and $c \circ \varphi$, where $\varphi \in \operatorname{Diff}(S^1)$ is a reparametrization. The reparametrization group $\operatorname{Diff}(S^1)$ is the diffeomorphism

group of the circle,

$$\mathrm{Diff}(S^1) = \left\{\varphi \in C^\infty(S^1, S^1) : \varphi' > 0\right\},$$

which is an infinite-dimensional regular Fréchet Lie group [43]. Reparametrizations act on curves by composition from the right, i.e., $c \circ \varphi$ is a reparametrization of c. The space

$$B_i(S^1, \mathbb{R}^d) = \mathrm{Imm}(S^1, \mathbb{R}^d) / \mathrm{Diff}(S^1),$$

of unparametrized curves is the orbit space of this group action. This space is not a manifold; it has singularities at any curve c with nontrivial isotropy subgroup [27]. We therefore restrict ourselves to the dense open subset $\mathrm{Imm}_f(S^1, \mathbb{R}^d)$ of curves upon which $\mathrm{Diff}(S^1)$ acts freely and define

$$B_{i,f}(S^1, \mathbb{R}^d) = \mathrm{Imm}_f(S^1, \mathbb{R}^d) / \mathrm{Diff}(S^1).$$

This restriction, albeit important for theoretical reasons, has no influence on the practical applications of Sobolev metrics, since $B_{i,f}(S^1, \mathbb{R}^d)$ is open and dense in $B_i(S^1, \mathbb{R}^d)$. We have the following result concerning the manifold structure of the orbit space and the descending properties of Sobolev metrics [15, 27, 52].

Theorem 3.5 *The space $B_{i,f}(S^1, \mathbb{R}^d)$ is a Fréchet manifold and the base space of the principal fibre bundle*

$$\pi : \mathrm{Imm}_f(S^1, \mathbb{R}^d) \to B_{i,f}(S^1, \mathbb{R}^d), \quad c \mapsto c \circ \mathrm{Diff}(S^1),$$

with structure group $\mathrm{Diff}(S^1)$. A Sobolev metric G on $\mathrm{Imm}_f(S^1, \mathbb{R}^d)$ induces a metric on $B_{i,f}(S^1, \mathbb{R}^d)$ such that the projection π is a Riemannian submersion.

The induced Riemannian metric on $B_{i,f}(S^1, \mathbb{R}^d)$ defines a geodesic distance, which can also be calculated using paths in $\mathrm{Imm}_f(S^1, \mathbb{R}^d)$ connecting c_0 to the orbit $c_1 \circ \mathrm{Diff}(S^1)$, i.e., for $\pi(c_0), \pi(c_1) \in B_{i,f}(S^1, \mathbb{R}^d)$ we have,

$$\mathrm{dist}\left(\pi(c_0), \pi(c_1)\right) = \inf\left\{L(c) : c(0) = c_0,\ c(1) \in c_1 \circ \mathrm{Diff}(S^1)\right\}.$$

To relate the geometries of $\mathrm{Imm}(S^1, \mathbb{R}^d)$ and $B_{i,f}(S^1, \mathbb{R}^d)$, one defines the vertical and horizontal subspaces of $T_c \mathrm{Imm}_f(S^1, \mathbb{R}^d)$,

$$\mathrm{Ver}_c = \ker(T_c\pi), \quad \mathrm{Hor}_c = (\mathrm{Ver}_c)^{\perp, G_c}.$$

As shown in [52] they form a decomposition of $T_c \mathrm{Imm}_f(S^1, \mathbb{R}^d)$,

$$T_c \mathrm{Imm}_f(S^1, \mathbb{R}^d) = \mathrm{Ver}_c \oplus \mathrm{Hor}_c,$$

as a direct sum. More explicitly,

$$\begin{aligned}
\mathrm{Ver}_c &= \left\{g.v_c \in T_c \mathrm{Imm}_f(S^1, \mathbb{R}^d) : g \in C^\infty(S^1)\right\} \\
\mathrm{Hor}_c &= \left\{k \in T_c \mathrm{Imm}_f(S^1, \mathbb{R}^d) : \langle a_0 k - a_1 D_s^2 k + a_2 D_s^4 k, v_c\rangle = 0\right\},
\end{aligned}$$

with $v_c = D_s c$ the unit tangent vector to c.

A geodesic c on $\mathrm{Imm}_f(S^1, \mathbb{R}^d)$ is called *horizontal* at t, if $\partial_t c(t) \in \mathrm{Hor}_{c(t)}$. It can be shown that if c is horizontal at $t = 0$, then it is horizontal at all t. Furthermore, geodesics on $B_{i,f}(S^1, \mathbb{R}^d)$ can be lifted to horizontal geodesics on $\mathrm{Imm}_f(S^1, \mathbb{R}^d)$ and the lift is unique if we specify the initial position of the lift; conversely, horizontal geodesics on $\mathrm{Imm}(S^1, \mathbb{R}^d)$ project down to geodesics on $B_{i,f}(S^1, \mathbb{R}^d)$.

What about long-time existence of geodesics on $B_{i,f}(S^1, \mathbb{R}^d)$? Using the correspondence between geodesics on $B_{i,f}(S^1, \mathbb{R}^d)$ and horizontal geodesics on $\mathrm{Imm}_f(S^1, \mathbb{R}^d)$ together with Thm. 3.3 we see that the horizontal lift of a geodesic can be extended for all times. However, it can leave the subset of free immersions and pass through curves with a non-trivial isotropy group. Thus the space $B_{i,f}(S^1, \mathbb{R}^d)$ is not geodesically complete, but we can regain geodesic completeness if we allow geodesics to pass through $B_i(S^1, \mathbb{R}^d)$.

The space $B_i(S^1, \mathbb{R}^d)$ inherits some of the completeness properties of $\mathrm{Imm}(S^1, \mathbb{R}^d)$. To formulate these properties we introduce the group $\mathcal{D}^2(S^1)$ of H^2-diffeomorphisms and the corresponding shape space of Sobolev immersions,

$$\mathcal{B}^2(S^1, \mathbb{R}^d) = \mathcal{I}^2(S^1, \mathbb{R}^d)/\mathcal{D}^2(S^1)\,.$$

It is not known whether this space is a smooth Banach manifold, it is however a metric length space. The structure of it is explained in more detail in the article [23], where the following completeness result is proven.

Theorem 3.6 *Let G be a second order Sobolev metric with constant coefficients.*

1. *The space $\big(\mathcal{B}^2(S^1, \mathbb{R}^d), \mathrm{dist}\big)$, where* dist *is the quotient distance induced by $\big(\mathcal{I}^2(S^1, \mathbb{R}^d), \mathrm{dist}\big)$, is a complete metric space, and it is the metric completion of $\big(B_{i,f}(S^1, \mathbb{R}^d), \mathrm{dist}\big)$.*
2. *Given two unparametrized curves $C_1, C_2 \in \mathcal{B}^2(S^1, \mathbb{R}^d)$ in the same connected component, there exist $c_1, c_2 \in \mathcal{I}^2(S^1, \mathbb{R}^d)$ with $c_1 \in \pi^{-1}(C_1)$ and $c_2 \in \pi^{-1}(C_2)$, such that*

$$\mathrm{dist}(C_1, C_2) = \mathrm{dist}(c_1, c_2)\,;$$

equivalently the infimum in

$$\mathrm{dist}(\pi(c_1), \pi(c_2)) = \inf_{\varphi \in \mathcal{D}^2(S^1)} \mathrm{dist}(c_1, c_2 \circ \varphi)$$

is attained.

3. *The metric space* $\left(\mathcal{B}^2(S^1,\mathbb{R}^d),\mathrm{dist}\right)$ *is a length space and any two shapes in the same connected component can be joined by a minimizing geodesic.*

In the last statement of the above theorem we have to understand a minimizing geodesic in the sense of metric spaces.

3.2.4 Euclidean motions

Curves modulo Euclidean motions are a natural object of consideration in many applications. The Euclidean motion group $SE(d) = SO(d)\ltimes\mathbb{R}^d$ is the semi-direct product of the translation group $\mathbb{R}^d$ and the rotation group $SO(d)$. These groups act on $\mathrm{Imm}(S^1,\mathbb{R}^d)$ by composition from the left. The metric (3.3) is invariant under these group actions,

$$G_{R.c+a}(R.h,R.k) = G_c(h,k) \qquad \forall (R,a)\in SE(d)\,.$$

As in the previous section we obtain an induced Riemannian metric on the quotient space

$$\mathcal{S}(S^1,\mathbb{R}^d) = \mathrm{Imm}_f(S^1,\mathbb{R}^d)/\operatorname{Diff}(S^1)\times SE(d) = B_{i,f}(S^1,\mathbb{R}^d)/SE(d)\,,$$

such that the projection $\pi : \mathrm{Imm}_f(S^1,\mathbb{R}^d)\to\mathcal{S}(S^1,\mathbb{R}^d)$ is a Riemannian submersion:

Theorem 3.7 *The space* $\mathcal{S}(S^1,\mathbb{R}^d)$ *is a Fréchet manifold and the base space of the principal fibre bundle*

$$\pi : \mathrm{Imm}_f(S^1,\mathbb{R}^d)\to\mathcal{S}(S^1,\mathbb{R}^d)\,,\quad c\mapsto SE(d).c\circ\operatorname{Diff}(S^1)\,,$$

with structure group $\operatorname{Diff}(S^1)\times SE(d)$*. A Sobolev metric* G *on* $\mathrm{Imm}_f(S^1,\mathbb{R}^d)$ *induces a metric on* $\mathcal{S}(S^1,\mathbb{R}^d)$ *such that the projection* π *is a Riemannian submersion.*

Note that the left action of $SE(d)$ commutes with the right action of $\operatorname{Diff}(S^1)$ and hence the order of the quotient operations does not matter. The induced geodesic distance on the quotient space is given by the infimum

$$\mathrm{dist}\left(\pi(c_0),\pi(c_1)\right) = \inf\left\{L(c)\colon c(0)=c_0,\, c(1)\in\pi(c_1) = SE(d).c_1\circ\operatorname{Diff}(S^1)\right\}\,,$$

with the infimum being taken over paths $c:[0,1]\to\mathrm{Imm}(S^1,\mathbb{R}^d)$.

Similarily as in the previous section geodesics on $\mathcal{S}(S^1,\mathbb{R}^d)$ can be lifted to horizontal geodesics on $\mathrm{Imm}_f(S^1,\mathbb{R}^d)$ and, conversely, horizontal geodesics on

$\mathrm{Imm}(S^1, \mathbb{R}^d)$ project down to geodesics on $\mathcal{S}(S^1, \mathbb{R}^d)$. Thus the space $\mathcal{S}(S^1, \mathbb{R}^d)$ inherits again some of the completeness properties of $\mathrm{Imm}(S^1, \mathbb{R}^d)$ and we obtain the equivalent of Thm. 3.6 also for the space $\mathcal{B}^2(S^1, \mathbb{R}^d)/SE(d)$.

Remark 3 The Sobolev metric (3.3) is not invariant with respect to scalings. However, this lack of invariance can be addressed by introducing weights depending on the length ℓ_c of the curve c. The modified metric

$$\widetilde{G}_c(h,k) = \int_{S^1} \frac{a_0}{\ell_c^3}\langle h,k\rangle + \frac{a_1}{\ell_c}\langle D_s h, D_s k\rangle + a_2\ell_c\langle D_s^2 h, D_s^2 k\rangle \,\mathrm{d}s$$

is invariant with respect to scalings. It induces a metric on the quotient space $\mathcal{S}(S^1, \mathbb{R}^d)/\mathbb{R}_+$, where $\mathbb{R}_+$ is the scaling group acting by multiplication $(\lambda, c) \mapsto \lambda.c$ on curves.

3.3 Discretization

In order to numerically compute geodesics, the infinite-dimensional space of curves must be discretized. The method we choose is standard: we construct an appropriate finite-dimensional function space and perform optimization therein. We choose B-splines among the many possible options because B-splines and their derivatives have piecewise polynomial representations and can be evaluated efficiently. This permits fast and simple computation of the energy functional and its derivatives. Furthermore, in contrast to standard finite-element discretization, it is possible to control the global regularity of the functions. For details regarding B-splines, their definition, efficient computations, etc., we refer to [73] and the vast literature on the subject.

For simplicity, we shall work only with *simple* B-splines, i.e., splines where all interior knots have multiplicity one. Hence the splines have maximal regularity at the knots. We will define splines of degrees n_t and n_θ in the variables $t \in [0,1]$ and $\theta \in [0, 2\pi]$, respectively. The corresponding numbers of control points are denoted by N_t and N_θ. For t we use a uniform knot sequence on the interval $[0,1]$ with full multiplicity at the boundary knots:

$$\Delta_t = \{t_i\}_{i=0}^{2n_t+N_t}, \qquad t_i = \begin{cases} 0 & 0 \le i < n_t \\ \dfrac{i-n_t}{N_t} & n_t \le i < n_t + N_t \\ 1 & n_t + N_t \le i \le 2n_t + N_t\,. \end{cases}$$

For θ we want the splines to be periodic on the interval $[0, 2\pi]$. Therefore we choose knots

$$\Delta_\theta = \{\theta_j\}_{j=0}^{2n_\theta+N_\theta}, \qquad \theta_j = \frac{j-n_\theta}{2\pi N_\theta}, \quad 0 \le j \le 2n_\theta + N_\theta .$$

The corresponding normalized B-spline basis functions are denoted by $B_i(t)$ and $C_j(\theta)$. Note that all interior knots have multiplicity one, i.e., the splines are simple. Therefore, they have maximal regularity at the knots,

$$B_i \in C^{n_t-1}([0,1]), \qquad C_j \in C^{n_\theta-1}(S^1), \qquad i = 1,\dots,N_t, \qquad j = 1,\dots,N_\theta .$$

Let $\mathcal{S}_{N_t}^{n_t}$ denote the orthogonal projection from $H^{n_t}([0,1])$ onto the span of the basis functions B_i. Similarly, let $\mathcal{S}_{N_\theta}^{n_\theta}$ denote the orthogonal projection from $H^{n_\theta}(S^1)$ onto the span of the basis functions C_j. Then

$$\lim_{N_t\to\infty} \|\mathcal{S}_{N_t}^{n_t} f - f\|_{H^{n_t}([0,1])} = 0, \qquad \lim_{N_\theta\to\infty} \|\mathcal{S}_{N_\theta}^{n_\theta} g - g\|_{H^{n_\theta}(S^1)} = 0,$$

holds for each $f \in H^{n_t}([0,1])$ and each $g \in H^{n_\theta}(S^1)$. This is a well-known result on the approximation power of one-dimensional splines (c.f. Lem. B.4); a detailed analysis can be found in [73].

The generalization of this statement to multiple dimensions involves tensor product splines and mixed-order Sobolev spaces. Tensor product splines are linear combinations of $B_i \otimes C_j$, where the basis functions B_i are interpreted as functions of t and C_j as functions of θ. To be explicit, a path of curves is represented as a tensor product B-spline with control points $c_{i,j} \in \mathbb{R}^d$ as follows:

$$c(t,\theta) = \sum_{i=1}^{N_t}\sum_{j=1}^{N_\theta} c_{i,j} B_i(t) C_j(\theta) . \tag{3.8}$$

Sobolev spaces of mixed order are Hilbert spaces defined for each $k, \ell \in \mathbb{N}$ as

$$\begin{aligned} H^{k,\ell}([0,1]\times S^1) &= \left\{ f \in L^2([0,1]\times S^1) : \exists f^{(k,0)}, f^{(0,\ell)}, f^{(k,\ell)} \in L^2([0,1]\times S^1) \right\}, \\ \langle f, g\rangle_{H^{k,\ell}} &= \langle f, g\rangle_{L^2} + \langle f^{(k,0)}, g^{(k,0)}\rangle_{L^2} + \langle f^{(0,\ell)}, g^{(0,\ell)}\rangle_{L^2} + \langle f^{(k,\ell)}, g^{(k,\ell)}\rangle_{L^2} . \end{aligned} \tag{3.9}$$

Function spaces of this type were first defined in [55, 56]. We refer to [83] and [68] for detailed expositions and further references. As before we define for each number of control points N_t, N_θ the spline approximation operator $\mathcal{S}_{N_t,N_\theta}^{n_t,n_\theta}$ to be the orthogonal projection from $H^{n_t,n_\theta}([0,1]\times S^1)$ onto the span of the tensor product splines $B_i \otimes C_j$. It can be shown that $\mathcal{S}_{N_t,N_\theta}^{n_t,n_\theta} = \mathcal{S}_{N_t}^{n_t} \otimes \mathcal{S}_{N_\theta}^{n_\theta}$.

Lemma 3.8 *For each $n_t \ge k, n_\theta \ge \ell$ and each $c \in H^{k,\ell}([0,1]\times S^1)$,*

$$\lim_{N_t,N_\theta\to\infty} \|c - \mathcal{S}_{N_t,N_\theta}^{n_t,n_\theta} c\|_{H^{k,\ell}([0,1]\times S^1)} = 0 .$$

The lemma is proven in App. B by showing that $H^{n_t,n_\theta}([0,1]\times S^1)$ is isometrically isomorphic to the Hilbert space tensor product of $H^{n_t}([0,1])$ and $H^{n_\theta}(S^1)$.

3.3.1 Discretization of the energy functional

The energy of a path of curves $c:[0,1]\times S^1\to\mathbb{R}^d$ is given by

$$E(c)=\int_0^1 G_c(\dot c,\dot c)\,\mathrm{d}t=\int_0^1\int_0^{2\pi} a_0|c'|\langle\dot c,\dot c\rangle+\frac{a_1}{|c'|}\langle\dot c',\dot c'\rangle+\frac{a_2}{|c'|^7}\langle c',c''\rangle^2\langle\dot c',\dot c'\rangle - \frac{2a_2}{|c'|^5}\langle c',c''\rangle\langle\dot c',\dot c''\rangle+\frac{a_2}{|c'|^3}\langle\dot c'',\dot c''\rangle\,\mathrm{d}\theta\,\mathrm{d}t\,, \tag{3.10}$$

as can be seen by combining (3.3) and (3.5). In the following let U denote the set of all paths $c\in H^{1,2}([0,1]\times S^1;\mathbb{R}^d)$ with nowhere vanishing spatial derivative, i.e., $c'(t,\theta)=\partial_\theta c(t,\theta)\neq 0$ holds for all $(t,\theta)\in[0,1]\times S^1$. Then U is an open subset of $H^{1,2}([0,1]\times S^1;\mathbb{R}^d)$ because $H^{1,2}([0,1]\times S^1;\mathbb{R}^d)$ embeds continuously into $C^{0,1}([0,1]\times S^1;\mathbb{R}^d)$ by Lem. B.3. The following lemma shows that the energy of a spline tends to the energy of the approximated curve as the number of control points tends to infinity.

Lemma 3.9 *If $n_t\geq 1$ and $n_\theta\geq 2$, then*

$$\lim_{N_t,N_\theta\to\infty} E(\mathcal{S}^{n_t,n_\theta}_{N_t,N_\theta}c)=E(c)$$

holds for each $c\in U$.

Proof. By Lem. 3.8 the spline approximations $\mathcal{S}^{n_t,n_\theta}_{N_t,N_\theta}c$ converge to c in $H^{1,2}([0,1]\times S^1)$. As U is open, $E(\mathcal{S}^{n_t,n_\theta}_{N_t,N_\theta}c)$ is well-defined for N_t,N_θ sufficiently large. The convergence $E(\mathcal{S}^{n_t,n_\theta}_{N_t,N_\theta}c)\to E(c)$ follows from the $H^{1,2}$-continuity of the energy functional.

To discretize the integrals in the definition of the energy functional we use Gaussian quadrature with m_t and m_θ quadrature points on each interval between consecutive knots. The total number of quadrature points is therefore $M_t=m_tN_t$ in time and $M_\theta=m_\theta N_\theta$ in space, and the discrete approximations of the Lebesgue measures on $[0,1]$ and S^1 are

$$\mu^{m_t}_{N_t}=\sum_{i=1}^{M_t} w_i\delta_{\bar t_i},\qquad \nu^{m_\theta}_{N_\theta}=\sum_{j=1}^{M_\theta}\omega_j\delta_{\bar\theta_j},$$

where w_i, ω_j are the Gaussian quadrature weights and $\bar{t}_i, \bar{\theta}_j$ the Gaussian quadrature points. We define the discretized energy $E^{m_t,m_\theta}_{N_t,N_\theta}(c)$ of a curve $c \in C^{1,2}([0,1]\times S^1) \cap U$ to be given by the right-hand side of (3.10) with $\mathrm{d}t\,\mathrm{d}\theta$ replaced by $\mu^{m_t}_{N_t}(\mathrm{d}t)\nu^{m_\theta}_{N_\theta}(\mathrm{d}\theta)$. The following theorem shows that the discretized energy of a path tends to the energy of the approximated path as the number of control points go to infinity, provided that the path is smooth enough.

Theorem 3.10 *If $n_t \geq 2$, $n_\theta \geq 3$, and $m_t, m_\theta \geq 1$, then*

$$\lim_{N_t,N_\theta\to\infty} E^{m_t,m_\theta}_{N_t,N_\theta}(\mathcal{S}^{n_t,n_\theta}_{N_t,N_\theta} c) = E(c)$$

holds for each $c \in U \cap H^{2,3}([0,1]\times S^1)$.

