Abstract

This dissertation presents some of the recent developments in the modelling of shape spaces. Forming the basis for a quantitative analysis of shapes, this is relevant for many applications involving image recognition and shape classification. All shape spaces discussed in this work arise from the general situation of a Lie group acting isometrically on some Riemannian manifold. The first chapter summarizes the most important results about this general set-up, which are well known in other branches of mathematics. A particular focus is laid on Hamiltonian methods that explore the relation of symmetry and conserved momenta. As a classical example these results are applied to Kendall’s shape space. More recent approaches of continuous shape models are then summarized and put in the same concise framework. In more detail the square root velocity shape representation, recently developed by Srivastava et al., is being discussed. In particular, the phenomenon of unclosed orbits under the action of reparametrization is addressed. This issue is partially resolved by an extended equivalence relation along with a well defined, non-degenerate, metric on the resulting quotient space.
Acknowledgements

I would like to express my thanks towards Dr Ian Jermyn who supervised this dissertation. His great support, not only in matters of finding the right references and critically asking about my work, did substantially contribute to this project. I would also like to thank Rafael Maldonado for proofreading this work. In particular, his advice on language related errors was very helpful.

Declaration

This dissertation is my own work except where marked otherwise. Material from the work of others has been acknowledged and quotations and paraphrases suitably indicated.

Maximilian Kähler

Durham, 5th September, 2012
Introduction

Usually, if mathematicians take a word from everyday life to name some specific mathematical object, it is not at all, or only vaguely, related to what a non mathematician might imagine the word to mean. For once, this is different. When we are talking about a shape, we actually mean shape, as one might find defined in the Oxford English Dictionary:

“the external form, contours, or outline of someone or something”

This is to say that we are not identifying coffee mugs with doughnuts, as one might do in topology, only because they both share the common feature of a hole. Nevertheless, one needs to take care to make the above idea into a solid mathematical definition and some identification of similar shapes may be wanted.

Most preliminary mathematical descriptions of shape have some superficial degree of freedom, which allows changing the object, but preserves what we might understand the actual shape to be. For instance, the usual approach to turn a picture into some mathematical object would be to draw a coordinate system onto our picture and extract the coordinates of all important features. The human eye will identify a shape as human on some picture regardless of how we rotate the picture. Even though we are not identifying mugs and doughnuts, we may want to identify objects that are related under such similarity transformations. Similarly, we could place an object in different corners of the picture, or magnify a shape. Thus, we will want a mathematical description of shape to be invariant under scaling, translation and rotation. In more complicated descriptions, there may be more such similarity transformations. For example, we will represent the contour of some object as a two dimensional parametrized curve. As we are only interested in the image of this curve, the parametrization has to be removed from the description, i.e. all definitions made have to be independent of the chosen parametrization. Thus, the first challenge of shape theory is defining a mathematical space of shapes, which entirely removes freedom due to similarity transformations.

![Figure 1: Which of these shapes should we consider the same?](image)

The second challenge of shape theory lies in devising a framework of shape comparison. Again, this is something the human eye does automatically and we intuitively recognize the shape of an elephant to be