Proof. The total error can be decomposed into a spline approximation error and a quadrature error:

$$|E^{m_t,m_\theta}_{N_t,N_\theta}(\mathcal{S}^{n_t,n_\theta}_{N_t,N_\theta} c) - E(c)| \leq |E^{m_t,m_\theta}_{N_t,N_\theta}(\mathcal{S}^{n_t,n_\theta}_{N_t,N_\theta} c) - E^{m_t,m_\theta}_{N_t,N_\theta}(c)| + |E^{m_t,m_\theta}_{N_t,N_\theta}(c) - E(c)|\,. \tag{3.11}$$

To show that the first summand on the right-hand side tends to zero, note that the spline approximations $\mathcal{S}^{n_t,n_\theta}_{N_t,N_\theta} c$ converge to c in $H^{2,3}([0,1]\times S^1)$ by Lem. 3.8. They also converge in $C^{1,2}([0,1]\times S^1)$ by Lem. B.3. Let $F(c)$ denote the integrand in (3.10). Then F is locally Lipschitz continuous when seen as a mapping from $U \cap C^{1,2}([0,1]\times S^1)$ to $C([0,1]\times S^1)$. Let L denote the Lipschitz constant of F near c. Then the first summand in (3.11) can be estimated for sufficiently large N_t, N_θ via

$$\begin{aligned} |E^{m_t,m_\theta}_{N_t,N_\theta}(\mathcal{S}^{n_t,n_\theta}_{N_t,N_\theta} c) - E^{m_t,m_\theta}_{N_t,N_\theta}(c)| &\leq \iint |F(\mathcal{S}^{n_t,n_\theta}_{N_t,N_\theta} c) - F(c)|\mu^{m_t}_{N_t}\nu^{m_\theta}_{N_\theta} \\ &\leq L\|\mathcal{S}^{n_t,n_\theta}_{N_t,N_\theta} c - c\|_{C^{1,2}([0,1]\times S^1)} \to 0\,. \end{aligned}$$

It remains to show that the second summand in (3.11) tends to zero. As the Gaussian quadrature rules $\mu^{m_t}_{N_t}$ and $\nu^{m_\theta}_{N_\theta}$ are of order $m_t, m_\theta \geq 1$, there is $K > 0$ such that the following estimates hold for all $f \in C^1([0,1])$ and $g \in C^1(S^1)$:

$$\int_{[0,1]} f(t)\big(\mu^{m_t}_{N_t}(\mathrm{d}t) - \mathrm{d}t\big) \leq K N_t^{-1}\|f'\|_{C([0,1])}\,, \qquad \int_{S^1} g(\theta)\big(\nu^{m_\theta}_{N_\theta}(\mathrm{d}\theta) - \mathrm{d}\theta\big) \leq K N_\theta^{-1}\|g'\|_{C(S^1)}\,.$$

See e.g. [22, Thm. 4.3.1] for this well-known result. Therefore, the second summand in (3.11) satisfies

$$\begin{aligned} |E^{m_t,m_\theta}_{N_t,N_\theta}(c) - E(c)| &= \left|\iint F(c)(t,\theta)(\mu^{m_t}_{N_t}(\mathrm{d}t)\nu^{m_\theta}_{N_\theta}(\mathrm{d}\theta) - \mathrm{d}t\,\mathrm{d}\theta)\right| \\ &\leq \left|\iint F(c)(t,\theta)\big(\mu^{m_t}_{N_t}(\mathrm{d}t) - \mathrm{d}t\big)\nu^{m_\theta}_{N_\theta}(\mathrm{d}\theta)\right| + \left|\iint F(c)(t,\theta)\,\mathrm{d}t\big(\nu^{m_\theta}_{N_\theta}(\mathrm{d}\theta) - \mathrm{d}\theta\big)\right| \\ &\leq K N_t^{-1}\|\partial_t F(c)\|_{C([0,1]\times S^1)} + K N_\theta^{-1}\|\partial_\theta F(c)\|_{C([0,1]\times S^1)} \to 0\,. \end{aligned} \tag{3.12}$$

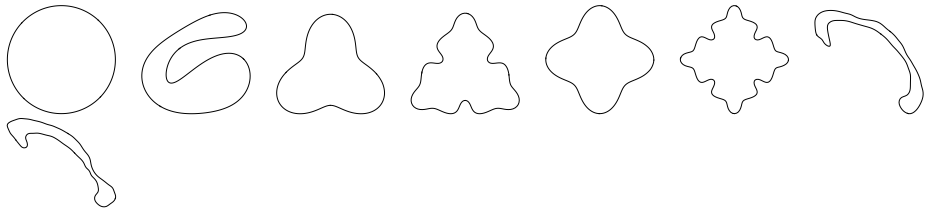


Figure 3.2: Curves that are used in the remainder of the section to test convergence of the proposed algorithms: circle, wrap, 3- and 4-bladed propellers without and with noise, and two corpus callosum shapes.[2]

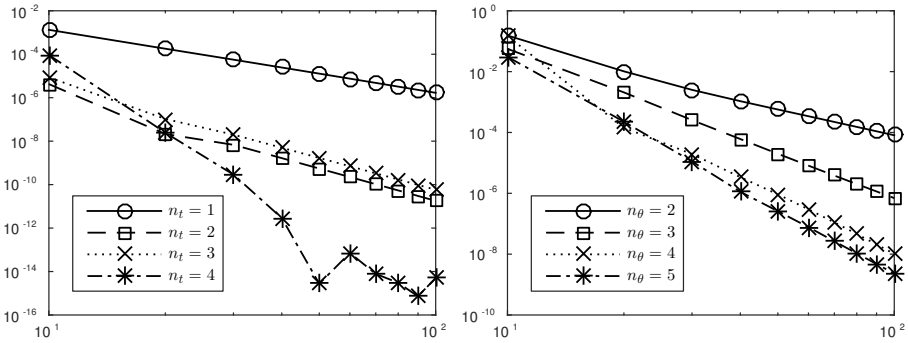


Figure 3.3: Convergence of the discrete energy: relative energy differences for increasing number of control points of the non-linear path $c(t,\theta) = c_0(\theta)\sin(1-t\pi/2)) + c_1(\theta)\sin(t\pi/2)$ connecting the circle c_0 to the wrap c_1. Left: varying N_t with fixed $n_\theta = 4$, $N_\theta = 60$. Right: varying N_θ and fixed $n_t = 3$, $N_t = 20$.

This shows that the total error (3.11) tends to zero as N_t, N_θ tend to infinity.

To confirm this theoretical result, we run a series of numerical experiments to test the convergence of the discrete energy, whose results are displayed in Fig. 3.3. The set of basic curves that we will use throughout the whole section in all numerical experiments is displayed in Fig. 3.2.

[2]The acquisition of the corpus callosum shapes is described in [37].

3.3.2 Boundary value problem for parameterized curves

Solving the geodesic boundary problem means, for given boundary curves c_0 and c_1, to find a path c which is a (local) minimum of the energy functional (3.5) among all paths with the given boundary curves. For existence of minimizers see Theorem 3.4. We will assume that the curves c_0, c_1 are discretized, i.e., given as linear combinations of the basis functions C_j. Should the curves be given in some other form, one would first approximate them by splines using a suitable approximation method.

The choice of full multiplicity for the boundary knots (in t) implies that the identity (3.8) for $t \in \{0, 1\}$ and a spline path c becomes

$$c(0,\theta) = \sum_{j=1}^{N_\theta} c_{1,j} C_j(\theta)\,, \quad c(1,\theta) = \sum_{j=1}^{N_\theta} c_{N_t,j} C_j(\theta)\,.$$

If the controls $c_{1,j}$ and $c_{N_t,j}$ are fixed, then (3.8) defines a family of paths between between the boundary curves $c_0(\theta) = \sum_{j=1}^{N_\theta} c_{1,j} C_j(\theta)$ and $c_1(\theta) = \sum_{j=1}^{N_\theta} c_{N_t,j} C_j(\theta)$. The family is indexed by the remaining control points $c_{2,j}, \ldots, c_{N_t-1,j}$. Discretizing the energy functional as described in Sect. 3.3.1 transforms the geodesic boundary value problem to the finite-dimensional optimization problem

$$\operatorname{argmin} E_{\text{discr}}(c_{2,1}, \ldots, c_{N_{t-1},N_\theta})\,. \tag{3.13}$$

where E_{discr} denotes the discretized energy functional $E^{m_t,m_\theta}_{N_t,N_\theta}$ applied to the spline defined by the control points $c_{i,j}$. This finite-dimensional minimization problem can be solved by conventional black-box methods, specifically we use Matlab's `fminunc` function. To speed up the optimization we analytically calculated the gradient and Hessian of the energy functional E. We notice that

$$\frac{\partial E_{\text{discr}}}{\partial c_{i,j}} = dE_c(B_i(t)C_j(\theta))\,.$$

The formulas for the derivative and the Hessian are provided in App. C.

Remark 4 For gradient-based optimization methods to work one must provide an initial path. An obvious choice for a path between two curves c_0, c_1 is the linear path $(1-t)c_0 + tc_1$. This path can always be constructed, but it is not always a valid initial path for the optimization procedure. For plane curves the space $\operatorname{Imm}(S^1, \mathbb{R}^2)$ is disconnected with the winding number of a curve determining the connected component [42]. The metric (3.2) is defined only for immersions, and a path leaving the space of immersion – for example

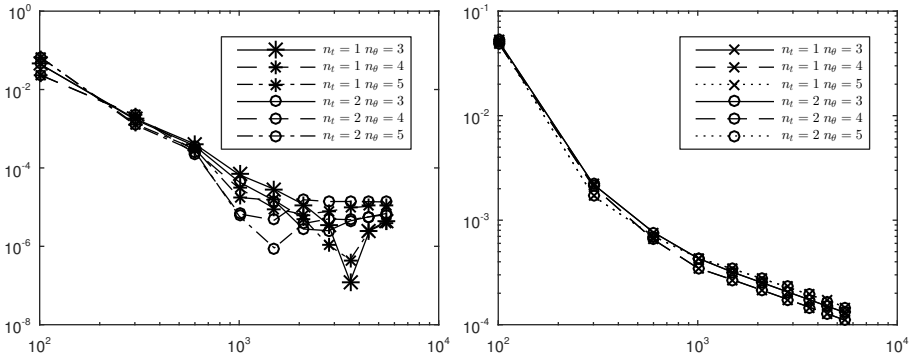


Figure 3.4: Left/Right: Relative energy difference $\frac{|E_i - E_{i-1}|}{E_{i-1}}$ and L^2-distance $\frac{\|c_i - c_{i-1}\|_{L^2}}{\|c_{i-1}\|_{L^2}}$, for the propeller shapes, as a function of increasing number of control points. The values of (N_t, N_θ) are $(10,10), (15,20), \ldots, (60,110)$.

as it passes from one connected component to another – will lead to a blow up of the energy (3.5). Hence an initial path connecting two curves must not leave $\mathrm{Imm}(S^1, \mathbb{R}^d)$. For most examples considered in this paper the linear path is a valid initial guess; for more complicated cases a different strategy might be needed.

Remark 5 Note that the tensor product structure in (3.8) allows us to evaluate $\mathrm{Log}_{c_0} c_1$ by taking a time derivative of the path $c(t,\theta)$ and evaluating it at $t = 0$ to obtain $\mathrm{Log}_{c_0} c_1 = \partial_t c(0,\cdot)$, where c is a solution of the geodesic boundary value problem.

Now we prove a result about Γ-convergence of the discrete energy functional. Before stating the theorem, we set up some notation. For brevity we denote the spline approximation operators in Lemma 3.8 by $\mathcal{S}_N$, the space $H^{1,2}([0,1] \times S^1; \mathbb{R}^d)$ by $H^{1,2}$, and the space $H^2(S^1; \mathbb{R}^d)$ by H^2. Let $\Omega_{c_0,c_1} H^{1,2}$ denote the closed subset of paths $c \in H^{1,2}$ for which $c(0,\cdot) = c_0$ and $c(1,\cdot) = c_1$. We extend the restriction $E|_{\Omega_{c_0,c_1} H^{1,2}}$ by ∞ to all of $H^{1,2}$, and denote this extension by E_Ω. We define the following discrete energy functionals on all of $H^{1,2}$ as follows

$$E_{N,\Omega}(c) = \begin{cases} E(c), & c \in S_N(\Omega_{c_0,c_1} H^{1,2}) \cap U, \\ \infty, & \text{otherwise.} \end{cases}$$

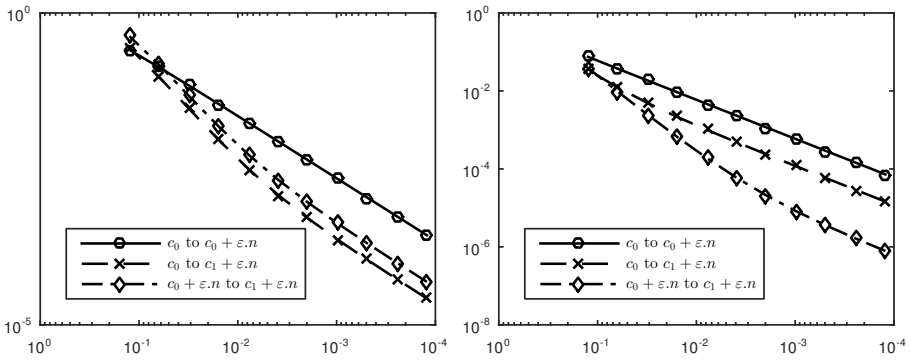


Figure 3.5: Continuity of the geodesic distance function. Left: c_0, c_1 are 3- and 4-bladed propeller shapes, perturbed by a sinusoidal displacement in the normal direction of the curve. Right: c_0, c_1 are corpus callosum shapes with the perturbation applied directly to the control points. The plots show the relative change in distance against the amplitude of the sinusoidal noise ε, i.e., $\operatorname{dist}(c_0, c_1 + \varepsilon.n)/\operatorname{dist}(c_0, c_1)$.

Here U is the domain of definition of E defined in section 3.3.1. That is, $E_{N,\Omega}$ is the restriction of E to the spline spaces which connects spline approximations of c_0 and c_1, and extended by ∞ elsewhere. Our result is the following

Theorem 3.11 *If $n_t \geq 1$ and $n_\theta \geq 2$, then the discretized energy functionals $E_{N,\Omega}$ are equi-coercive with respect to the weak $H^{1,2}$ topology and Γ-converge with respect to the weak $H^{1,2}$ topology to the energy functional E_Ω as $N_t, N_\theta \to \infty$. It follows that the sequence of minimizers of the discretized energy functionals $E_{N,\Omega}$ has a subsequence that converges weakly to a minimizer of E_Ω.*

We refer to [28, Definitions 4.1 and 7.6] for the concepts of equi-coercivity and Γ-convergence and to [28, Chapter 8] for Γ-convergence under weak topologies.

Proof. First we show that the functionals $E_{N,\Omega}$ are equi-coercive with respect to the weak topology on $H^{1,2}$. This means that for each $r > 0$ there is a weakly compact set K_r such that for each N, $\{c \in H^{1,2} : E_{N,\Omega}(c) \leq r^2/2\} \subseteq K_r$. To see this let $r > 0$ and $c \in H^{1,2}$ with $E_{N,\Omega}(c) \leq r^2/2$. Then $E(c) \leq r^2/2$ and consequently $L(c) \leq r$. Therefore, $\operatorname{dist}(c(0), c(t)) \leq r$ holds for all $t \in [0,1]$. By [23, Prop. 3.5 and Lem. 4.2] there exists constants $C_1, C_2 > 0$ such that $\|h\|^2_{H^2} \leq C_1 G_{\tilde{c}}(h,h)$ and $\|c(0) - \tilde{c}\|_{H^2} \leq C_2 \operatorname{dist}(c(0), \tilde{c})$ holds for all $\tilde{c} \in \mathcal{I}^2(S^1, \mathbb{R}^d)$ satisfying $\operatorname{dist}(c(0), \tilde{c}) \leq r$ and all $h \in H^2$. Since we have full multiplicity at the ends we have $c(0) = S^{n_\theta}_{N_\theta}(c_0)$, by classical approximation results there exists a

constant C_3 such that $\|S_{N_\theta}^{n_\theta}(c_0)\|_{H^2} \le C_3\|c_0\|_{H^2}$ where C_3 doesn't depend on N_θ. Then there is $C_4 > 0$ such that

$$\begin{aligned}\|c\|_{H^{1,2}}^2 &= \int_0^1 \|c(t)\|_{H^2}^2 + \|\dot{c}(t)\|_{H^2}^2 dt \\ &\le C_2(\|c(0)\|_{H^2} + r^2) + C_1 E(c) \\ &\le C_4(\|c_0\|_{H^2} + 3r^2/2) =: R_r^2.\end{aligned}$$

This shows that $\{c \in H^{1,2} : E_{N,\Omega}(c) \le r^2/2\}$ is contained in the set K_r, which we define as the ball of radius R_r in $H^{1,2}$. As this ball is weakly compact, the functionals $E_{N,\Omega}$ are equi-coercive.

The equi-coercivity, which we have just shown, and [28, Prop. 8.16] give the following sequential characterization of Γ-convergence: the functionals $E_{N,\Omega}$ Γ-converge to E_Ω if

$$\forall c\ \forall c_N \rightharpoonup c: \qquad E_\Omega(c) \le \liminf_{N\to\infty} E_{N,\Omega}(c_N), \tag{3.14}$$

$$\forall c\ \exists c_N \rightharpoonup c: \qquad E_\Omega(c) = \lim_{N\to\infty} E_{N,\Omega}(c_N). \tag{3.15}$$

To prove (3.14) let $c_N \rightharpoonup c$ in $H^{1,2}$. First we consider the case $c \in \Omega_{c_0,c_1}H^{1,2}$. Then $E(c) = E_\Omega(c)$. Notice that $E(c) \le E_{N,\Omega}(c)$ always. It was shown in the proof of [23, Thm 5.2] that E is sequentially weakly lower semicontinuous, so we obtain

$$E_\Omega(c) = E(c) \le \liminf_{n\to\infty} E(c_N) \le \liminf_{n\to\infty} E_{N,\Omega}(c_N).$$

Now if $c \notin \Omega_{c_0,c_1}H^{1,2}$ then either $c(0) \ne c_0$ or $c(1) \ne c_1$, w.l.o.g we can assume the former, and $E_\Omega(c) = \infty$. By Lem B.3 we have $c_N \to c$ in $C^{0,1}$, especially $c_N(0) \nrightarrow c_0$ in C^0. For big enough N we must then have $c_N \notin S_N(\Omega_{c_0,c_1}H^{1,2})$, so $E_{N,\Omega}(c_N) = \infty$ and (3.14) is satisfied. Equation (3.15) follows directly from Lem. 3.9. Thus, we have shown that the functionals $E_{N,\Omega}$ are equi-coercive and Γ-converge to E_Ω in the weak $H^{1,2}$ topology. The statement about the convergence of minimizers is well-known and can be found in [28, Thm. 7.23].

Note that the result concerns the energy functional restricted to spline spaces, but evaluation of integrals is assumed to be exact. For the case where we approximate the integrals by gaussian quadrature, we were not able to prove a Γ-convergence result. However, in numerical experiments we still observe convergence for the solution of the boundary value problem, as can be seen in Fig. 3.4 for varying numbers of control points. Convergence holds for both the optimal energy and the L^2-norm of the minimizing paths. In Fig. 3.5 we show that the geodesic distance function is continuous: by adding a sinusoidal displacement in the normal direction to the curves, the geodesic distance converges to 0 as the noise becomes smaller.

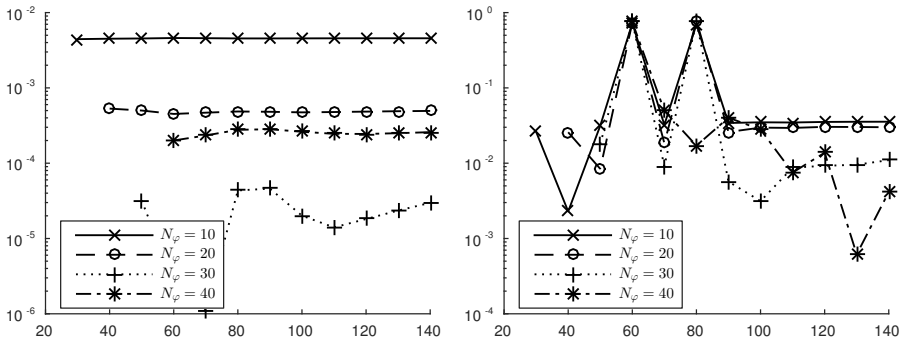


Figure 3.6: Symmetry of the geodesic distances for the 3- and 4-bladed propeller shapes on the left and the corpus callosum shapes on the right. The relative difference $|\operatorname{dist}(c_0, c_1) - \operatorname{dist}(c_1, c_0)| / \operatorname{dist}(c_0, c_1)$ is plotted against N_θ for different choices of N_φ.

3.3.3 Boundary value problem for unparameterized curves

To numerically solve the boundary value problem on the space of unparametrized curves, we first have to discretize the diffeomorphism group. By the identification of S^1 with $\mathbb{R}/[0, 2\pi]$, diffeomorphisms $\varphi: S^1 \to S^1$ can be written as $\varphi = \mathrm{Id} + f$, where f is a periodic function. Periodic functions can be discretized as before using simple knot sequences with periodic boundary conditions. This leads to the spline representation

$$\varphi(\theta) = \sum_{i=1}^{N_\varphi} \varphi_i D_i(\theta) = \sum_{i=1}^{N_\varphi} (\xi_i + f_i)\, D_i(\theta)\,.$$

Here D_i are B-splines of degree n_φ, defined on a uniform periodic knot sequence, f_i are the control points of f, i.e., $f(\theta) = \sum_{i=1}^{N_\varphi} f_i D_i(\theta)$, and ξ_i are the *Greville abscissas*, i.e., control points of the identity represented in a B-spline basis, $\mathrm{Id} = \sum_{i=1}^{N_\varphi} \xi_i D_i$.

The constraint that φ is a diffeomorphism is $\varphi' > 0$. By the fact that the B-spline basis functions are nonnegative and by the recursive formula for the derivatives of B-splines, see [68, Chap. 4], a sufficient condition to ensure that $\varphi' > 0$ is

$$f_{i-1} - f_i < \xi_i - \xi_{i-1}\,. \tag{3.16}$$

This is a linear inequality constraint. To speed up convergence, we introduce an additional variable $\alpha \in \mathbb{R}$ representing constant shifts of the reparametrization.

The resulting redundancy is eliminated by the constraint

$$\sum_{i=1}^{N_\varphi} f_i = 0\,, \tag{3.17}$$

which ensures that the average shift of φ is 0.

We have to minimize the energy functional (3.5) over all paths $c\colon [0,1]\times[0,2\pi] \to \mathbb{R}^2$, and diffeomorphisms φ, subject to the constraints

$$c(0,\cdot) = c_0\,, \qquad c(1,\cdot) = c_1 \circ \varphi\,.$$

It is important to note that the reparametrization $(c,\varphi) \mapsto c\circ\varphi$ does not preserve splines: if c_1 and φ are represented by splines, then the function $c_1 \circ \varphi$ is in general not. To overcome this difficulty we approximate the reparameterized curve $c_1 \circ \varphi$ by a new spline in each optimization step. This then leads to a finite-dimensional constrained minimization problem

$$\operatorname{argmin} E_{\text{discr}}(c_{2,1},\dots,c_{N_t-1,N_\theta}, f_1,\dots,f_{N_\varphi},\alpha)\,, \tag{3.18}$$

where $f_1,\dots,f_{N_\varphi}$ are the controls used to construct the diffeomorphism φ and α is the constant shift in the parametrization. Similar to the unconstrained problem (3.13), we can analytically compute the gradient and hessian and then solve this by standard methods for constrained minimization problems, specifically we use Matlab's `fmincon` function. In order to use `fmincon` we replace (3.16) by $f_{i-1} - f_i \le \xi_i - \xi_{i-1} + \varepsilon$ with ε small.

From a mathematical point of view we would expect the geodesic distance between two shapes to be symmetric, i.e., interchanging the curves c_0 and c_1 should have no effect on the resulting geodesic distance. This is, however, only approximately true numerically: in our numerical examples the relative error is below 5% if sufficiently many grid points are chosen, see Fig. 3.6. Other articles on numerical methods for computing geodesic distances between unparametrized curves seem to sidestep this question, see e.g. [77, 54, 31]. An example of a forward and backward geodesic is plotted in Fig. 3.7.

3.3.4 Boundary value problem on shape space

To numerically solve the boundary value problem on shape space, it remains to discretize the finite-dimensional motion group. To simplify the presentation we will assume in the following that $d = 2$, so that we can parametrize rotations by the one-dimensional parameter β. We have to add translations and rotations to the minimization problem, i.e., minimize the energy functional (3.5) over all

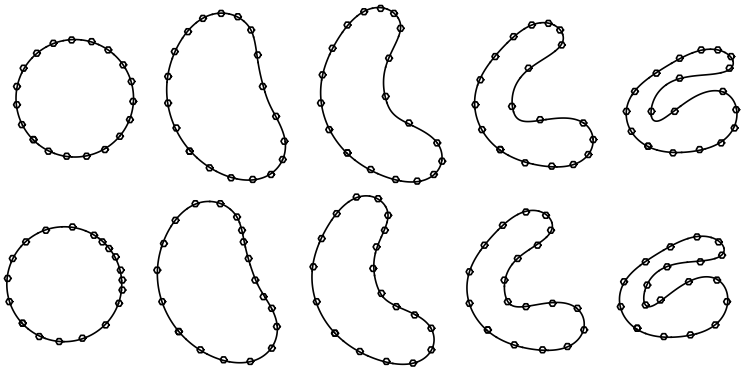


Figure 3.7: Symmetry of the geodesic boundary value problem on the space of unparametrized curves. For the circle and the wrap the geodesic boundary value problem is solved forwards and backwards. To better compare the results the second geodesic is plotted backwards in time. The plot markers visualize the optimal parametrization of the curves.

paths $c\colon [0,1] \times [0,2\pi] \to \mathbb{R}^2$, diffeomorphisms φ, rotations R_β and translations a, subject to the constraints

$$c(0,\cdot) = c_0\,, \qquad c(1,\cdot) = R_\beta(c_1 \circ \varphi + a)\,.$$

Note that rotations and translations preserve splines: if c_1 is represented by a spline then for any rotation R_β and translations a the function $R_\beta(c_1 + a)$ is a spline of the same type. This then leads to a finite-dimensional constrained minimization problem

$$\operatorname{argmin} E_{\text{discr}}(c_{2,1}, \dots, c_{N_t-1,N_\theta}, f_1, \dots, f_{N_\varphi}, \alpha, \beta, a)\,, \tag{3.19}$$

where $f_1, \dots, f_{N_\varphi}$ are the controls used to construct the diffeomorphism φ, α is the constant shift in the parametrization, β the rotation angle and a the translation vector.

3.3.5 Initial value problem

To solve the geodesic initial value problem we use the variational discrete geodesic calculus developed in [67]. For a discrete path $(c_0, \dots, c_K)$, $K \in \mathbb{N}$, one defines the discrete energy

$$E_K(c_0, \dots, c_K) = K \sum_{k=1}^{K} W(c_{k-1}, c_k)\,,$$

where $W(c,\tilde{c})$ is an approximation of $\operatorname{dist}(c,\tilde{c})^2$. Since our Riemannian metric G is smooth, it approximates the squared distance sufficiently well in the sense that $G_c(c-\tilde{c},c-\tilde{c}) - \operatorname{dist}(c,\tilde{c})^2 = O(\operatorname{dist}(c,\tilde{c})^3)$, and we can take the approximation to be

$$W(c,\tilde{c}) = \frac{1}{2} G_c(c-\tilde{c},c-\tilde{c})\,.$$

We call $(c_0,\dots,c_K)$ a discrete geodesic if it is a minimizer of the discrete energy with fixed endpoints c_0, c_K. To define the discrete exponential map we consider discrete paths (c_0,c_1,c_2) consisting of three points. The discrete energy of such a path is

$$E_2(c_0,c_1,c_2) = G_{c_0}(c_1-c_0,c_1-c_0) + G_{c_1}(c_2-c_1,c_2-c_1)\,.$$

Given c_0, c_1, we define $c_2 = \operatorname{Exp}_{c_0} c_1$ if (c_0,c_1,c_2) is a discrete geodesic, in other words, if $c_1 = \operatorname{argmin} E_2(c_0,\cdot,c_2)$. Given an initial curve c_0, an initial velocity v_0, and a number K of time steps, our solution of the geodesic initial value problem is $c_K = \operatorname{Exp}_{c_{K-2}} c_{K-1}$, where the intermediate points $c_1,\dots,c_{K-1}$ are defined iteratively via

$$c_1 = c_0 + \frac{1}{K} v_0\,, \quad c_2 = \operatorname{Exp}_{c_0} c_1\,, \quad c_3 = \operatorname{Exp}_{c_1} c_2\,, \quad \dots\,, \quad c_{K-1} = \operatorname{Exp}_{c_{K-3}} c_{K-2}\,.$$

To compute a discrete geodesic we need to find minima of the function $E_2(c_0,\cdot,c_2)$. Differentiating E_2 with respect to c_1 leads to the following system of nonlinear equations

$$2G_{c_0}(c_1-c_0,\cdot) - 2G_{c_1}(c_2-c_1,\cdot) + D_{c_1}G_\cdot(c_2-c_1,c_2-c_1) = 0\,.$$

This system has to be solved for c_1, with the argument replaced by all basis functions C_j defining the spline space. We use the solver `fsolve` in Matlab to solve this system of equations. Some examples of discrete geodesics are depicted in Fig. 3.12. The discretizations of the geodesic initial and boundary value problems are compatible as demonstrated in Fig. 3.8.

3.3.6 Karcher mean

The Karcher mean $\bar{c}$ of a set $\{c_1,\dots,c_n\}$ of curves is the minimizer of

$$F(c) = \frac{1}{n}\sum_{j=1}^{n} \operatorname{dist}(c,c_j)^2\,. \tag{3.20}$$

It can be calculated by a gradient descent on $(\operatorname{Imm}(S^1,\mathbb{R}^d),G)$. Letting $\operatorname{Log}_c c_j$ denote the Riemannian logarithm, the gradient of F with respect to G is [60]

$$\operatorname{grad}^G F(c) = -\frac{2}{n}\sum_{j=1}^{n} \operatorname{Log}_c c_j\,. \tag{3.21}$$

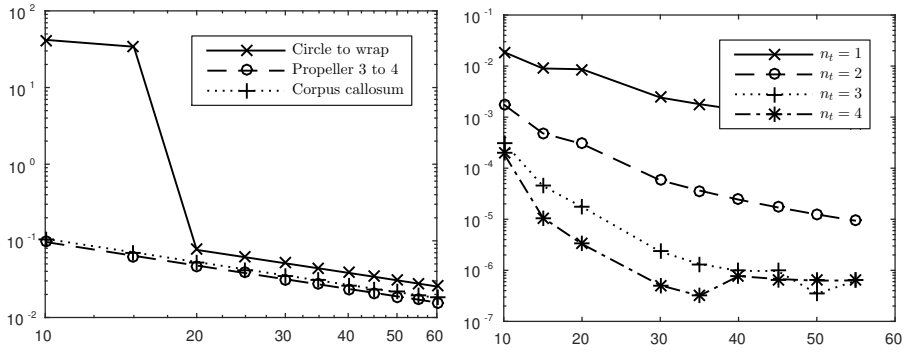


Figure 3.8: Compatibility of the geodesic IVP and BVP with increasing N_t. On the left one computes $c_1 = \mathrm{Exp}_{c_0}(v)$ for given c_0, v and then solves the BVP for $\tilde{v} = \mathrm{Log}_{c_0}(c_1)$. The plot shows the relative distance $\|v - \tilde{v}\|_{c_0} / \|v\|_{c_0}$ agains N_t. On the right one computes $v = \mathrm{Log}_{c_0}(c_1)$ for given c_0, c_1 and plots the relative difference $\|v - \tilde{v}\|_{c_0} / \mathrm{dist}(c_0, c_1)$ between two consecutive (w.r.t. N_t) initial velocities $v, \tilde{v}$ against N_t.

Fig. 3.9 illustrates the computation of the Karcher mean of 8 propeller shapes, which have all been modified by adding a 10% uniform noise to their control points.

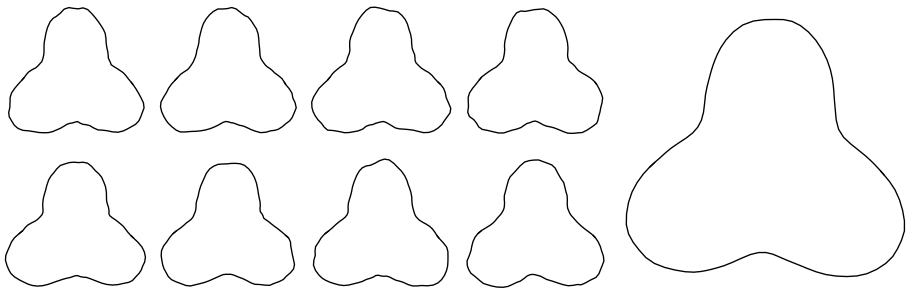


Figure 3.9: Eight propellers with 10% uniform noise added to their control points, along with their Karcher mean.

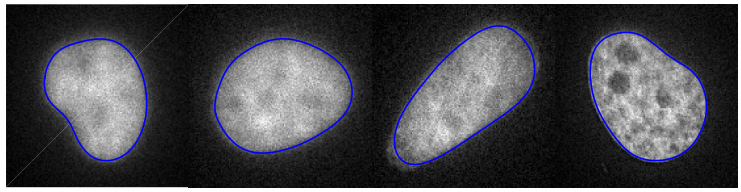


Figure 3.10: Examples of HeLa cell nuclei and the spline representation of the boundary.

3.4 Shape analysis of HeLa cells

We used second order metrics to characterize the nuclear shape variation in HeLa cells. The data consists of fluorescence microscope images of HeLa cell nuclei[3] (87 images in total). The acquisition of the images is described in [20].

To extract the boundary of the nucleus we apply a thresholding method [58] to obtain a binary image, and then we fit – using least squares – a spline with $N_\theta = 12$ and $n_\theta = 4$ to the longest 4-connected component of the thresholded image. This provides a good balance between capturing shape details and not overfitting the image noise; see Fig. 3.10. Then we rescale all curves by the same factor to arrive at an average length $\bar{\ell}_c = 2\pi$. The choice $\bar{\ell}_c = 2\pi$ has the following nice property: if a curve c has $\ell_c = 2\pi$ and c has a constant speed parametrization, then $|c'| = 1$, and the arc length derivative $D_s h$ coincides with the regular derivative h'. The scaling matters because the metric we work with is not scale invariant. Had we decided to work with curves of a different average length we would have to change the constants a_j of the metric in order to arrive at the same results.

For the subsequent analysis we use splines with $N_\theta = 40$ and $n_\theta = 3$. The increased number of control points compared to the data acquisition allows us to preserve shape information even after reparametrizing the curves. To parametrize the diffeomorphism group we use splines with $N_\varphi = 20$ and $n_\varphi = 3$. This leaves us with roughly $2 \cdot 40 - 20 - 2 - 1 = 57$ degrees of freedom to represent the population of 87 given shapes of cell nuclei. The influence of the number of control points on the geodesic BVP can be seen in Fig. 3.11. All analysis is performed modulo translations, rotations, and reparametrizations.

The choice of constants a_0, a_1, and a_2 of the Riemannian metric has a significant impact on the results; see Fig 3.14. One constant may be chosen freely, so we set $a_0 = 1$. To simplify the interpretation of the results, we set $a_1 = 0$; our

[3]The dataset was downloaded from `http://murphylab.web.cmu.edu/data`.

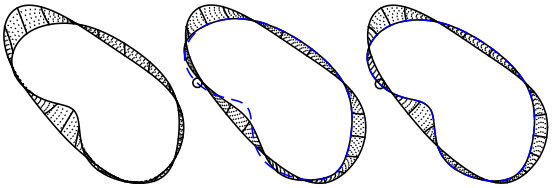


Figure 3.11: Geodesic between two cells (solid lines); the dashed line shows the exact endpoint before reparametrization. The geodesic is computed between parametrized curves with $N_\theta = 12$ (left), unparametrized curves with $N_\theta = 12$ (middle) and $N_\theta = 40$ (right).

metric shall have no H^1-part. This leaves us with one more parameter, which we choose by looking at the L^2- and H^2-contributions to the energy of geodesics between shapes in the dataset. For a geodesic c between two curves c_0 and c_1 these contributions are

$$\operatorname{dist}(c_0, c_1)^2 = E_{L^2}(c) + E_{H^2}(c) = \int_0^1 \int_{S^1} |c_t|^2 \,\mathrm{d}s\,\mathrm{d}t + a_2 \int_0 \int_{S^1} \left|D_s^2 c_t\right|^2 \,\mathrm{d}s\,\mathrm{d}t\,.$$

The relative contribution of the H^2-term to the total energy is $\varrho_{H^2} = E_{H^2}/(E_{L^2} + E_{H^2})$. We denote the population mean and standard deviation of the variable ϱ_{H^2} by $\bar{\varrho}_{H^2}$ and σ, respectively. The following table shows that the choices $a_2 = 2^{-12} \approx 0.00024$ and $a_2 = 2^{-8} \approx 0.0039$ both lead to balanced energy contributions of the zero and second order terms:

$$\begin{aligned} a_2 &= 2^{-12}\,, & \bar{\varrho}_{H^2} &= 0.032\,, & \sigma &= 0.027\,, \\ a_2 &= 2^{-8}\,, & \bar{\varrho}_{H^2} &= 0.203\,, & \sigma &= 0.119\,. \end{aligned}$$

We will use these parameter choices in our subsequent analysis. Note that from a physical point of view the parameter a_2 has units $[m^4]$, m being meters.

The average shape of the nucleus can be captured by the Karcher mean $\bar{c}$. To solve the minimization problem (3.20) for the Karcher mean of the 87 nuclei we use a conjugate gradient method on the Riemannian manifold of curves as implemented in the Manopt library [21]. For each choice of parameters the optimization is performed until the gradient of the objective function $F(\bar{c})$ satisfies $\|\operatorname{grad}^G F(\bar{c})\|_{\bar{c}} < 10^{-3}$.

Having computed the mean $\bar{c}$, we represent each nuclear shape c_j by the initial velocity $v_j = \operatorname{Log}_{\bar{c}}(c_j)$ of the minimal geodesic from $\bar{c}$ to c_j. We perform principal component analysis with respect to the inner product $G_{\bar{c}}$ on the set of initial velocities $\{v_j : j = 1, \ldots, 87\}$. Geodesics from the mean in the first five directions can be seen in Fig. 3.12. A projection of the dataset onto the subspace spanned by the first two principal components is depicted in Fig. 3.13.

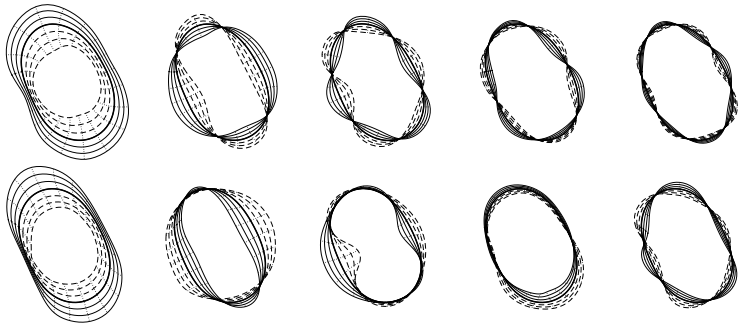


Figure 3.12: Geodesics from the mean in the first five principal directions. The curves show geodesics at times $-3, -2, \ldots, 2, 3$; the mean is shown in bold. One can see characteristic deformations of the curve: expansion, stretching, compressing and bending. The first row shows principal components calculated for curves modulo reparametrizations; the second row for parametrized curves.

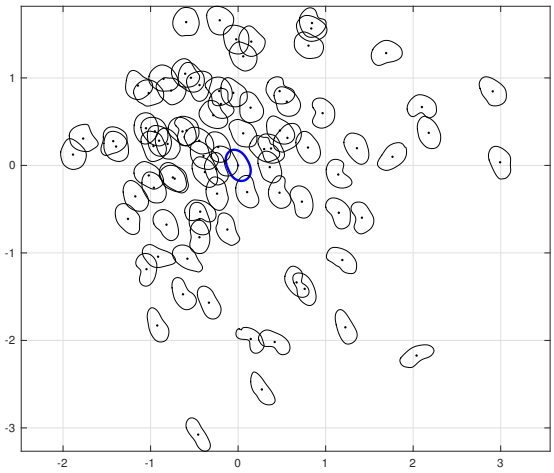


Figure 3.13: Cell nuclei projected to the plane in the tangent plane, spanned by the first two principal components. The mean (in blue) is situated at the origin. The units on the coordinate axes are standard deviations.

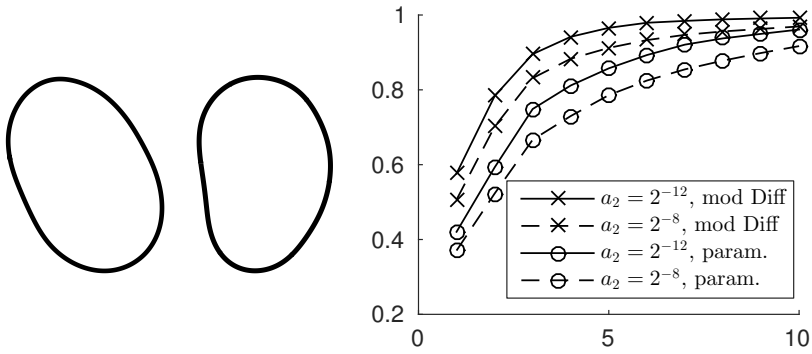


Figure 3.14: Left: the mean shape of cell nuclei. Middle: for comparison, the mean shape as computed in [66] via the Christensen–Rabbitt–Miller method [32]. Right: the proportion of the total variance explained by the first 10 eigenvectors.

For unparametrized curves and for the parameter choice $a_2 = 2^{-12}$ the first five principal components explain 57.6%, 78.3%, 90.0%, 94.2% and 98.0% of the total variance; see Fig. 3.14. Under the choice $a_2 = 2^{-8}$ the first five principal components explain only 93.3% of the variance as compared to 98.0% in the previous case. This demonstrates that approximation power of the principal components depends on the choice of the metric. Fig. 3.14 also shows that fewer principal components are needed to explain the same amount of variance when the reparametrization group is factored out.

The results we obtain are comparable to those of [66], where diffeomorphic matching was used to compare cells. It turns out that the mean shape with respect to our metrics is symmetric, while the mean shape obtained in [66] is bent towards one side; see Fig. 3.14.

3.5 Further applications

Now we provide some additional applications taken from the peer-reviewed conferences papers [6] and [7]. The primary contributions of these papers was to provide the first working implementations of the geodesic BVP on parametrized and unparametrized curves respectively. The content of this section is a copy of the appropriate sections in each paper respectively, adjusted to previous sections of this chapter.

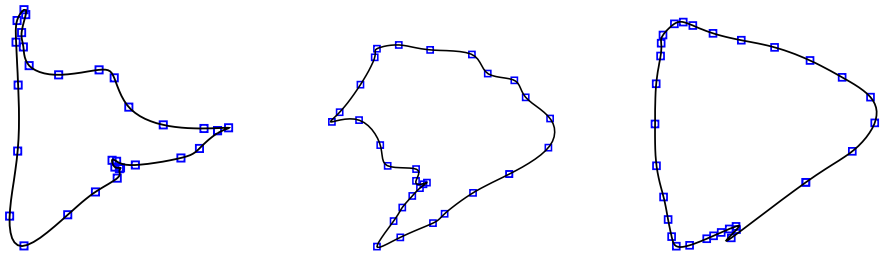


Figure 3.15: Projections to a two-dimensional barycentric subspace of 30 images from the cardiac cycles of three patients. Cubic splines interpolation of degree $n_\theta = 3$ with $N_\theta = 30$ control points is used.

3.5.1 Traces of cardiac images

In our second application we study curves that are obtained from images of the cardiac cycle. More precisely, we consider a sequence of 30 cardiac images, taken at equispaced time points along the cardiac cycle. Each image is projected to a barycentric subspace of dimension two, yielding a closed curve in the two-dimensional space of barycentric coordinates. After normalizing the coordinates [61, Sect. 3] we obtain a closed, plane curve – with the curve parameter representing time – to which we can apply the methods presented in Sect. 3.3. Details regarding the acquisition and projection of the images can be found in [82, 49]; barycentric subspaces on manifolds are described in [61].

The data consists of 10 cardiac cycles of patients with Tetralogy of Fallot and 9 patients from a control group. Each cardiac cycle is originally represented by three-dimensional homogeneous coordinates $x_1 : x_2 : x_3$, sampled at 30 time points. We project the homogeneous coordinates onto the plane $x_1 + x_2 + x_3 = 1$ and choose a two-dimensional coordinate system for this plane. Then we use spline interpolation with degree $n_\theta = 3$ and $N_\theta = 30$ control points to reconstruct the planar curves from the data points; see Fig. 3.15.

The parameters a_0, a_1, and a_2 in the metric are chosen similarly to Sect. 3.4; however, the scale of the curves is not changed and we use equal weighting between the L^2-, H^1- and H^2-parts of the average energy for linear paths. This leads to parameters $a_0 = 1$, $a_1 = 0.1$, and $a_2 = 10^{-9}$. To see if the metric structure derived from the Sobolev metric enables us to distinguish between diseased patients and the control group, we compute all 171 pairwise distances between the 19 curves; this takes about 15 minutes on a 2 GHz single core processor. Multi-dimensional scaling of the distance matrix shows that the metric separates healthy and diseased patients quite well (Fig. 3.16a). Indeed, a cluster analysis

based on the distance matrix recovers exactly – with exception of one outlier (patient 4) – the subgroups of healthy and diseased patients (Fig. 3.16b).

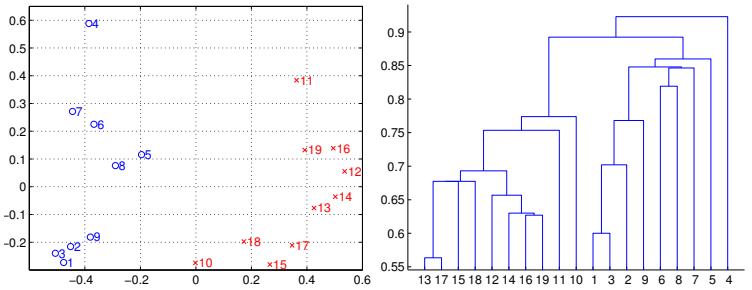


Figure 3.16: (a) Two dimensional representation of the data using multi dimensional scaling of the pairwise distance matrix. (b) A dendrogram of clusters computed from the pairwise distance matrix using the single linkage criterion. Healthy patients are labelled 1–9 and diseased ones 10–19.

The Karcher means of the healthy and diseased subgroups as well as of the entire population are depicted in Fig. 3.17. The mean was computed using a gradient descent method as described in Sect. 3.3.6 with a threshold of 10^{-4} for the norm of the gradient. The average distance from the mean for the diseased group is 0.6853 with a variance of 0.0149, and for the control group the distance is 0.7708 with a variance of 0.0083.

To investigate the variability of the observed data, we performe principal component analysis on the initial velocities of the minimal geodesics connecting curves to the respective means (c.f. Sect. 3.4). Fig. 3.18 shows the initial velocities projected to the subspace spanned by the first two principal directions. Within the healthy and sick subgroups, less then 30% of the principal components are needed to explain 90% of the shape variation. If, in contrast, principal components are analyzed for the entire dataset based on the global Karcher mean, then 35% of the principal components are needed to explain 90% of the shape variation.

3.5.2 The Kimia database

We also tested our implementation on a dataset of shapes collected by the Computer Vision Group at Brown university [18]. The dataset consists of black and white images of physical objects. It is natural to represent the objects by their boundaries using unparametrized curves. In addition to factoring out

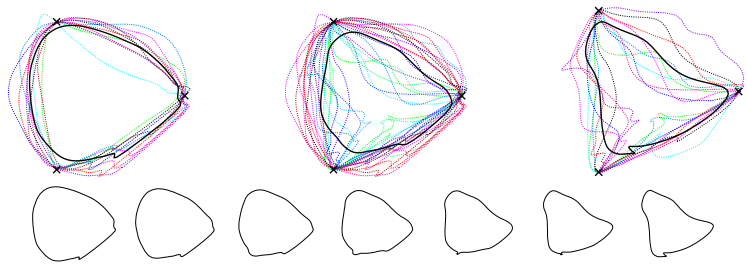


Figure 3.17: First row: Karcher means of pathological cardiac cycles (left), all cycles (middle), and healthy cycles (right). Second row: geodesic connecting the Karcher mean of pathological cycles to the Karcher mean of healthy cycles. The crosses denote the position of images, with respect to whom the barycentric projection was computed.

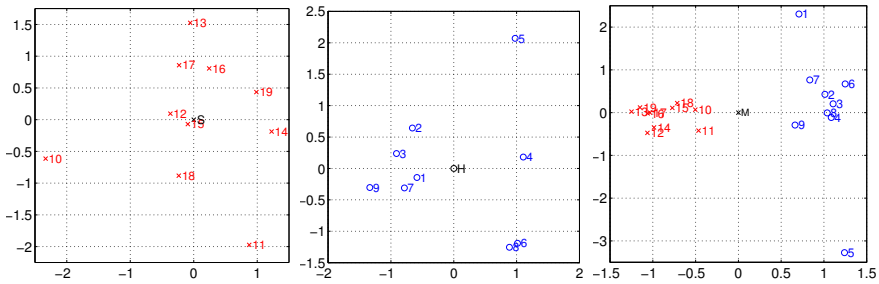


Figure 3.18: (a) Initial velocities of minimizing geodesics projected to the subspace spanned by first two principal components for the diseased group. (b) The same picture for the control group. (c) The same picture for the whole population.

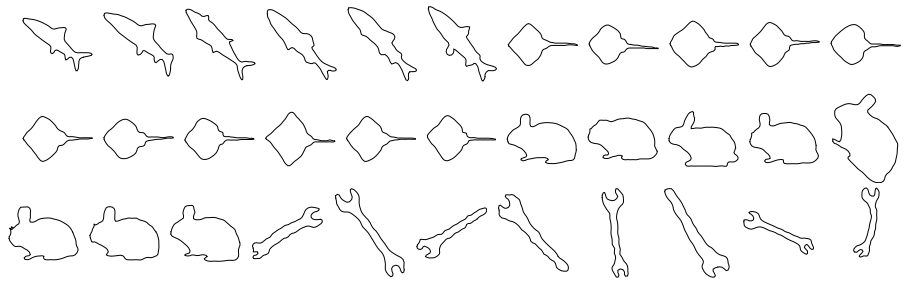


Figure 3.19: Selection of shapes from the dataset [18].

reparametrizations, we also factor out translations and rotations because we are not interested in the position of the curves in space. Some of the resulting curves are depicted in Fig. 3.19. We used splines of degree $n_\theta = n_\phi = 3$ with $N_\theta = 60$ and $N_\phi = 20$ controls in space and of degree $n_t = 2$ with $N_t = 20$ controls in time.

We used the following ad-hoc strategy for choosing the constants: we computed the average L^2-, H^1- and H^2-contributions $\overline{E}_0$, $\overline{E}_1$, $\overline{E}_2$ to the energy of linear paths between each pair of curves in the dataset. Then we normalized a_0 to 1 and chose a_1 and a_2 such that

$$a_0\overline{E}_0 : a_1\overline{E}_1 : a_2\overline{E}_2 = 1:1:1 \quad \text{and} \quad \overline{E} = a_0\overline{E}_0 + a_1\overline{E}_1 + a_2\overline{E}_2 = 100\,.$$

This resulted in the parameter values $a_1 = 250$ and $a_2 = 0.004$. It remains open how to best choose the constants depending on the data under consideration.

Comparison to non-elastic metrics

Riemannian metrics on spaces of unparametrized curves are often called elastic metrics, as they allow both for bending and stretching of the curve. In the elastic case, solving the boundary value problem for geodesics involves optimizing over the $\mathrm{Diff}(S^1)$-orbit of the initial or final shape. This is a computationally expensive and difficult task since the diffeomorphism group is infinite-dimensional.

An alternative and simpler approach is to parametrize the curves by unit-speed and to calculate geodesics in the space of parametrized curves. This could be called a non-elastic approach. (Of course, one might wish to also factor out rigid transformations and constant shifts of the parametrization, but this is much simpler because these groups are finite dimensional.)

Our experiments suggest that using the more involved elastic approach pays off. The two approaches yield different results, and in particular in cases involving large amounts of stretching the geodesics found using the first approach appear more natural. However, in cases involving mainly bending of the curves the results are very similar (see Fig. 3.20).

Clustering and principal component analysis

We investigated if pairwise geodesic distances can be used to cluster shapes into meaningful groups. To this aim, we calculated all pairwise distances between the 33 shapes presented in Fig. 3.19. Solving the corresponding 528 boundary value problems took about two hours on a 3GHz processor with four cores. The

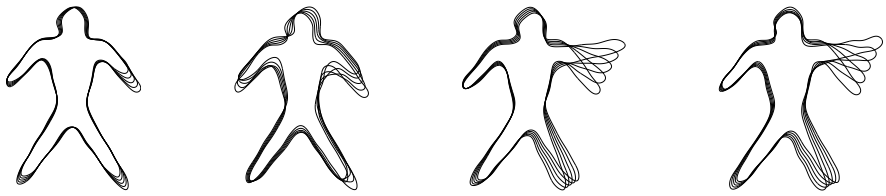


Figure 3.20: Geodesics in the space of unparametrized (1st, 3rd) versus parametrized (2nd, 4th) curves modulo rotations and translations. Note that since we also optimize over translations and rotations of the target curve, the curves in the first two panels are aligned differently.

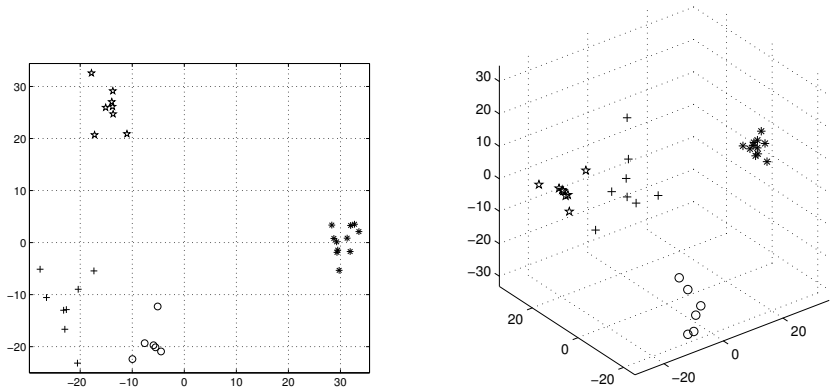


Figure 3.21: The matrix of geodesic distances between the shapes in Fig. 3.19, visualized using multi-dimensional scaling in two and three dimensions. The labels are: fish ○, sting rays ☆, bunnies ∗, tools +.

resulting distance matrix is visualized in Fig. 3.21 using multi-dimensional scaling; the plot suggests that objects of the same group lie close together. Indeed, agglomerative clustering with 4 clusters reproduces exactly the subgroups of the database.

Finally, we studied within-group variations using non-linear principal component analysis. To this aim, we first computed the Karcher mean of each group. The corresponding optimization problem (3.20) was solved using a conjugate gradient method, as implemented in the Manopt library [21], on the finite-dimensional spline approximation of the Riemannian manifold of curves. The mean shapes for the groups of fish and humans can be seen in Fig. 3.22. Next, we represented each shape in the group by the initial velocity from the mean $\bar{c}$ using the inverse

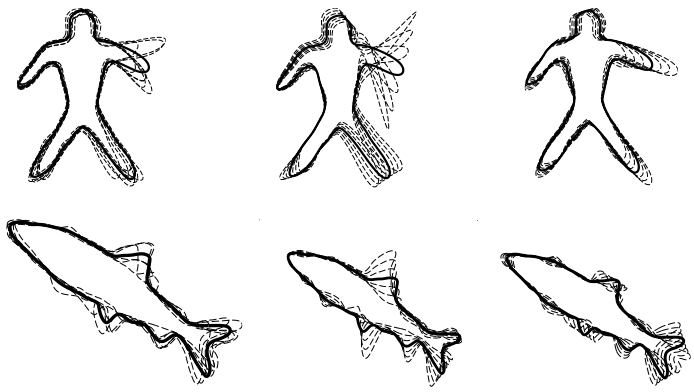


Figure 3.22: First column: Karcher means (bold) of the groups of fish and humans. Second and third column: geodesics from the mean in the first and second principal direction at the times $-3, -2, \ldots, 2, 3$; the bold curve is the mean.

of the Riemannian exponential map. We then performed a principal component analysis with respect to the inner product $G_{\bar{c}}$ on the set of initial velocities. In the group of human figures, the first three eigenvalues capture 67%, 22%, and 6% of within-group variation. In the group of fish, the first three eigenvalues capture only 40%, 25%, and 16% of within-group variation. Geodesics from the mean in the directions of the first two principal directions can be seen in Fig. 3.22. In the group of humans the first principal direction encodes bending of the arms and legs, whereas the second direction reflects stretching in the extremities.

Chapter 4

Shape Optimization

Shape optimization is a broad topic covering many topics. In broad terms it is the topic concerned with find shapes that optimizes a functional. The functional in question can describe many things, and there are a wide range of applications coming from physics, image analysis, computer graphics, etc. Our motivation will come from functionals arising in physics. Here one wants to find physical shapes that are optimal in some contexts. A classical problem comes from aerodynamic shape optimization, where one tries to find the shape of an airplane which minimizes drag, while maintaining a necessary lift. These problems are usually constrained by some physics described by a Partial Differential Equation (PDE), for aircraft the equations governing the dynamics are the Navier-Stokes equations for the interaction of air and craft. When trying to solve problem in practice, there are two paradigms, which can roughly be summarized as follows.

- *Discretize-first*: Discretize functional and state equation to obtain finite dimensional system. Then use classical optimization algorithms.
- *Optimize-first*: Write up optimality conditions in an appropriate functional space. Then discretize the resulting equations.

In concrete optimization problems one parametrizes the feasible shapes of the problem by their boundary. A simple, but important observation is that any

solution of such a problem is not dependent of the choice parametrization. This invariance is exactly the equivalence of curves under reparametrization studied in Chapter 2 for shape spaces. In this way, it is natural to consider shape functionals defined on planar objects, as a function on the manifold of regular curves or shape space. This point of view have only been considered very recently in a series of papers, starting with [69]. In this way the non-linear structure of the feasible shapes are captured, and we can hopefully use the structure to gain more information about the optimization procedure. The choice of a Riemannian metric on immersion allow us to use generalized versions of classical optimization algorithms like steepest-descent and BFGS, on a way that naturally turns the regularity constraint into an unconstrained problem. To make matters concrete we focus our efforts on a simple model problem:

Example 4.1 (Model Problem) *Let $f, y_0 \in H^1(\mathbb{R}^2, \mathbb{R})$ be fixed.*

$$\min_{\Omega} J(u, \Omega) = \int_{\Omega} \frac{1}{2} |y(x) - y_0(x)|^2 dx \tag{4.1}$$

$$-\Delta y = f \quad \text{in } \Omega, \tag{4.2}$$

$$y = 0 \quad \text{on } \Gamma. \tag{4.3}$$

Over all Ω, bounded by a simple curve.

We will derive procedures to solve this specific problem numerically, but this methodology can in principle be extended to any type of shape optimization problem which uses parametrized boundaries.

4.1 Shape calculus

Shape functionals are often not defined on linear vector spaces or manifolds, and so there is an extensive theory devoted to giving meaning to and computing directional derivatives of such shape functionals. In this section we will define the *shape derivative* of a functional depending on a domain, and compute it for our concrete problem in Example 4.1. Let $\Omega \subset \mathbb{R}^2$ be a domain, and $V(x,t)$ be a sufficiently smooth vector field, with solution map $T_t(x) = T(t,x)$,

$$\frac{dT(t,x)}{dt} = V(T(t,x)), \quad T(0,x) = x.$$

Define $\Omega_t = T(t,\Omega)$. The *Eulerian derivative* of J is defined as

$$dJ(\Omega, V) = \lim_{t \to 0} \frac{J(\Omega_t) - J(\Omega)}{t}$$

If the map $V \mapsto dJ(\Omega, V)$ is continuous we will say that $dJ(\Omega, V)$ is the *shape derivative* of the shape functional J in the direction of V.

For our example problem let $y(\Omega)$ denote the solution of the Poisson equation with Dirichlet conditions

$$\begin{aligned} -\Delta y &= f \text{ in } \Omega, \\ y &= 0 \text{ on } \Gamma. \end{aligned} \tag{4.4}$$

with $f \in H^1(\mathbb{R}^2)$ fixed. We define the shape function

$$J(\Omega) = \frac{1}{2}\int_\Omega |y(\Omega) - y_d|^2 dx,$$

for a fixed $y_d \in H^1(\mathbb{R}^2)$. The solution of (4.4) is also the solution of the weak form, which is given by

$$E(\Omega, y, \varphi) = \int_\Omega (\langle \nabla y, \nabla \varphi \rangle - f\varphi) dx = 0, \quad \forall \varphi \in H_0^1(\Omega). \tag{4.5}$$

This is called the *state equation* of the problem. The objective function is defined as

$$F(\Omega, \varphi) = \frac{1}{2}\int_\Omega |\varphi - y_d|^2 dx,$$

so we may write

$$J(\Omega) = F(\Omega, y(\Omega)).$$

Now we will compute an expression for the shape derivative of $J(\Omega)$. One approach is to try and invoke an implicit function theorem. The problem is that $y(\Omega)$ lives in different Sobolev spaces as we vary Ω, so first we would have to give meaning to $y'(\Omega)$, the derivative of the state w.r.t to a deformation of Ω, from which we could derive an expression like $dJ(\Omega, V) = dF_\varphi(\Omega, y(\Omega))y'(\Omega)$. This is technique pursued in [75]. We will use an alternative method, that avoids having to characterize a derivative $y'(\Omega)$, and which in general is more flexible that the implicit approach as it requires less smoothness of the domain. The first step is setting the problem up as an appropriate control theory type problem, by defining a Lagrangian functional with the state equation added with a Lagrange multiplier ψ

$$G(\Omega, \varphi, \psi) = F(\Omega, \varphi) + E(\Omega, \varphi, \psi). \tag{4.6}$$

Observing that

$$\sup_{\psi \in H_0^1(\Omega)} G(\Omega, \varphi, \psi) = \begin{cases} F(\Omega, y(\Omega)) & \psi = y(\Omega), \\ \infty & \psi \neq y(\Omega), \end{cases}$$

we can express the objective function as a saddle point of the Lagrangian

$$J(\Omega) = \inf_{\varphi \in H_0^1(\Omega)} \sup_{\psi \in H_0^1(\Omega)} G(\Omega, \varphi, \psi).$$

The Lagrangian is continuous w.r.t to both φ and ψ, and convex in φ, and defined on $H_0^1(\Omega)$ which is closed and convex. Under these conditions G has a saddle point at $(y,p) \in H_0^1(\Omega)^2$ if and only if the saddle point equations have a solution,

$$dG(\Omega,y,p)(0,\psi) = dE(\Omega,y,p)(0,\psi) = 0, \forall \psi \in H_0^1(\Omega),$$
$$dG(\Omega,y,p)(\varphi,0) = dF(\Omega,y)(\varphi) + dE(\Omega,y,p)(\varphi,0) = 0, \forall \varphi \in H_0^1(\Omega).$$

In this case E is linear in both arguments. Writing these equations out we get

$$\int_\Omega (\langle \nabla y, \nabla \psi \rangle - f\psi) dx = 0 \quad \forall \psi \in H_0^1(\Omega).$$
$$\int_\Omega ((y - y_d)\varphi + \langle \nabla p, \nabla \varphi \rangle) dx = 0 \quad \forall \varphi \in H_0^1(\Omega).$$

The first is the state equation, and the second is the *adjoint equation*, which in strong form reads

$$\begin{aligned} -\Delta p &= y - y_0 \text{ in } \Omega, \\ y &= 0 \text{ on } \Gamma. \end{aligned} \tag{4.7}$$

These equations have a unique solution, so G has a saddle point. Now given a vectorfield $V \in \mathcal{D}^1(\mathbb{R}^2, \mathbb{R}^2)$ (compactly supported vector fields) we want to calculate $dJ(\Omega, V)$. For t small enough such that the flow of V is defined we have the domains $\Omega_t = T_t(\Omega)$ on which we can evaluate J,

$$J(\Omega_t) = \inf_{\varphi \in H_0^1(\Omega_t)} \sup_{\psi \in H_0^1(\Omega_t)} G(\Omega_t, \varphi, \psi).$$

Notice that we compute the saddle point of G in different spaces $H_0^1(\Omega_t)$ for each t. We want to remove this time dependence. There are at least two methods for computing the derivative $\partial_t|_{t=0} J(\Omega_t)$ as $t \to 0$.

- The *function space parametrization*
- The *function space embedding.*

In the first case we can bring all the analysis back to Ω by the mapping

$$\phi \mapsto \phi \circ T_t^{-1} : H_0^1(\Omega) \to H_0^1(\Omega_t),$$

and defining a Lagrangian on Ω alone

$$\tilde{G}(t, \varphi, \psi) = G(T_t(\Omega), \varphi \circ T_t^{-1}, \psi \circ T_t^{-1}).$$

Now one can find a expression for derivative of $g(t) = J(\Omega_t)$

$$dg(0) = \lim_{t \to 0} \frac{g(t) - g(0)}{t}.$$

For this to work we need to know how to differentiate a saddle point w.r.t. to a paramater t. In our case we know that for each t there is a unique saddle point, but in a more general situation there could be several. Will need this freedom in a little while. We state a theorem on the existence of the above limit under some conditions. We need to define some notation. Consider the general type of functional

$$G : [0, \tau] \times X \times Y \to \mathbb{R}$$

For some $\tau > 0$ and sets X, Y. Define

$$g(t) = \inf_{x \in X} \sup_{y \in Y} G(t, x, y),$$
$$h(t) = \sup_{y \in Y} \inf_{x \in X} G(t, x, y)$$

and the sets

$$X(t) = \{x \in X : \sup_{y \in Y} G(t, x, y) = g(t), Y(t) = \{y \in Y : \inf_{x \in X} G(t, x, y) = h(t).$$

Then we have $h(t) \leq g(t)$, and the of saddle points is defined as

$$S(t) = \{(x, y) \in X \times Y : h(t) = G(t, x, y) = g(t)\}.$$

The theorem gives a condition under which $dg(0)$ exists and how to compute it.

Theorem 4.1 (Correa and Seeger) *Let $\tau > 0$ and set X, Y and a functional*

$$G : [0, \tau] \times X \times Y$$

be given. Under the following assumptions

- *For each t, there exists at least one saddle point: $S(t) \neq \varnothing$.*
- *For all t, $\partial_t G$ exists for all elements in $X(t) \times Y(0)$ and $X(0) \times Y(t)$.*
- *For every sequence (t_n) with $t_n \searrow 0$, there exists a subsequence (t_{n_k}) and an element $x_0 \in X(0), x_{n_k} \in X(t_{n_k})$ such that for all $y \in Y(0)$*

$$\lim_{\substack{k \to \infty \\ t \searrow 0}} \partial_t G(t, u_{n_k}, p) = \partial_t G(0, x_0, y)$$

- *For every sequence* (t_n) *with* $t_n \searrow 0$*, there exists a subsequence* (t_{n_k}) *and an element* $y_0 \in Y(0), y_{n_k} \in Y(t_{n_k})$ *such that for all* $x \in X(0)$

$$\lim_{\substack{k\to\infty \\ t\searrow 0}} \partial_t G(t, x, y_{n_k}) = \partial_t G(0, x, y_0)$$

Then there exists a saddle point (x^0, y^0) *such that*

$$\begin{aligned} dg(0) &= \inf_{x\in X(0)} \sup_{y\in Y(0)} \partial_t G(0, x, y) \\ &= \sup_{y\in Y(0)} \inf_{x\in X(0)} \partial_t G(0, x, y) \\ &= \partial_t G(0, x^0, y^0) \end{aligned} \tag{4.8}$$

In this way we can compute $\partial_t G(0, x, y)$ and then look for elements in $X(0)$ and $Y(0)$ which satisfy the saddle point condition. In our case this is very simple since there is only one. In cite [88] it is proved that the assumptions are verified for our model problem. We have $g(t) = J(\Omega_t)$ and computing $\partial_t G(t, \varphi, \psi)$ can be done explicitly by pulling all integrals back to Ω and differentiating the resulting time dependent integrand. We will leave the details out, and just state the final form of the shape derivative

$$\begin{aligned} dJ(\Omega, V) &= dg(0) \\ &= \int_\Omega \frac{1}{2}(y - y_d)^2 \operatorname{div} V - (y - y_d)\langle \nabla y_d, V\rangle \\ &\quad + (\operatorname{div} V I - DV^T - DV)\langle \nabla y, \nabla p\rangle \\ &\quad + \operatorname{div} V \cdot (yp - fp) - \langle \nabla f, V\rangle \cdot p \, \mathrm{d}x \end{aligned} \tag{4.9}$$

Where DV is the Jacobian of V. We can derive a simpler expression, by simpler calculations by using the function space embedding, which is similar in its use of the same theorem, as we shall see. This comes at the cost of added smoothness requirements of Ω, we will comment on this at the end. With this method, we extend all functions to a larger domain D which contain all Ω_t, for simplicity we choose $D = \mathbb{R}^2$. This can be done, since there exists extension operators

$$E_\Omega : H^m(\Omega) \to H^m(\mathbb{R}^2),$$

Consider the functional defined on this larger function space

$$J(\Omega_t) = \min_{\Phi\in H^1(\mathbb{R}^2)} \max_{\Psi\in H^1(\mathbb{R}^2)} G(\Omega_t, \Phi, \Psi).$$

Now the set of saddle points $S(t)$ is not empty since

$$\begin{aligned} X(t) &= \{\Phi \in H^1(\mathbb{R}^2) : \Phi|_{\Omega_t} = y(\Omega_t)\} \\ Y(t) &= \{\Psi \in H^1(\mathbb{R}^2) : \Psi|_{\Omega_t} = p(\Omega_t)\} \end{aligned}$$

are the sets of all extensions $\mathbb{R}^2$ of the solutions to the state and adjoint equations on Ω_t, which are clearly not empty. We can still apply the theorem of Correa and Seeger: G is the same but extended,

$$G(t,\Phi,\Psi) = \int_{\Omega_t} \frac{1}{2}|\Phi - y_d|^2 + \langle \nabla\Phi, \nabla\Psi\rangle - f\Psi \,\mathrm{d}x,$$

and computing $\partial_t G$ is simple since Φ and Ψ are now fixed function on $\mathbb{R}^2$ so have no dependence on t. Contrast this to $\tilde{G}(t,\varphi,\phi)$ where the dependence on t is in all the variables. The only time dependence is on Ω_t so the formula for differentiating along the moving domains becomes very simple

$$\partial_t G(t,\Phi,\Psi) = \int_{\Gamma_t} \left(\frac{1}{2}|\Phi - y_d|^2 + \langle \nabla\Phi, \nabla\Psi\rangle - f\Psi\right) \langle V, n_t\rangle \,\mathrm{d}x.$$

This doesn't depend on the values of Φ and Ψ outside of Ω_t, so the derivative (4.8) can be evaluated at any extension. Now $\Phi = 0$ on Γ so $\nabla\Phi = \frac{\partial p}{\partial n} n$ on Γ, and likewise for Ψ, and we end up with

$$dJ(\Omega, V) = \int_\Gamma \left\{\frac{1}{2}|y_d|^2 + \frac{\partial y}{\partial n}\frac{\partial p}{\partial n}\right\} \langle V, n\rangle ds. \tag{4.10}$$

where (y,p) satisfies the state and adjoint equations

$$-\Delta y = f \text{ in } \Omega,\ y = 0 \text{ on } \Gamma.$$
$$-\Delta p = y - y_d \text{ in } \Omega,\ p = 0 \text{ on } \Gamma.$$

This confirms that the derivative only depends on the deformation of the boundary of Ω. All of this is only the formal computations. We note that one can make this formal. We sum up the result in

Lemma 4.2 *If Ω_t is a $C^{1,1}$ for all t, then the assumptions of 4.1 holds, hence the limit $dg(0)$ exists.*

Proof. In [88] it is proved that the assumptions of the Correa-Seeger theorem holds if the solution $y(\Omega_t) \in H^2(\Omega_t)$ for each t and the limit $dg(0)$ exists. Classical regularity of elliptic equations imply that a $C^{1,1}$ domain has a $H^2(\Omega)$ regular solution, see [35].

4.2 Riemannian optimization

Riemannian optimization is an area of optimization theory concerned with creating and analysing optimization algorithms on Riemannian manifolds. The

motivation comes from the fact that some types of constrained minimization problems can naturally be considered as problems on manifolds, e.g. submanifolds on $\mathbb{R}^n$ like the sphere, or quotient manifolds. A great deal of the research has gone into producing efficient algorithms on matrix groups. Using the gradient $\operatorname{grad} f$ to obtain descent directions and the exponential map $\exp_x(\xi)$ to move around the manifold, one can generalize steepest descent, conjugate gradients, BFGS methods etc. from their counterpart in $\mathbb{R}^n$. An introduction can be found in [1], where the theory is presented exclusively on finite dimensional manifolds, especially matrix groups. As we are concerned with infinite dimensional manifolds, the results in this reference are not readily applicable. In vectors spaces there are ways to generalize optimization methods to infinite dimensional Hilbert spaces, see [48, 29]. In [65] Riemannian BFGS methods and conjugate gradients on possibly infinite dimensional manifolds are treated. As we have seen in Chap. 2 there are problems when generalizing geometry to infinite dimensions, and it is assumed implicitly in this reference that the metric is strong. In this section we shall give a brief introduction to Riemannian optimization methods, to the extend that we can apply the techniques in for shape optimization in Sec. 4.5. We will briefly mention convergence analysis of the methods.

4.2.1 Line-search methods

In the following let (M, g) be a Riemannian manifold, and $f : M \to \mathbb{R}$ be a sufficiently smooth function. Recall the definition of a gradient

$$g_x(\operatorname{grad} f, v) = df_x(v), \quad \forall v \in T_x M. \tag{4.11}$$

Where $df_x(v) = v_x(f)$ is the differential of f. It follows that $\operatorname{grad} f$ is g-orthogonal to the level sets of f, as in the euclidean case, but with the notion of orthogonality depending on the choice of g. Now a line-search method on (M, g) is the construction of a sequence $\{x_k\} \in M$ which converges to a critical point x^* of f, i.e. $df_{x^*} = 0$. Notice that the notion of a critical point is independent of the choice of metric. On $\mathbb{R}^n$ a line-search method is generated by the update

$$x_{k+1} = x_k + \alpha_k p_k,$$

where t_k is the step size and η_k the search direction. We could possibly have η_k as $\operatorname{grad} f$. This is update not defined on a general manifold, when x_k is a point in M and η_k a direction in the tangent space $T_{x_k} M$. In order do define this procedure on M we need to introduce a way to move in the direction of η_k while staying in M, which we define as,

Definition 4.3 (Retraction) A *retraction* on a manifold M is a smooth

map $R : TM \to M$ with the following properties. Let $R_x : T_xM \to M$ be the restriction to T_xM.

1. $R_x(0_x) = x$.
2. $DR_x(0_x) = \mathrm{id}_{T_xM}$.

Here we have used the identification $T_{0_x}T_xM \simeq T_xM$.

On any (M, g) the exponential map $\exp_x(v)$ defines a retraction. Notice that condition 2 on a Hilbert manifold makes R_x a local diffeomorphism. The reason for not restricting to geodesic retractions is that it can be computationally expensive to compute the exponential map, and that it does not change the qualitative convergence of the method. In this way, if a simple retraction is available it can produce a great speed up of the method. Much of the research on matrix groups is devoted to finding efficient retractions. Now the update formula looks like

$$x_{k+1} = R_{x_k}(\alpha p_k),$$

where $\eta_k \in T_{x_k}M$, and $\alpha > 0$. In Alg. 1, the procedure is given as a general procedure.

Algorithm 1 Riemannian Line-Search

Input: $f : M \to \mathbb{R}$, $x_0 \in M$, $k = 0$.
for $k = 1, 2, \ldots$ **do**
  Choose descent direction $p_k \in T_{x_k}M$.
  Choose a Retraction $R_{x_k} : T_{x_k}M \to M$.
  Choose a step length $\alpha_k \in \mathbb{R}$.
  Set $x_{k+1} = R_{x_k}(\alpha_k p_k)$.
  if $f(x_{k+1})$ satisfies stop criteria. **then**
    return x_{k+1}
  end if
end for

When choosing the step-length α_k a simple backtracking line-search method find the Armijo point:

Definition 4.4 (Armijo Point) Given a point $x \in M$, $\eta \in T_xM$, $\gamma > 0$, $\beta, \sigma \in (0, 1)$. The Armijo point is $p^A = \alpha^A p = \beta^m \gamma p$, where m is the smallest nonnegative integer such that

$$f(x) - f(R_x(\beta^m \gamma p)) \geq -\sigma \langle \mathrm{grad}\, f(x), \beta^m \gamma p \rangle$$

The acceptance condition correspond to the first part of the classical Wolfe conditions in $\mathbb{R}^n$ and is used to show that such a line-search methods produces a convergent algorithm.

4.2.2 The Quasi-Newton BFGS method

The usual way to improve the convergence rate, is to take higher order derivatives into account. The classical Newton method does this by picking the step p_k as the solution of

$$\operatorname{Hess} f(p_k, v) = -\langle \operatorname{grad} f(x_k), v\rangle, \ \forall v \in T_{x_k}M,$$

which under certain assumptions on f gives a quadratic convergence to a critical point. On a Riemannian manifold the definition of $\operatorname{Hess} f$ given by

$$\operatorname{Hess} f(X, Y) = X(Yf) - (\nabla_X Y)f,$$

In practice for many problems, it is unreasonable to calculate second order derivatives, especially the Riemannian Hessian which also needs the covariant derivative to be computed. Instead in Quasi-Newton method one solves a modified equation

$$B_k(p_k, v) = -\langle \operatorname{grad} f(x_k), v\rangle, \tag{4.12}$$

which gives a descent direction as long as B_k is a positive definite bilinear form. The idea is now to generate a sequence of operators B_k that should approximate the Hessian of f. Since the Hessian can be interpreted as a rate of change of the gradient, if we can compare gradients along of sequence of iterates, we can use this information to generate B_k as an approximation of $\operatorname{Hess} f$. To formalize this, we need a way to compare gradients from different tangent spaces on M. On a Riemannian manifold, the parallel transport can be used to transport vectors between tangent spaces. We want to be a bit more flexible, and a general notion of vector transport can be defined as follows

Definition 4.5 (Vector Transport) A vector transport on a manifold M is smooth mapping

$$TM \otimes TM \to TM : (\eta_x, \xi_x) \to T_{\eta_x}(\xi_x),$$

satisfying the following properties

1. (Associated Retraction) There exists a retraction R such that the following diagram commutes

$$\begin{array}{ccc} (\eta_x, \xi_x) & \xrightarrow{T} & T_{\eta_x}(\xi_x) \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\ \eta_x & \xrightarrow{R} & R(\eta_x) \end{array}$$

2. (Consistency) $T_{0_x}\xi_x = \xi_x$
3. (Linearity) $T_{\eta_X}(a\xi_x + b\zeta_x) = aT_{\eta_z}(\xi_x) + bT_{\eta_x}(\zeta_x)$

On any Riemannian manifold, the choice of parallel transport along geodesics is a vector transport. Since parallel transport can be a computationally expensive task, the above definition allows us to choose more efficient ways to compare tangent spaces. If we impose a generalized secant condition on B_k

$$B_{k+1}(T_k \operatorname{grad} f(x_k), v) = g_{x_{k+1}}(\operatorname{grad} f(x_{k+1}) - T_{\eta_k}(\operatorname{grad} f(x_k)), v)$$

with $T_k : T_{x_k}M \to T_{x_{k+1}}M$ being the vector transport, one can derive a generalization of the classical BFGS methods, where B_{k+1} is updated according to the formula

$$\begin{aligned} s_k &= \alpha_k p_k = R_{x_k}^{-1}(x_{k+1}) \in T_{x_k}M, \\ y_k &= \operatorname{grad} f(x_{k+1}) - T_{\eta_k}(\operatorname{grad} f(x_k)) \in T_{x_{k+1}}M., \\ B_{k+1}(T_k v, T_k w) &= B_k(v,w) - \frac{B_k(s_k,v)B_k(s_k,w)}{B_k(s_k,s_k)} \\ &\quad + \frac{g_{x_{k+1}}(y_k, T_k v) g_{x_{k+1}}(y_k, T_k w)}{g_{x_{k+1}}(y_k, T_k s_k)}, \ \forall v, w \in T_{x_k}M. \end{aligned} \tag{4.13}$$

In [65] the BFGS method, along with a Fletcher-Reeves Nonlinear Conjugate Gradient method, is analyzed on strong Riemannian manifolds and superlinear convergence is proved under a set of conditions. The pullback function $f \circ R$ is required to be locally convex, and the initial guess B_0 is symmetric, bounded and coercive. The most important condition is the vector transport need to be an isometry of tangent spaces, and satisfy a rather restrictive bound on a nuclear norm which severely restricts the geometry of the underlying manifold. Under these conditions the BFGS method converges super-linearly to an optimum x^*, i.e.

$$\lim_{k\to\infty} \frac{\operatorname{dist}(x_{k+1}, x^*)}{\operatorname{dist}(x_k, x^*)} = 0.$$

In practice, one wants to avoid solving (4.12) at every iteration. In finite dimensions B_k can be represented by a matrix, and there is an update formula for the inverse $H_k = B_k^{-1}$, and the system can be solved by a matrix multiplication. On a strong infinite dimensional Riemannian manifold, consider the Riesz representative of the operator $\hat{B}_k : T_xM \to T_xM$

$$B_k(u,v) = g_{x_k}(u, \hat{B}_k v)$$

Then one can similarly define an update for the inverse $H_k = \tilde{B}_k^{-1}$ using the Sherman-Morrison formula

$$H_{k+1} = T_k \circ \left(J^* H_k J + \frac{g_{x_k}(s_k, \cdot) s_k}{g_{x_{k+1}}(y_k, T_k s_k)} \right) \circ T_k^*, \tag{4.14}$$
$$J = \left(\mathrm{Id} - \frac{g_{x_k}(s_k, \cdot)}{g_{x_{k+1}}(y_k, T_k s_k)} y_k \right)$$

Inserting the expression for H_{k-1} etc., gives a recursive formula for H_{k+1}. This is utilized in the *Limited-Memory BFGS* method, to represent H_{k+1} by a limited number of vector pairs, which is very efficient for large scale problems. In the end the number of variables in our problem wont be too big, so for simplicity we will use the update (4.14).

4.3 Isogeometric analysis

To evaluate the functional and shape derivative in our model problem we need to compute the solution of PDE: the state and adjoint equations, both are simple Poisson equations with Dirichlet conditions. A standard way to solve this is to use a Finite Element Method (FEM). Looking ahead to section 4.5, we want to combine the numerics for Sobolev metrics on curves with Riemannian optimization methods, hence parametrize the boundary of our domains by B-splines. If using FEM, we would have to triangulate our curves and domain every time we need to solve the state and adjoints equations. This is time-consuming and we lose some control over the error in the FEM method, since we wont have a conforming FEM (the boundary of the domain is not a polygon). Instead we opt to use Isogeometric Analysis (IGA), which was recently proposed in [38], which has the advantage that the geometry of the physical domain can be given by exactly the same B-splines that we use for numerics for Sobolev metrics. In this section we will give a brief introduction to the essentials of IGA, see [17] for an overview papers.

Like FEM, IGA formulate a Galerkin projection which can be used to solve any type of PDE in a weak formulation. For our needs a Poisson equation with Dirichlet conditions is enough, so we will restrict our focus to that. The classical FEM method is summed up in figure 4.1: A physical domain Ω is triangulated, and basis for a finite dimensional subspace V_h of $H_0^1(\Omega)$ is constructed as piece-wise linear functions, or higher order polynomials in the *hp*-FEM method, with support on each triangle. The solution u_h of the weak equation restricted to V_h is an approximation of the true solution u. For elliptic problems this can be shown to provide a projection of the true solution to V_h, and the error of

the method can be related to the approximation power of V_h. We thank Jens Gravesen for providing all figures in this section.

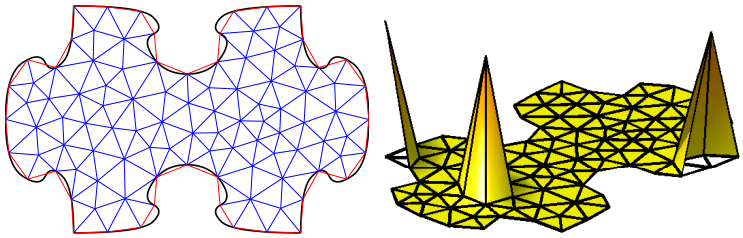


Figure 4.1: Trianglulation of domain Ω, and 3 basis functions for the FE space.

In IGA the starting point is not a triangulation of the domain Ω, but a parametrization from a simple fixed domain $\mathbf{x} : [0,1]^2 \to \Omega$. We use B-splines to define a parametrisation,i.e.

$$\mathbf{x}(\xi,\eta) = \sum_{i=1}^{m_\xi}\sum_{j=1}^{m_\eta} \mathbf{x}_{i,j}\, B^{n_\xi}_{\boldsymbol{\xi};i}(\xi)\, B^{n_\eta}_{\boldsymbol{\eta};j}(\eta) := \sum_{i,j} \mathbf{x}_{i,j} B_{i,j}(\xi,\eta).$$

such that $\mathbf{x}^{-1}$ exists everywhere and is differentiable. Here $\mathbf{x}_{i,j} \in \mathbb{R}^2$ are vectors defining a tensor product spline in each coordinate. The functions $B^{n_\xi}_{\boldsymbol{\xi},i}$ and $B^{n_\eta}_{\boldsymbol{\eta},j}$ are B-splines of degree n_ξ and n_η defined on knots vectors $\boldsymbol{\xi} = (\xi_1,\ldots,\xi_{m_\xi})$, $\boldsymbol{\eta} = (\eta_1,\ldots,\eta_{m_\eta})$. For simplicity we will assume full multiplicity at the boundary. This allows us to specify the boundary of the domain only by the outer most control points: $\mathbf{x}_{1,j}, \mathbf{x}_{m_\xi,j}, \mathbf{x}_{i,1}, \mathbf{x}_{i,m_\eta}$. They define 4 univariate splines that make up the boundary. Also we will assume simple interior knots, so the parametrization is $C^{n_\xi-1,n_\eta-1}([0,1]^2)$ regular. An example of a parametrization is given in Fig. 4.2. When solving a PDE, we consider the weak form of the equation. For simplicity we assume Dirichlet boundary conditions,

$$a(u,\varphi) = b(\varphi), \quad \forall \varphi \in H^1_0(\Omega).$$

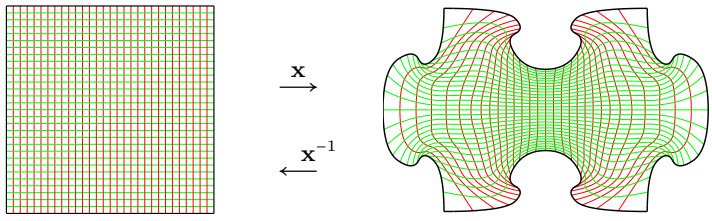


Figure 4.2: Parametrization of Ω

For the Poisson equation

$$\begin{aligned} -\Delta u &= f, \quad \text{in } \Omega, \\ u &= 0, \quad \text{on } \Gamma, \end{aligned}$$

we have the weak form

$$a(u,\varphi) = \int_\Omega \langle \nabla u, \nabla \varphi \rangle \, \mathrm{d}x, \quad b(\varphi) = \int_\Omega f\varphi \, \mathrm{d}x.$$

Classical regularity theory tells us that the problem on Ω is regular if the interior angles of the corners of the domain is less that π, otherwise the solution might have singularities of a certain type, for more information see [35]. The problem has a solution if the angles between The idea is now to pull the solution back to I^2 using the parametrization $\mathbf{x}$ and construct a finite dimensional subspace $V_h \subseteq H_0^1(I^2)$ to approximate the solution of the weak equation in, as in standard Galerkin methods. To this end we consider the set B-splines on the parameter domain $[0,1]^2$

$$N_{k,\ell} : (\xi,\eta) \mapsto N^p_{\boldsymbol{\xi};k}(\xi)\, N^q_{\boldsymbol{\eta};\ell}(\eta),$$

Where the right hand side are B-splines of degree p and q defined on the same knot vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ as $\mathbf{x}$. It is also possible to use refined knot vectors to improve the numerical stability. Using the parametrization we can construct a set of basis function in Ω

$$M_{k,\ell} : (x,y) \mapsto N_{k,\ell}\left(\mathbf{x}^{-1}(x,y)\right).$$

And we will look for a solution u_h in

$$V_h = \operatorname{span}\{M_{i,j}\}, i = 1 \dots m_\xi, j = 1 \dots m_\eta.$$

Which satisfy

$$a(u_h, v_h) = b(v_h), \quad \forall v_h \in V_h.$$

Here is h is a parameter specifying the *mesh size* of the B-splines solution space V_h. To be explicit, our solution u_h is of the form

$$u_h(x,y) = \sum_{i=1}^{m_\xi}\sum_{j=1}^{m_\eta} u_{i,j} M_{i,j}(x,y) = \sum_{k=1}^{m_\xi m_\eta} u_k M_k(x,y)$$

Where we have made a reordering of coefficients, $k = i + m(j+1)$. We assumed full multiplicity for the boundary knots in $\boldsymbol{\xi}, \boldsymbol{\eta}$, ensuring that we can enforce the Diriclet condition by setting the boundary control points of $u_{i,j}$ to zero, thus ignoring the boundary B-splines. $M_{i,j}$ is of the same smoothness as $N_{i,j}$.

See Fig. 4.3 for a visualization of the basis functions $M_{k,\ell}$. Using the chain rule

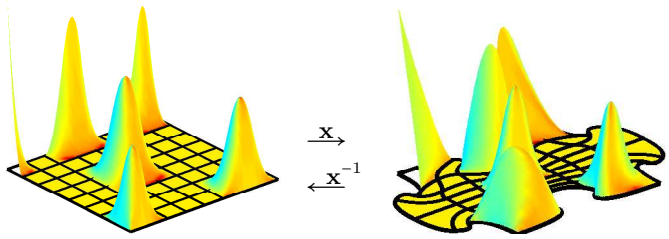


Figure 4.3: Left: $N_{i,j}$ basis. Right: $M_{i,j}$ basis

$$\begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \end{pmatrix} = \begin{pmatrix} \frac{\partial \xi}{\partial x} & \frac{\partial \eta}{\partial x} \\ \frac{\partial \xi}{\partial y} & \frac{\partial \eta}{\partial y} \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial \xi} \\ \frac{\partial}{\partial \eta} \end{pmatrix} = \begin{pmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial}{\partial \xi} \\ \frac{\partial}{\partial \eta}. \end{pmatrix}$$

That is

$$\nabla M = J^{-T} \nabla N,$$

where J is the Jacobian of $\mathbf{x}$, and J^{-T} is the inverse transposed. For the Poisson equation, this gives is the weak pulled back equation with $u = \hat{u} \circ \mathbf{x}$ and $\varphi = \hat{\varphi} \circ \mathbf{x}$,

$$\int_0^1 \int_0^1 \langle J^{-T} \nabla u, J^{-T} \nabla \varphi \rangle \det J \, \mathrm{d}\xi \, \mathrm{d}\eta = \int_0^1 \int_0^1 f \varphi \det J \, \mathrm{d}\xi \, \mathrm{d}\eta$$

For $u_h \in V_h$ and letting v_h vary of all M_k, we get the linear system

$$A[u] = b$$

with $[u]$ being the vector of control points of u and

$$A_{ij} = \int_0^1 \int_0^1 \langle J^{-T} \nabla N_i(\xi, \eta), J^{-T} \nabla N_j(\xi, \eta) \rangle \det J \, \mathrm{d}\xi \, \mathrm{d}\eta,$$
$$b_i = \int_0^1 \int_0^1 f(\mathbf{x}(\xi, \eta)) \, N_i(\xi, \eta) \det J \, \mathrm{d}\xi \, \mathrm{d}\eta$$

In practice we will evaluate the integrals using a Gaussian quadrature scheme. One can obtain a similar error analysis as for standard Galerkin methods, we just mention the results

Theorem 4.6 *There exists a unique solution solution u_h, and the approximation power to the unique solution u of (4.3) satisfy the classical Cea type lemma,*

$$\|u - u_h\|_{H^1(\Omega)} \leq \frac{M}{\alpha} \inf_{v_h \in V_h} \|u - v_h\|_{H^1(\Omega)} \tag{4.15}$$

Where M and α are the continuity and coercivity constants of a, i.e.

$$|a(u, v)| \leq M \|u\| \|v\|$$
$$a(v, v) \geq \alpha \|v\|$$

The right hand side is the approximation error of the subspace V_h. If we define the mesh of the parametrization as

$$Q_{i,j} = (\xi_i, \xi_{i+1}) \times (\eta_j, \eta_{j+1})$$

and $h = \max_{i,j} \operatorname{diam} \mathbf{x}(Q_{i,j})$. Then the approximation power of V_h is given by

Theorem 4.7 *Assuming the degree $p = q$, There exists an interpolation operator $i_h : H^s(\Omega) \to V_h$, and a constant such that for all $u \in H^s(\Omega)$ it holds that*

$$\|u - i_h(u)\|_{H^s(\Omega)} \leq C h^{p-s} \|u\|_{H^s(\Omega)}$$

See [17].

4.3.1 The Parametrization problem

For a general domain it is quite difficult to determine coefficients $\mathbf{x}_{i,j}$ such that the map $\mathbf{x}$ defines a valid parametrization. In general we only have the boundary curve given, and then we want to extend this to a parametrization of the whole domain. See Fig. 4.4. The problem is, given $\mathbf{y} : \partial I^2 \to \partial\Omega$ find a map $\mathbf{x}$ such

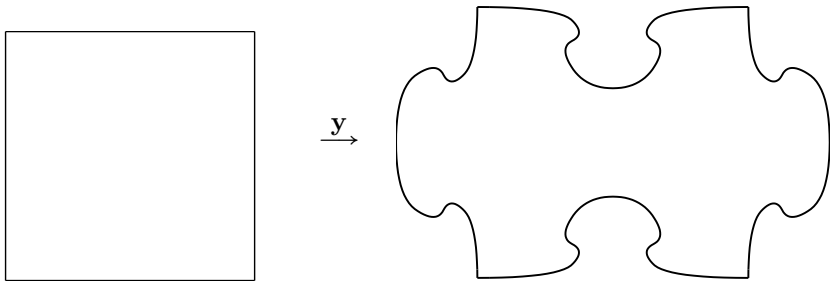


Figure 4.4: How to extend $\mathbf{y}$ to all of Ω?

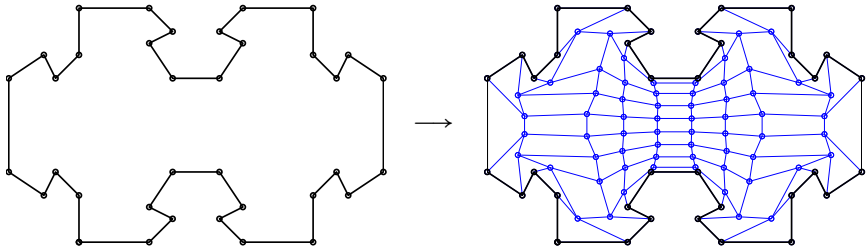
that

$$\mathbf{x} : I^2 \to \Omega, \quad \text{such that } \mathbf{x}|_{\partial I^2} = \mathbf{y}, \quad \det J > 0.$$

In terms of our B-splines basis, we have assumed that the geometry knot sequences have full multiplicity at the boundary, so the boundary control points

$$\mathbf{x}_{1,j}, \mathbf{x}_{N,j}, \mathbf{x}_{i,1}, \mathbf{x}_{i,M}$$

Figure 4.5: An extension of $\mathbf{y}$



completely determine the four boundary curves. Now we need determine the inner control points

$$\mathbf{x}_{i,j}, \quad i = 2, \dots, m-1, \quad j = 2, \dots, n-1.$$

See Fig. 4.5, for an illustration. This is equivalent to the meshing problem for FEM. Several methods have been proposed to solve this problem. The arguably simplest method, and the one we will use is the *spring model*. We will think of all the control points as connected by springs, fixing the boundary, and then letting the system go and settle in an equilibrium. In this way, each control points will be the average of its connected neighbours,

$$\mathbf{x}_{i,j} = \frac{1}{4}(\mathbf{x}_{i-1,j} + \mathbf{x}_{i+1,j} + \mathbf{x}_{i,j-1} + \mathbf{x}_{i,j+1}).$$

This is a linear problem, and straight forward to solve. On the other hand, it does not guarantee that the obtained parametrization is valid, and it will fail for a lot of domains. For simple domains, it does work reasonably well and gives valid parametrizations. We refer to Section 4.6 for examples of solutions to this system. Other more advanced methods have been proposed, to find parametrizations that are guaranteed to be valid and give a good numerical analysis. These can be quite difficult to solve, and usually involves a non-linear optimization problem to be solved, see [34].

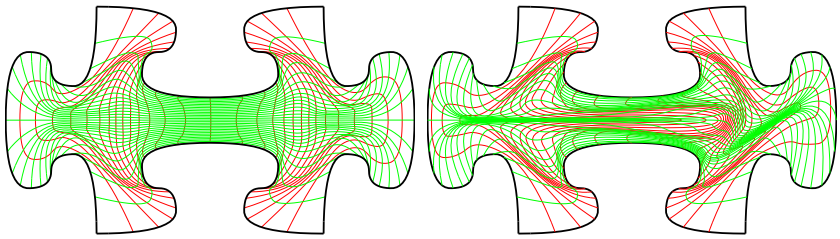


Figure 4.6: Two different parametrizations of Ω. Color is visualizing $|\det J|$

4.4 Discretize-first

Recall that an alternative to *Optimize-first* is the *Discretize-first* paradigm. In this approach we discretize our shape functional first, to obtain a problem on finite dimensional subspace. Using the IGA method of Section 4.3, we get the functional

$$\min_{\mathbf{x}_{ij}} J(\mathbf{x}) = \frac{1}{2}\int_{I^2} |y(\mathbf{x}(\xi,\eta)) - y_d(\mathbf{x}(\xi,\eta))|^2 \det J \,\mathrm{d}\xi\,\mathrm{d}\eta,$$
$$s.t. \quad A(\mathbf{x})[y] = b(\mathbf{x}).$$

Where $y = \sum_k y_k M_k$, so $y \circ \mathbf{x} = \sum_k y_k N_k$, and $[y]$ denotes the vector of control points. It is not a problem to define the derivative of the solution u to the state equation w.r.t the control points, so we can apply the implicit differentiation. Let $' = \partial_{\mathbf{x}^k_{i,j}}$ denote differentiation w.r.t the k coordinate of the control point $\mathbf{x}_{i,j}$, then

$$J(\mathbf{x})' = \int_{I^2} |y - y_0|(y' - \nabla y_d \cdot \mathbf{x}') + \frac{1}{2}|y - y_0|^2 \operatorname{tr}(J^{-1}J') \det J \,\mathrm{d}\xi\,\mathrm{d}\eta \tag{4.16}$$

where y' solves

$$A(\mathbf{x})y' = b'(\mathbf{x}) - A'(\mathbf{x})y. \tag{4.17}$$

With the derivatives of the components of the weak equation given by

$$\begin{aligned} A'_{ij} = \int_0^1 \int_0^1 &\big(\langle J^{-T}J'^{T}J^{-T}\nabla N_i(\xi,\eta), J^{-T}\nabla N_j(\xi,\eta)\rangle \\ &+ \langle J^{-T}\nabla N_i(\xi,\eta), J^{-T}J'^{T}J^{-T}\nabla N_j(\xi,\eta)\rangle \\ &+ \langle J^{-T}\nabla N_i(\xi,\eta), J^{-T}\nabla N_j(\xi,\eta)\rangle \operatorname{tr}\big(J^{-1}J'\big) \det J \,\mathrm{d}\xi\,\mathrm{d}\eta, \\ b'_i = \int_0^1 \int_0^1 &(\langle \nabla f(\mathbf{x}(\xi,\eta)), \mathbf{x}'(\xi,\eta)\rangle \\ &+ f(\mathbf{x}(\xi,\eta)) \operatorname{tr}\big(J^{-1}J'\big) N_i(\xi,\eta) \det J \,\mathrm{d}\xi\,\mathrm{d}\eta \end{aligned}$$

and the derivatives of the parametrization are

$$\mathbf{x}' = \delta_{ij}\boldsymbol{e}_k B_{lk}(\xi,\eta)$$
$$J' = \boldsymbol{e}_k(\nabla N_{i,j})^T$$

Where $\boldsymbol{e}_1 = (1,0), \boldsymbol{e}_2 = (0,1)$ is the standard basis of $\mathbb{R}^2$. This is a finite dimensional optimization problem, which can be solved by standard methods.

4.5 Riemannian shape optimization

In this section we collect results from previous sections, in an effort to solve the model shape optimization problem. Our problem functional only depends on the boundary of Ω, so for regular enough boundaries $J(\Omega)$ can be identified with a function $J : \mathrm{Imm}^n(S^1, \mathbb{R}^2) \to \mathbb{R}$, which we denote by the same symbol. The shape derivative can be interpreted as the differential $dJ(c)$, and the metric G_c opens up the use of Riemannian optimization methods to solve the problem. Using a Riemannian metric to turn a shape derivative into a gradient was first proposed in [69], on shape functionals not constrained by a PDE. This was extended to PDE constrained problems in subsequent papers, [71, 70, 72]. However there are still many unanswered questions, we mention a few:

- Error in definition of Horizontal bundles. This influences the computation of gradients.
- The influence of choice of metric is not investigated.
- All derivation and analysis is done formally on smooth curves. No discussion of the applicability of convergence results.
- No comparison to non-Riemannian methods.

The implementation in these papers is based on FEM, which is not good if we want to study the effect of higher order terms in the metric. Any simple closed continuous curve bounds an interior domain Ω, which we can identify with its boundary Γ, which can be parametrized by an embedding $c : S^1 \to \Gamma$. In this way we obtain a map

$$J : \mathrm{Emb}^n(S^1, \mathbb{R}^2) \to \mathbb{R}$$
$$J(c) = J(\mathrm{int}(\Gamma))$$

Where J is the function defined in (4.1). Classical existence results for elliptic PDE's state that the state equation has a weak solution in $H_0^1(\Omega)$ as long as the boundary is Lipschitz. Hence $J(c)$ is well-defined for $n \geq 2$ by the Sobolev Embedding Theorem. J is a function on the Riemannian manifold $(\mathrm{Emb}^n(S^1, \mathbb{R}^2), G_c)$ where G_c is chosen as a strong Riemannian metric. If J is smooth enough, we can apply algorithms from Sec. 4.2 to minimize it. We can prove that for smooth enough boundaries J is at least continuously directionally differentiable,

Lemma 4.8 *The functional $J : \mathrm{Emb}^n(S^1, \mathbb{R}^d)$ is Gateux-differentiable for $n \geq 3$.*

PROOF. To show differentiability we will use the results from shape calculus in Sec. 4.1. The condition $n \geq 3$ ensures that the boundary of Ω is C^2, hence the solution y of the Poisson equation (4.4) is in $H^2(\Omega)$. Now take a differentiable path $c : (-\epsilon, \epsilon) \to \mathrm{Emb}^n(S^1, \mathbb{R}^2)$ with $c(0) = c_0$ and $c_t(0) = h$. Let Ω_t denote the interior of $c(t)$, by Lemma 4.2, the derivative of J along c exists and is given by the boundary formula, (4.10) with $V(0) = h$,

$$dJ_(h) = \int_{S^1} \left\{ \frac{1}{2}|y_d|^2 + \frac{\partial y}{\partial n}\frac{\partial p}{\partial n} \right\} \langle h, n \rangle \, \mathrm{d}s. \tag{4.18}$$

where (y, p) satisfies the state and adjoint equations

$$-\Delta y = f \text{ in } \Omega,\ y = 0 \text{ on } \Gamma.$$
$$-\Delta p = y - y_d \text{ in } \Omega,\ p = 0 \text{ on } \Gamma.$$

This is linear and continuous in h, which is the statement. □ .

REMARK 6 We conjecture that the result also holds for $n = 2$, but at the time of writing we have no complete proof. The reason to believe this is that the domain formula (4.9) is well-defined for any $c \in H^2(S^1, \mathbb{R}^2)$ and $V \in C^0((-\epsilon, \epsilon), C^1_c(\mathbb{R}^2, \mathbb{R}^2))$, and is continuously linear in V. If we can show how to extend a tangent vector $h \in H^2(S^1, \mathbb{R}^2)$ linearly and continuously to a field V, such that the flow of c along V is in $H^2(S^1, \mathbb{R}^2)$, then J is Gateux-differentiable on $\mathrm{Imm}^2(S^1, \mathbb{R}^2)$. For the boundary formula (4.10) to be valid, the solution y needs to be in $H^{3/2}(\Omega)$ to have a normal derivative on the boundary. This is satisfied for a $C^{1,1}$ domain, so there are H^2-immersions for which (4.18) cannot be valid.

REMARK 7 To show that J is twice-differentiable we would have to compute the shape derivative of (4.18), w.r.t a new vectorfield W. This can be done, but requires more smoothness assumptions on the solution y, hence smoothness of c. The BFGS method requires J to be twice continuously Frechét differentiable, but we have no proof of this. The second derivative is also not needed for the implementation of the optimization procedure. Showing first and second order Frechet differentiability is left for future work.

To compute the $\mathrm{grad}\, J \in T_c \mathrm{Imm}^n(M, \mathbb{R}^2)$, we need to solve the equation

$$G_c(\mathrm{grad}\, J, h) = dJ_c(h).$$

Note that by definition J is invariant under the action of the diffeomorphism group, so $\mathrm{grad}\, J$ is orthoginol to the action of reparametrizations of curve: it is a horizontal vector. If c is smooth, $dJ_c(h)$ defines a distribution on S^1 so

we cannot immediately construct a smooth gradient vectorfield $\operatorname{grad} J$ which is dual to $dJ_c(h)$ through G_c. This is the reason we need a strong Riemannian metric to guarantee that the gradient will be in the right tangent space, for the case of regular curves this restricts us to consider Sobolev curves. Note that if c is Sobolev regularity $n \geq 3$ then by Lemma 2.4 the action of $\operatorname{Diff}(S^1)$ is differentiable, and $\nabla J_G J$ will be G_c-orthogonal to the orbits of $\operatorname{Diff}(S^1)$. For the simplest L^2 metric, this is equivalent to being point-wise normal to c, for higher order metrics the condition is more complicated.

As explained in Section 4.2, we want to avoid computing geodesics, so we choose a retraction $\operatorname{Imm}^n(S^1, \mathbb{R}^2)$. The arguably simplest retraction we can pick is

$$R_c(th) = c + th, \quad c \in \operatorname{Imm}^n(S^1, \mathbb{R}^2),\ h \in T_c \operatorname{Imm}^n(S^1, \mathbb{R}^2)$$

For t small enough this is still a regular curve. Associated to this retraction is the vector transport

$$T_{c,h}(v) = v \in T_{R(c+h)} \operatorname{Imm}^n(S^1, \mathbb{R}^2).$$

This is isomorphism of vector spaces, but not an isometry, which is assumed in the convergence analysis of the BFGS method in [65]. In the same paper the numerical examples also used this relaxed condition, where convergence was still seen in practice, but no convergence was proven. An important property of the manifold is completeness. By the results of Chapter 1, we know that we have completeness for metrics of order $n \geq 2$. The final algorithm for solving the optimization problem is given in Alg. 1. Note that from the Riemannian

Algorithm 2 Riemannian BFGS Shape Optimization

Input: $f, y_d \in H^1(\mathbb{R}^2)$, $c_0 \in \operatorname{Emb}^n(S^1, \mathbb{R}^2)$, $k = 0, B_0 = \operatorname{Id}$.
for $k = 1, 2, \ldots$ **do**
  Update H_k according to (4.14).
  Descent direction: $p_k = H_k \operatorname{grad} J$.
  Do line-search:
  while $c_k + tp_k \notin \operatorname{Emb}^n(S^1, \mathbb{R}^d)$ **or** satisfies Def 4.4(Armijo Point) **do**
    $t \leftarrow \beta t$
  end while
  $c_{k+1} \leftarrow c_k + tp_k$
  if $J(c_{k+1})$ satisfies stop criteria. **then**
    return c_{k+1}
  end if
end for

perspective, the problem is not constrained. The embedding condition is built in naturally in the retraction. The openness of embeddings in all H^n-curves guarantees that backtracking will find a valid curve. Contrast the setting to a classical BFGS method where the problem must be unconstrained.

4.5.1 Discretization

We need to compute $dJ_c(h)$, G_c, grad J and H_k. Again we will use a discretization by B-splines. To use IGA (in a single patch formulation), our geometry parametrization is defined on the unit square $\mathbf{x} : I^2 \to \mathbb{R}^2$, and the boundary is defined by four splines. Such a parametrization can have singularities at the corner points, which contradicts our assumption that the boundary can be parametrized by a regular curve. We remedy this by assuming that only one boundary is free, which is what will use in the examples in Section 4.6. So we consider the space

$$\mathrm{Emb}_0^n([0,1],\mathbb{R}^d) = \{c \in H^n(S^1,\mathbb{R}^d) : |c'| > 0, c(0) = p_0, c(1) = p_1\},$$

with $p_0, p_1 \in \mathbb{R}^d$. The manifold structure of this space is constructed the is the same as for periodic curves, expect some care has to taken on the boundary, we leave out the details here. Sobolev metrics can be defined on this space by altering the boundary condition for the defining differential operator. Metric completeness for higher order metrics on this space is not proven, but we conjecture that the result still holds for fixed end points. For notation, we us define the function

$$B_p(\xi) = p_0 B^{n_\xi}_{\boldsymbol{\xi},1}(\xi) + p_1 B^{n_\xi(\xi)}_{\boldsymbol{\xi},m_\xi}$$

$$\mathcal{S} = \{c(\xi) = B_p(\xi) + \sum_{i=2}^{m_\xi - 1} \mathbf{c}_i B^{n_\xi}_{\boldsymbol{\xi},i}(\xi) : \mathbf{c}_i \in \mathbb{R}^2, \mathbf{c}_1 = p_0, \mathbf{c}_{m_\xi} = p_1\}$$

the space of all spline curves connecting p_0 and p_1, which is an affine space. Then we can parametrize a finite dimensional submanifold of Imm_0^n by

$$\mathcal{S}_e = \mathcal{S} \cap \mathrm{Emb}_0^n([0,1],\mathbb{R}^d).$$

The tangent space is given by

$$T_c\mathcal{S}_e = \left\{ h = \sum_{i=1}^{m_\xi - 2} \mathbf{h}_i B^{n_\xi}_{\boldsymbol{\xi},i+1}(\xi) : \mathbf{h}_i \in \mathbb{R}^2 \right\}$$

Note that since $\mathcal{S}_e$ is locally linear, the tangent space description doesn't depend on the footpoint c. We use the induced metric on $\mathcal{S}_e$, so at c the metric can be represented by the matrix

$$G_{ij} = G_c(B^{n_\xi}_{\boldsymbol{\xi},i+1}, B^{n_\xi}_{\boldsymbol{\xi},j+1})$$

At the level of controlpoints, computing the retraction and vector transport is trivial

$$R_c(th) = B_p(\xi) + \sum_{i=1}^{m_\xi - 2} (\mathbf{c}_i + t\mathbf{h}_i) B^{n_\xi}_{\boldsymbol{\xi},i}(\xi)$$

$$T_{c,h_1}(h_2) = h_1 + h_2, \quad h_1, h_2 \in T_c\mathcal{S}_e$$

When computing the retraction $R_c(th)$ we need to check if $c+th$ is an embedding: no self-intersections and no singular points. For spline curves there are advanced algorithms available to find self-intersections and zeros, see **XX**. [59]. In practice we will use two simple heuristics. For a spline curve c, $|c'|^2$ is also a spline of higher degree, and we can compute $\min |c'|^2$ exactly. To locate self-intersections we compute the turning angle of curve from p_0 to p_1,

$$\alpha = \int_0^1 \kappa(s)\,\mathrm{d}s,$$

using Gaussian quadrature. As mentioned in Sect. 4.3 the elliptic equation only has a solution if the angles of the corners are less than π, so if $|\alpha| > \pi$ the curve must have turned more than once, and thus contain a self-intersection. In Alg. 3 this checking procedure is described with a tolerance level τ for how point-wise singular we want to accept spline curves.

Algorithm 3 Check c is an immersion

Input: $c_0 \in \mathcal{S}_e$, $\tau > 0$.
Compute $\alpha = \int_0^1 \kappa(s)\,\mathrm{d}s$,
Evaluate $\min |c'|^2$ at
if $|\alpha| < \pi$ **and** $\min |c'|^2 > \tau^2$ **then**
  return `true`
else
  return `false`
end if

4.6 Numerical example

We will consider the following example, where the solution is known. Choose y_d, f as

$$y_d = e^{1-(x^2+y^2)} - 1$$
$$f = -4(x^2 + y^2 - 1)e^{1-(x^2+y^2)}$$

Then the solution is unit disc, and the optimal value of the functional is zero. The initial guess of the free boundary is a straight line connecting $(1, 0)$ and $(0, 1)$. The fixed part of boundary is chosen as the L^2 minimizer to the standard unit-speed parametrization of the circle. The initial mesh obtained by the spring method is vizualized in Fig. 4.7: Red nodes are fixed, green nodes are the free parameters and yellow nodes are computed from the spring model. Note

that according with the central theme of this thesis, there are many different parametrizations of the optimal solution, so we cannot directly measure the distance to the circle. We could use the numerical methods in Chapter 2, but we would get different values for different choices of metrics. Instead we will compare the optimization history, computation time and resulting parametrization between *Discretize-first* and Riemannian Shape optimization. We use 10 control points in our test.

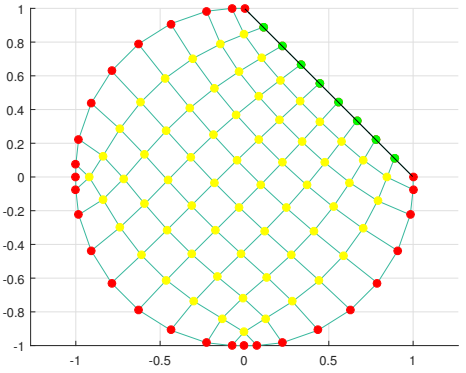


Figure 4.7: Initial mesh

We use Alg. 3 as a non-linear constraint on the space of control points, and implement the functional and gradient given by the *Discretize-first* method from Sect. 4.4, together in Matlab's `fmincon`, using a BFGS interior-point method. The resulting optimization history is depicted in Fig. 4.8 for two different values of the tolerance τ in the embedding constraint. We use the termination criteria $|\nabla J| < 10^{-6}$. For $\tau = 0.1$ the optimization stops at the gradient criteria, the computation time on a 2.4Ghz laptop running Matlab v2015b is 210 sec. For $\tau = 0.01$ the optimization does not terminate by the gradient condtition, but after taking a step smaller than 10^{-12}.

Now we show results using Riemannian optimization, all optimization methods are terminated when $|\operatorname{grad} J| < 10^{-6}$. In Fig 4.10 the optimization history of a Riemmanian steepest descent, which correspond to the choice of $H_k = \mathrm{Id}$ for all k. The history of Riemannian BFGS shape optimization as implemented in Alg. 2, is shown in in Fig. 4.11 and 4.12. Here the influence of different choices of the constants a_0, a_1, a_2 is investigated. The value of functional is show on the left, and the value of the norm of the gradient on the right. The computational time In Fig 4.9 we compare the final optimized meshes from `fmincon` and Riemannian BFGS with $a_0 = 1, a_1 = 0.1, a_2 = 0$. The computational times of the problem is summed up in table 4.1.

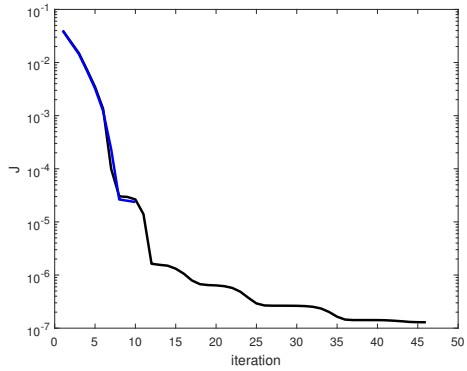


Figure 4.8: `fmincon` optijmization history. Black: $\tau = 0.1$, blue $\tau = 0.01$.

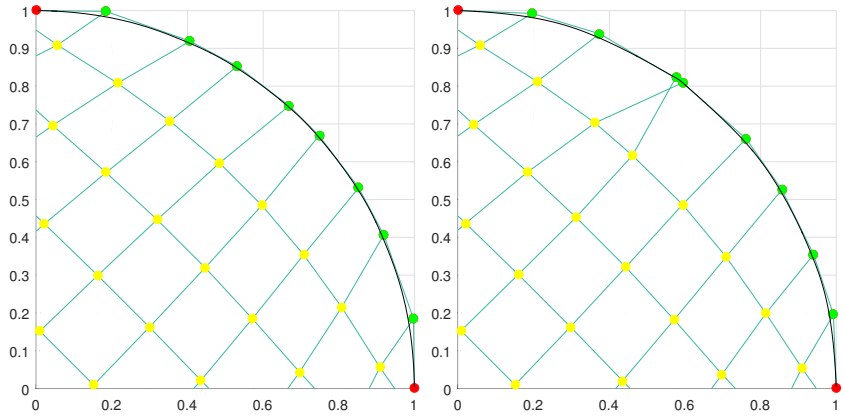


Figure 4.9: Zoom in on solution mesh. Left: Riemannian Optimization.$a_0 = 1, a_1 = 0.1$ Right: `fmincon`, $\tau = 0.1$

4.7 Discussion

In the case of our numerical example, the Riemannian shape optimization seems to be a good alternative to an to a simple discretize-first approach. The parametrization of the solution mesh produced by `fmincon`, see Fig. 4.9 is very close to being singular: two control points have switched places along the circle, here the parametrization is still regular but very close to being singular. This is a typical problem for the solutions produced by `fmincon` when trying other initial guesses and number of control points. The optimization has a tendency to develop singular points on the boundary and so it fails to evaluate the functional. This is typical of shape optimization problems in general, where one

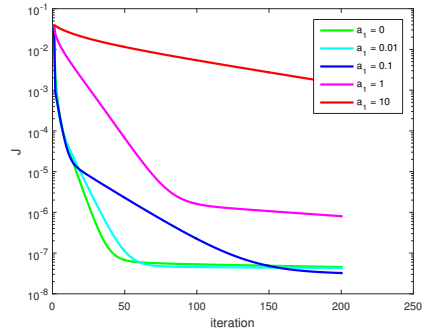


Figure 4.10: Steepest Descent history

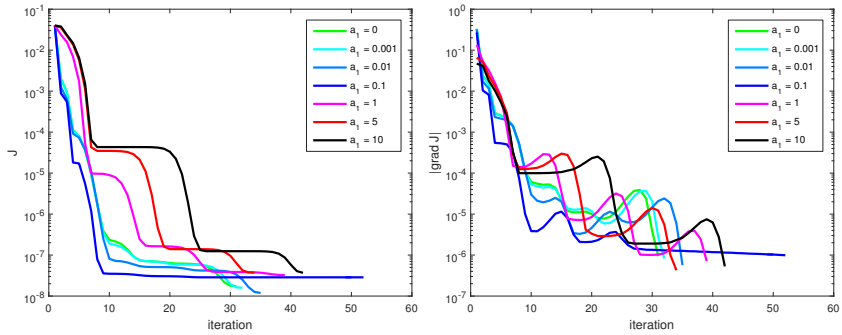


Figure 4.11: Riemannian BFGS optimization history. $a_0 = 1, a_2 = 0$

often has to overcome problems with degenerate meshes. In [57] clustering of control points is also observed in a fluid dynamics problem, and overcome by adding a term to J which penalizes non constant-speed parametrizations.

The mesh shown in Fig. 4.9 for the Riemannian shape optimization is characteristic for any choice of parameters a_i: we see a uniform distribution of control points along the solution circle, and this is without adding a regularization term to J. We could interpret this as the Riemannian metric acting as a regularization. As mentioned before, from the Riemannian perspective, the problem is not constrained by the regularity of the boundary, this is build into the definition of the underlying manifold of immersions.

An interesting observation was that computing the gradient using discretize-first (4.16) along with (4.17), and the shape derivative (4.18), gave numerically very similar results. We mean this in the sense that using (4.18) for the gradient in `fmincon` provided the exact same optimization history. The difference in computational time is somewhat significant, around a factor of 3 faster to compute

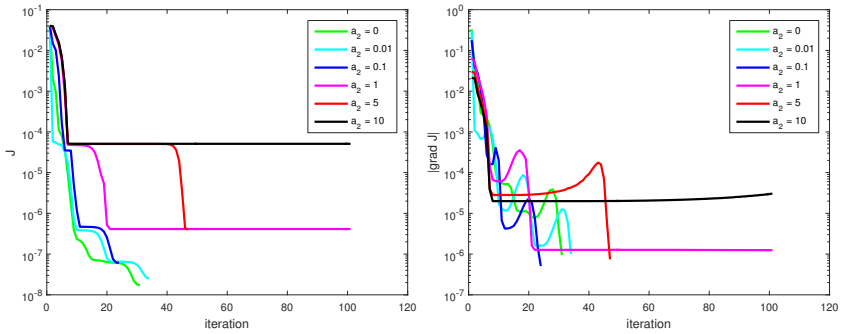


Figure 4.12: Riemmanuan BFGS optimization history, $a_0 = 1, a_1 = 0$

	0	0.01	0.1	1	5	10
a_1	19.3	21.2	122	24.0	73	83.0
a_2	19.2	20.8	14.5	MAX	MAX	MAX

Table 4.1: Computational time in sec. of Riemannian BFGS method on 2.4Ghz laptop. MAX means that the methods stopped after reacting the maximum iteration of 100.

the shape derivative reducing the optimization time from 210 to 70 sec. The difference is due to the fact that computing $A'(x)$ is somewhat slow. This could of course be due to differences in implementation, so we will be reluctant in making any conclusions.

The computational time of one Steepest Descent iteration is roughly the same as a BFGS step, so comparing Fig. 4.10 with Fig. 4.11 and 4.12, we see that BFGS provides a significant speed up compared to a simple first order method. It is difficult to compare computational time across different methods, but taking the above comment into consideration we see that Riemannian BFGS provides a small speed up in computation over `fmincon`. This is only case that the constants in the metric is chosen small. Computing grad J involves inverting the metric G, and as G contains a differential operator, this gives a smoothing effect on the resulting gradient. For large values of a_1 and a_2, inverting G becomes very sensible, so the directions obtained may be quite poor for performing line-search. This is reflected in the tableaus of the history, where the line-search can only find a very small decrease in the objective function, until a good direction is found and the decrease is bigger. For large a_2 the method doesn't converge within the maximum number of allowed iterations. For the choice $a_1 = 0.1$ we see a spike in computational time, where much time is spend in the line-search part, but we have no explanation for this behaviour. It seems best to use $a_1 = a_2 = 0$, or small values of a_1. This is in contrast to the theory which tells us the metric

is strong and the manifold complete only for $a_2 > 0$. A L^2 or H^1 type metric is only weak, so obtaining theoretical results for this case would be very difficult.

Appendix A

Sobolev theory on the circle

Here we will present a quick overview of square integrable Sobolev spaces on the circle, and some results about elliptic operators between them. Integer order Sobolev spaces are standard material in any course on PDE's, but real valued, espcially negative power Sobolev space might not be, so we will present them in a way which includes these. For the circle everything can be done very neatly in terms of Fourier series. For general non-compact manifolds the theory becomes much more complicated and we will not mention this here. The results will follow Chap. 4 in [19], and we refer here for proofs. Another more thorough presentation without elliptic operators can be found in [33].

For any finite sums of exponentials $u = \sum_n \alpha_n e^{inx}$ and $v = \sum_n \beta_n e^{inx}$ define the inner product

$$\langle u, v \rangle_s = \sum_{n \in \mathbb{Z}} (1 + n^2)^s \alpha_n \beta_{-n} \tag{A.1}$$

The space $H^s(S^1)$ is defined as the completion of this space. In this way, $H^s(S^1)$ is simply a space of Fourier series, which converges in a weighted ℓ^2 norm. For positive s the weights grow with n, so the series has to go to zero. For negative s the weights go towards zero with n, so series can be allowed to grow with n.

We have the following classical embedding result.

Lemma A.1 (Sobolev Embedding) *For $s < \frac{1}{2}+k$, the space $H^s(S^1)$ embeds continuously into $C^k(S^1)$.*

We will also need that for regular enough functions, Sobolev spaces define an algebra under multiplication

Lemma A.2 (Sobolev Algebra) *For $s > \frac{1}{2}$, and $f, g \in H^s(S^1)$ we have a well-defined product which satisfies*

$$\|f \cdot g\|_{H^s} \leq C\|f\|_{H^s}\|g\|_{H^s}$$

The spaces embed continuously into each other.

Lemma A.3 *For $s < t$ the embedding $H^t \hookrightarrow H^s$ is continuous.*

An operator of order m $L : H^n \to H^{n-m}$ of the form

$$L(f) = \sum_{k=0}^{n} a_k(x)\frac{d^k f}{dx^k}$$

is said to be elliptic if there exists a constant C such that

$$(-1)^{m/2} a_m(x)\xi^m \geq C|\xi|^m$$

We can characterize invertability of such an operator.

Theorem A.4 *Let L be an elliptic operator of order m, the equation*

$$Lu = f$$

has a solution for any $f \in H^{-\infty}$, if and only if

$$Lu = 0$$

has only the trivial solution $u = 0$.

Appendix B

Convergence of spline approximations

The Hilbert space tensor product $H^k([0,1])\widehat{\otimes}H^\ell(S^1)$ is the completion of the algebraic tensor product $H^k([0,1]) \otimes H^\ell(S^1)$ with respect to the uniform cross norm

$$\beta\Big(\sum_i f_i \otimes g_i\Big)^2 = \sum_{i,j} \langle f_i, f_j\rangle_{H^k([0,1])} \langle g_i, g_j\rangle_{H^\ell(S^1)}\,. \tag{B.1}$$

The following result connects the mixed order Sobolev space (3.9) to a Hilbert space tensor product.

Lemma B.1 *$H^{k,\ell}([0,1]\times S^1)$ is isometrically isomorphic to $H^k([0,1])\widehat{\otimes}H^\ell(S^1)$.*

A similar result for $H^{k,\ell}(\mathbb{R}\times\mathbb{R})$ is shown in [74, Thm. 2.1]. Our proof follows the lines of [47, Thm. 1.39], where the result is shown for the case $k = \ell = 0$.

Proof. Each tensor $c = \sum_i f_i \otimes g_i \in H^k([0,1]) \otimes H^\ell(S^1)$ defines a function $Jc \in H^{k,\ell}([0,1]\times S^1)$, via $Jc(t_1,t_2) = \sum_i f_i(t_1)g_i(t_2)$. It is not hard to verify that J is an isometric embedding of $H^k([0,1]) \otimes H^\ell(S^1)$ in $H^{k,\ell}([0,1]\times S^1)$, i.e., $\beta(c) = \|c\|_{H^{k,\ell}([0,1]\times S^1)}$. To complete the proof, we show that J is onto. Being an isometry, the range of J is closed and so it suffices to show that its

orthogonal complement is trivial in $H^{k,\ell}([0,1]\times S^1)$. Let $c \in H^{k,\ell}([0,1]\times S^1)$ and suppose that $\langle c, f\otimes g\rangle_{H^{k,\ell}([0,1]\times S^1)} = 0$ for all $f \in H^k([0,1])$ and $g \in H^\ell(S^1)$. Let $\langle c, g\rangle_{H^\ell(S^1)}$ denote the function $t_1 \mapsto \int_{S^1} c(t_1,t_2)g(t_2)dt_2$. Then $\langle c, g\rangle_{H^\ell(S^1)} \in H^k(S^1)$ with $\partial_{t_1}^k\langle c, g\rangle_{H^\ell(S^1)} = \langle \partial_{t_1}^k c, g\rangle_{H^\ell(S^1)}$. It follows that

$$\langle c, f\otimes g\rangle_{H^{k,\ell}([0,1]\times S^1)} = \langle f, \langle c, g\rangle_{H^\ell(S^1)}\rangle_{H^k([0,1])} = 0\,.$$

As f is arbitrary, it follows that $\langle c, g\rangle_{H^\ell(S^1)}$ vanishes at almost every t_1. Similarly, since g is arbitrary, c vanishes at almost every t_1, t_2. Therefore, $c = 0$ in $H^{k,\ell}([0,1]\times S^1)$.

Corollary B.2 *The multiplicatively decomposable functions $(t,\theta) \mapsto f(t)g(\theta) = (f\otimes g)(t,\theta)$ with $f \in H^k([0,1])$, $g \in H^\ell(S^1)$, span a dense subspace of $H^{k,\ell}([0,1]\times S^1)$.*

Proof. This follows from the denseness of $H^k([0,1])\otimes H^\ell(S^1)$ in $H^k([0,1])\otimes H^\ell(S^1)$ and Lem. B.1.

The following lemma shows that the Sobolev embedding theorem in one dimension extends to higher dimensions via tensor products.

Lemma B.3 *For each $k,\ell \geq 0$, the space $H^{k+1,\ell+1}([0,1]\times S^1)$ is continuously embedded in the space $C^{k,\ell}([0,1]\times S^1)$.*

Proof. Let $\{f_i\}$ be an orthonormal basis of $H^{k+1}([0,1])$ and $\{g_j\}$ an orthonormal basis of $H^{\ell+1}(S^1)$. Then $\{f_i\otimes g_j\}$ is an orthonormal basis of $H^{k+1}([0,1])\widehat{\otimes}H^{\ell+1}(S^1)$. By Lem. B.1 this space is equal to $H^{k+1,\ell+1}([0,1]\times S^1)$. Therefore, any element in this space can be expressed as $c = \sum_{i,j} c_{i,j} f_i \otimes g_j$ for some $c_{i,j} \in \mathbb{R}$. By the Sobolev embedding theorem in one dimension there exists $C > 0$ such that

$$\begin{aligned}\|\partial_t^k\partial_\theta^l c\|_\infty^2 &= \left\|\sum_{ij} c_{ij}(\partial_t^k f_i)\otimes(\partial_\theta^l g_j)\right\|_\infty^2 \leq \sum_{ij} c_{ij}^2 \|\partial_\theta^k f_i\|_\infty^2 \|\partial_t^l g_j\|_\infty^2 \\ &\leq C\sum_{ij} c_{ij}^2 \|f_i\|_{H^{k+1}([0,1])}^2 \|g_j\|_{H^{\ell+1}(S^1)}^2 = C\|c\|_{H^{k+1,\ell+1}([0,1]\times S^1)}^2\,.\end{aligned}$$

Similar estimates hold for lower derivatives of c. This shows that the $C^{k,\ell}$-norm is bounded by the $H^{k+1,\ell+1}$-norm.

To prove Lem. 3.8 we need a result on the approximation power of one-dimensional splines.

Lemma B.4 *Let $I = [0,1]$ or $I = S^1$, $n,k \in \mathbb{N}$ with $n \geq k$, and $f \in H^k(I)$. Then*

$$\lim_{N\to\infty} \|f - \mathcal{S}_N^n f\|_{H^k(I)} = 0\,.$$

PROOF. The set of smooth functions in dense in $H^k(I)$. Therefore, there is for each $\varepsilon > 0$ a function $g \in C^\infty(I)$ such that $\|f - g\|_{H^k(I)} < \varepsilon/2$. If N is sufficiently large, there is a spline h of order n defined on the uniform grid with N points such that $\|g - h\|_{H^k(I)} < \varepsilon/2$. This follows from [73, Cor. 6.26]. By the best approximation property of the orthogonal projection $\mathcal{S}_N^n$,

$$\|f - \mathcal{S}_N^n f\|_{H^k(I)} \le \|f - h\|_{H^k(I)} \le \|f - g\|_{H^k(I)} + \|g - h\|_{H^k(I)} < \varepsilon\,.$$

Since ε was arbitrary, this shows that $\mathcal{S}_N^n f \to f$ in $H^n(I)$ as $N \to \infty$.

Collecting these results we are able to prove Lem. 3.8.

Proof of Lem. 3.8. Let $c \in H^{k,\ell}([0,1] \times S^1)$ and $\varepsilon > 0$. By Cor. B.2 there exist functions $f_i \in H^k([0,1])$ and $g_i \in H^\ell(S^1)$, $i = 1, \dots, n$, such that

$$\left\| c - \sum_{i=1}^n f_i \otimes g_i \right\|_{H^{k,\ell}([0,1]\times S^1)} < \varepsilon/2\,.$$

By Lem. B.4 and by the fact that the tensor norm is a reasonable cross norm (i.e., $\|f_i \otimes g_i\|_{H^{k,\ell}([0,1]\times S^1)} \le \|f_i\|_{H^k([0,1])} \|g_i\|_{H^\ell(S^1)}$) it is possible to choose N_t and N_θ large enough such that

$$\left\| \sum_{i=1}^n f_i \otimes g_i - \sum_{i=1}^n \mathcal{S}_{N_t}^{n_t} f_i \otimes \mathcal{S}_{N_\theta}^{n_\theta} g_i \right\|_{H^{k,\ell}([0,1]\times S^1)} < \varepsilon/2\,.$$

These two estimates and the best approximation property of the orthogonal projection $\mathcal{S}_{N_t,N_\theta}^{n_t,n_\theta}$ yield

$$\left\| c - \mathcal{S}_{N_t,N_\theta}^{n_t,n_\theta} c \right\|_{H^{k,\ell}([0,1]\times S^1)} \le \left\| c - \sum_{i=1}^n \mathcal{S}_{N_t}^{n_t} f_i \otimes \mathcal{S}_{N_\theta}^{n_\theta} g_i \right\|_{H^{k,\ell}([0,1]\times S^1)} < \varepsilon\,.$$

Appendix C

Derivatives of the energy functional

In this appendix we list the derivatives of the energy functional (3.5). The first derivative is

$$dE_c(k) = \int_0^1 \int_0^{2\pi} t_1\langle c', k'\rangle + t_2\left(\langle c'', k'\rangle + \langle c', k''\rangle\right) + t_3\langle \dot{c}, \dot{k}\rangle + t_4\langle \dot{c}', \dot{k}'\rangle + t_5\left(\langle \dot{c}'', \dot{k}'\rangle + \langle \dot{c}', \dot{k}''\rangle\right) + t_6\langle \dot{c}'', \dot{k}''\rangle \,\mathrm{d}\theta\,\mathrm{d}t\,,$$

with

$$t_1 = \frac{a_0}{|c'|}\langle \dot{c}, \dot{c}\rangle - \frac{a_1}{|c'|^3}\langle \dot{c}', \dot{c}'\rangle - 7\frac{a_2}{|c'|^9}\langle c', c''\rangle^2\langle \dot{c}', \dot{c}'\rangle + 10\frac{a_2}{|c'|^7}\langle c', c''\rangle\langle \dot{c}', \dot{c}''\rangle - 3\frac{a_2}{|c'|^5}\langle \dot{c}'', \dot{c}''\rangle\,,$$
$$t_2 = 2\frac{a_2}{|c'|^7}\langle c', c''\rangle\langle \dot{c}', \dot{c}'\rangle - 2\frac{a_2}{|c'|^5}\langle \dot{c}', \dot{c}''\rangle\,, \quad t_3 = 2a_0|c'|\,, \quad t_4 = 2\frac{a_1}{|c'|} + 2\frac{a_2}{|c'|^7}\langle c', c''\rangle\,,$$
$$t_5 = -2\frac{a_2}{|c'|^5}\langle c', c''\rangle\,, \quad t_6 = 2\frac{a_2}{|c'|^3}\,.$$

The Hessian is

$$
\begin{aligned}
d^2E_c(h,k) = \int_0^1 \int_0^{2\pi} & w_1\langle c',h'\rangle\langle c',k'\rangle \\
& + w_2\left(\langle c'',h'\rangle\langle c',k'\rangle + \langle c',h'\rangle\langle c'',k'\rangle + \langle c',h''\rangle\langle c',k'\rangle + \langle c',k''\rangle\langle c',h'\rangle\right) \\
& + w_3\left(\langle c'',h'\rangle\langle c'',k'\rangle + \langle c',h''\rangle\langle c',k''\rangle + \langle c',h''\rangle\langle c'',k'\rangle + \langle c',k''\rangle\langle c'',h'\rangle\right) \\
& + w_4\left(\langle \dot c,\dot h\rangle\langle c',k'\rangle + \langle \dot c,\dot k\rangle\langle c',h'\rangle\right) + w_5\left(\langle \dot c',\dot h'\rangle\langle c',k'\rangle + \langle \dot c',\dot k'\rangle\langle c',h'\rangle\right) \\
& + w_6\left(\langle \dot c'',\dot h'\rangle\langle c',k'\rangle + \langle \dot c'',\dot k'\rangle\langle c',h'\rangle + \langle \dot c',\dot h''\rangle\langle c',k'\rangle + \langle \dot c',\dot k''\rangle\langle c',h'\rangle\right) \\
& + w_7\left(\langle \dot c',\dot h'\rangle\langle c'',k'\rangle + \langle \dot c',\dot k'\rangle\langle c'',h'\rangle + \langle \dot c',\dot h'\rangle\langle c',k''\rangle + \langle \dot c',\dot k'\rangle\langle c',h''\rangle\right) \\
& + w_8\big(\langle \dot c'',\dot h'\rangle\langle c'',k'\rangle + \langle \dot c'',\dot k'\rangle\langle c'',h'\rangle + \langle \dot c',\dot h''\rangle\langle c'',k'\rangle + \langle \dot c',\dot k''\rangle\langle c'',h'\rangle \\
& \qquad + \langle \dot c'',\dot h'\rangle\langle c',k''\rangle + \langle \dot c'',\dot k'\rangle\langle c',h''\rangle + \langle \dot c',\dot h''\rangle\langle c',k''\rangle + \langle \dot c',\dot k''\rangle\langle c',h''\rangle\big) \\
& + w_9\left(\langle \dot c'',\dot h''\rangle\langle c',k'\rangle + \langle \dot c'',\dot k''\rangle\langle c',h'\rangle\right) \\
& + t_1\langle h',k'\rangle + t_2\left(\langle h'',k'\rangle + \langle h',k''\rangle\right) + t_3\langle \dot h,\dot k\rangle + t_4\langle \dot h',\dot k'\rangle \\
& + t_5\left(\langle \dot h'',\dot k'\rangle + \langle \dot h',\dot k''\rangle\right) + t_6\langle \dot h'',\dot k''\rangle\,\mathrm{d}\theta\,\mathrm{d}t\,,
\end{aligned}
$$

with

$$
\begin{aligned}
w_1 &= -a_0\frac{1}{|c'|}\langle \dot c,\dot c\rangle + a_1\frac{3}{|c'|^5}\langle \dot c',\dot c'\rangle + a_2\frac{63}{|c'|^{11}}\langle c',c''\rangle^2\langle \dot c',\dot c'\rangle \\
&\quad - a_2\frac{70}{|c'|^9}\langle c',c''\rangle\langle \dot c',\dot c''\rangle + a_2\frac{15}{|c'|^7}\langle \dot c'',\dot c''\rangle\,, \\
w_2 &= -a_2\frac{14}{|c'|^9}\langle c',c''\rangle\langle \dot c',\dot c'\rangle + a_2\frac{10}{|c'|^7}\langle \dot c',\dot c''\rangle\,, \quad w_3 = a_2\frac{2}{|c'|^7}\langle \dot c',\dot c'\rangle\,, \quad w_4 = a_0\frac{2}{|c'|}\,, \\
w_5 &= -a_1\frac{2}{|c'|^3} - a_2\frac{14}{|c'|^9}\langle c',c''\rangle^2\,, \quad w_6 = a_2\frac{10}{|c'|^7}\langle c',c''\rangle\,, \quad w_7 = a_2\frac{4}{|c'|^7}\langle c',c''\rangle\,, \\
w_8 &= -a_2\frac{2}{|c'|^5}\,, \quad w_9 = -a_2\frac{6}{|c'|^5}\,.
\end{aligned}
$$

Bibliography

[1] P. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

[2] C. J. Atkin. The Hopf-Rinow theorem is false in infinite dimensions. *Bull. London Math. Soc.*, 7(3):261–266, 1975.

[3] C. J. Atkin. Geodesic and metric completeness in infinite dimensions. *Hokkaido Math. J.*, 26(1):1–61, 1997.

[4] M. Bauer and M. Bruveris. A new Riemannian setting for surface registration. In *3nd MICCAI Workshop on Mathematical Foundations of Computational Anatomy*, pages 182–194, 2011.

[5] M. Bauer, M. Bruveris, P. Harms, and P. W. Michor. Vanishing geodesic distance for the Riemannian metric with geodesic equation the KdV-equation. *Ann. Global Anal. Geom.*, 41(4):461–472, 2012.

[6] M. Bauer, M. Bruveris, P. Harms, and J. Møller-Andersen. Curve matching with applications in medical imaging. In *5th MICCAI workshop on Mathematical Foundations of Computational Anatomy*, 2015.

[7] M. Bauer, M. Bruveris, P. Harms, and J. Møller-Andersen. Second order elastic metrics on the shape space of curves. In *1st Workshp on Differential Geometry in Computer Vision for Analysis of Shapes, Images and Trajectories*, 2015.

[8] M. Bauer, M. Bruveris, P. Harms, and J. Møller-Andersen. A Numerical Framework for Sobolev Metrics on the Space of Curves. *SIAM Journal of Imaging*, to appear, 2016.

[9] M. Bauer, M. Bruveris, S. Marsland, and P. W. Michor. Constructing reparameterization invariant metrics on spaces of plane curves. *Differential Geom. Appl.*, 34:139–165, 2014.

[10] M. Bauer, M. Bruveris, and P. W. Michor. R -transforms for Sobolev H^2 -Metrics on Spaces of Plane Curves. pages 1–35, 2013.

[11] M. Bauer, M. Bruveris, and P. W. Michor. Overview of the geometries of shape spaces and diffeomorphism groups. *J. Math. Imaging Vis.*, 50:60–97, 2014.

[12] M. Bauer, M. Bruveris, and P. W. Michor. R-transforms for Sobolev H^2-metrics on spaces of plane curves. *Geom. Imaging Comput.*, 1(1):1–56, 2014.

[13] M. Bauer, M. Eslitzbichler, and M. Grasmair. Landmark-guided elastic shape analysis of human character motions. 2014.

[14] M. Bauer and P. Harms. Metrics on spaces of immersions where horizontality equals normality. *Differential Geom. Appl.*, 39:166–183, 2015.

[15] M. Bauer, P. Harms, and P. W. Michor. Sobolev metrics on shape space of surfaces. *J. Geom. Mech.*, 3(4):389–438, 2011.

[16] M. Bauer, P. Michor, and O. Müller. Riemannian geometry of the space of volume preserving immersions. 24625:1–22, 2016.

[17] L. Beirão da Veiga, A. Buffa, G. Sangalli, and R. Vázquez. *Mathematical analysis of variational isogeometric methods*, volume 23. 2014.

[18] Benjamin Kimia, Computer Vision Group at LEMS at Brown University. Database of 99 binary shapes. `https://vision.lems.brown.edu/content/available-software-and-databases`, 2015.

[19] L. Bers, F. Jorn, and M. Schechter. *Partial Differential Equations*, volume 3. American Mathematical Society, 1964.

[20] M. V. Boland and R. F. Murphy. A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells. *Bioinformatics*, 17(12):1213–1223, 2001.

[21] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre. Manopt, a Matlab toolbox for optimization on manifolds. *J. Mach. Learn. Res.*, 15:1455–1459, 2014.

[22] H. Brass and K. Petras. *Quadrature theory: the theory of numerical integration on a compact interval*, volume 178 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2011.

[23] M. Bruveris. Completeness properties of Sobolev metrics on the space of curves. *J. Geom. Mech.*, 7(2), 2015.

[24] M. Bruveris. Notes on Riemannian Geometry on Manifolds of Maps. (July):1–18, 2016.

[25] M. Bruveris, P. W. Michor, and D. Mumford. Geodesic Completeness for Sobolev Metrics on the Space of Immersed Plane Curves. *Forum of Mathematics*, 2:1–33, 2014.

[26] M. Bruveris, P. W. Michor, and D. Mumford. Geodesic completeness for Sobolev metrics on the space of immersed plane curves. *Forum Math. Sigma*, 2:e19, 2014.

[27] V. Cervera, F. Mascaró, and P. W. Michor. The action of the diffeomorphism group on the space of immersions. *Differential Geom. Appl.*, 1(4):391–401, 1991.

[28] G. Dal Maso. *An introduction to* Γ*-convergence*. Progress in Nonlinear Differential Equations and their Applications, 8. Birkhäuser Boston, Inc., Boston, MA, 1993.

[29] J. C. De los Reyes. *Numerical PDE-Constrained Optimization*. Springer, 2015.

[30] J. A. Dieudonne. *Foundations of Modern Analysis*. Academic Press, New York, 1969.

[31] M. Eslitzbichler. Modelling character motions on infinite-dimensional manifolds. *The Visual Computer*, pages 1–12, 2014.

[32] M. I. M. G. E. Christensen, R. D. Rabbitt. Deformable templates using large deformation kinematics. *IEEE Trans. Image Process.*, 5:1435–1447, 1996.

[33] P. Garrett. Functions on circles : Fourier series , I. Technical report, 2013.

[34] P. Gravesen, Jens Evgrafov, Anton Dang Manh, Nguyen Nørtoft. Planar Parametrization in Isogeometric Analysis and Shape Optimization. In *Proceedings of the "Eighth International Conference on Mathematical Methods for Curves and Surfaces*, 2012.

[35] P. Griswald. *Elliptic Problems in Non-smooth Domains*. 1984.

[36] R. S. Hamilton. The inverse function theorem of Nash and Moser. *Bulletin of the American Mathematical Society*, 7(1):65–222, 1982.

[37] R. K. Hiess, R. Alter, S. Sojoudi, B. A. Ardekani, R. Kuzniecky, and H. R. Pardoe. Corpus callosum area and brain volume in autism spectrum disorder: quantitative analysis of structural mri from the abide database. *J. Autism. Dev. Disord.*, 45(10):3107–3114, 10 2015.

[38] T. J. R. Hughes, J. A. Cottrell, and Y. Bazilevs. Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement. *Computer Methods in Applied Mechanics and Engineering*, 194(39-41):4135–4195, 2005.

[39] H. Inci, T. Kappeler, and P. Topalov. On the regularity of the composition of diffeomorphisms. 2012.

[40] E. Klassen, A. Srivastava, W. Mio, and S. H. Joshi. Analysis of planar shapes using geodesic paths on shape spaces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(3):372–383, march 2004.

[41] W. Klingenberg. *Riemannian Geometry*. Walter de Gruyter, 1982.

[42] H. Kodama and P. W. Michor. The homotopy type of the space of degree 0-immersed plane curves. *Rev. Mat. Complut.*, 19(1):227–234, 2006.

[43] A. Kriegl and P. W. Michor. *The convenient setting of global analysis*, volume 53 of *Mathematical Surveys and Monographs*. American Mathematical Society, 1997.

[44] H. Krim and A. Yezzi, Jr., editors. *Statistics and Analysis of Shapes*. Modeling and Simulation in Science, Engineering and Technology. Birkhäuser Boston, 2006.

[45] H. Laga, S. Kurtek, A. Srivastava, and S. J. Miklavcic. Landmark-free statistical analysis of the shape of plant leaves. *J. of Theor. Biol.*, 363:41–52, 2014.

[46] S. Lang. *Fundamentals of differential geometry*. 1999.

[47] W. A. Light and E. W. Cheney. Approximation theory in tensor product spaces. *Lecture Notes in Mathematics*, 1169, 1985.

[48] D. G. Luenberger. *Optimization by Vector Space Methods*. 1969.

[49] K. McLeod, M. Sermesant, P. Beerbaum, and X. Pennec. Spatio-temporal tensor decomposition of a polyaffine motion model for a better analysis of pathological left ventricular dynamics. *IEEE Trans. Med. Imaging*, 2015.

[50] P. W. Michor and A. Kriegl. *The Convenient Setting of Global Analysis*, volume 53. American Mathematical Society, 1997.

[51] P. W. Michor and D. Mumford. Riemannian geometries on spaces of plane curves. *J. Eur. Math. Soc.*, 8:1–48, 2006.

[52] P. W. Michor and D. Mumford. An overview of the Riemannian metrics on spaces of curves using the Hamiltonian approach. *Applied and Computational Harmonic Analysis*, 23(2007):74–113, 2007.

[53] W. Mio, A. Srivastava, and S. Joshi. On shape of plane elastic curves. *International Journal of Computer Vision*, 73(3):307–324, 2007.

[54] G. Nardi, G. Peyré, and F.-X. Vialard. Geodesics on shape spaces with bounded variation and sobolev metrics. *SIAM Journal on Imaging Sciences*, 9(1):238–274, 2016.

[55] S. M. Nikolsky. On boundary properties of differentiable functions of several variables. *Dokl. Akad. Nauk SSSR*, 146(3):542–545, 1962.

[56] S. M. Nikolsky. On stable boundary values of differentiable functions of several variables. *Mat. Sb.*, 61:224–252, 1965.

[57] P. Nørtoft and J. Gravesen. Isogeometric shape optimization in fluid mechanics. *Structural and Multidisciplinary Optimization*, 48(5):909–925, 2013.

[58] N. Otsu. A threshold selection method from gray-level histograms. *IEEE T. Syst. Man Cyb.*, 9(1):62–66, 1979.

[59] D. Pekerman, G. Elber, and M. S. Kim. Self-intersection detection and elimination in freeform curves and surfaces. *CAD Computer Aided Design*, 40(2):150–159, 2008.

[60] X. Pennec. Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. *J. Math. Imaging Vision*, 25(1):127–154, 2006.

[61] X. Pennec. Barycentric subspaces and affine spans in manifolds, 2015. To appear in the proceeding of Geometric Science of Information, 2015.

[62] S. C. Preston. The Geometry of Whips. (2):1–24.

[63] S. C. Preston. The motion of whips and chains. *Journal of Differential Equations*, 251(3):504–550, aug 2011.

[64] S. C. Preston and R. Saxton. An H^1 Model for Inextensible Strings. *Discrete and Continuous Dynamical Systems*, 33(5):2065–2083, 2012.

[65] W. Ring and B. Wirth. Optimization methods on Riemannian manifolds and their application to shape space. *SIAM Journal on Optimization*, 22(2):596–627, 2012.

[66] G. K. Rohde, A. J. S. Ribeiro, K. N. Dahl, and R. F. Murphy. Deformation-based nuclear morphometry: capturing nuclear shape variation in HeLa cells. *Cytometry Part A*, 73A(4):341–350, 2008.

[67] M. Rumpf and B. Wirth. Variational time discretization of geodesic calculus. *IMA J. Numer. Anal.*, 35(3):1011–1046, 5 2015.

[68] H.-J. Schmeisser. Recent developments in the theory of function spaces with dominating mixed smoothness. In *Nonlinear Analysis, Function Spaces and Applications*, pages 145–204. Institute of Mathematics of the Academy of Sciences of the Czech Republic, 2007.

[69] V. Schulz. A riemannian view on shape optimization. *Foundations of Computational Mathematics*, 14(3):483–501, apr 2014.

[70] V. Schulz, M. Siebenborn, and K. Welker. Structured Inverse Modeling in Parabolic Diffusion Problems. pages 1–17.

[71] V. Schulz, M. Siebenborn, and K. Welker. Towards a Lagrange-Newton Approach for PDE Constrained Shape Optimization. pages 1–13.

[72] V. Schulz, M. Siebenborn, and K. Welker. A novel Steklov-Poincar\'e type metric for efficient PDE constrained optimization in shape spaces. pages 1–17, 2015.

[73] L. L. Schumaker. *Spline functions: basic theory*. Cambridge Mathematical Library. Cambridge University Press, third edition, 2007.

[74] W. Sickel and T. Ullrich. Tensor products of Sobolev–Besov spaces and applications to approximation from the hyperbolic cross. *J. Approx. Theory*, 161(2):748–786, 2009.

[75] J. Sokolowski and J.-P. Zolésio. *An Introduction to Shape Optimization*. Springer, 1992.

[76] A. Srivastava, E. Klassen, S. Joshi, and Jermyn. I. Shape analysis of elastic curves in euclidean spaces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(7):1415 –1428, 2011.

[77] A. Srivastava, E. Klassen, S. H. Joshi, and I. H. Jermyn. Shape analysis of elastic curves in Euclidean spaces. *IEEE T. Pattern Anal.*, 33(7):1415–1428, 2011.

[78] J. Su, S. Kurtek, E. Klassen, and A. Srivastava. Statistical analysis of trajectories on Riemannian manifolds: bird migration, hurricane tracking and video surveillance. *Ann. Appl. Stat.*, 8(1):530–552, 03 2014.

[79] J. Su, A. Srivastava, F. D. M. de Souza, and S. Sarkar. Rate-invariant analysis of trajectories on Riemannian manifolds with application in visual speech recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 620–627, 6 2014.

[80] G. Sundaramoorthi, A. C. Mennucci, S. Soatto, and A. Yezzi. A new geometric metric in the space of curves, and applications to tracking deforming objects by prediction and filtering. *SIAM J. Imaging Sci.*, 4(1):109–145, 2011.

[81] G. Sundaramoorthi, A. Yezzi, and A. C. Mennucci. Coarse-to-fine segmentation and tracking using Sobolev active contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5):851–864, 2008.

[82] C. Tobon-Gomez, M. De Craene, K. Mcleod, L. Tautz, W. Shi, A. Hennemuth, A. Prakosa, H. Wang, G. Carr-White, S. Kapetanakis, A. Lutz, V. Rasche, T. Schaeffter, C. Butakoff, O. Friman, T. Mansi, M. Sermesant, X. Zhuang, S. Ourselin, H. O. Peitgen, X. Pennec, R. Razavi, D. Rueckert, A. F. Frangi, and K. Rhode. Benchmarking framework for myocardial tracking and deformation algorithms: an open access database. *Medical Image Analysis*, 17(6):632–648, 2013.

[83] J. Vybıral. Function spaces with dominating mixed smoothness. *Diss. Math*, 436:73, 2006.

[84] Q. Xie, S. Kurtek, and A. Srivastava. Analysis of AneuRisk65 data: elastic shape registration of curves. *Electron. J. Stat.*, 8:1920–1929, 2014.

[85] A. Yezzi and A. Mennucci. Conformal metrics and true" gradient flows" for curves. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1. Vol. 1. IEEE*, 2005.

[86] L. Younes. Spaces and manifolds of shapes in computer vision: an overview. *Image Vision Comput.*, 30(6):389–397, 2012.

[87] L. Younes, P. W. Michor, J. Shah, and D. Mumford. A metric on shape space with explicit geodesics. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur.*, 19(1):25–57, 2008.

[88] J.-P. Zolésio and M. C. Delfour. *Shapes and Geometries: Metrics, Analysis, Differential Calculus and Optimization.* SIAM, second edi edition, 2009.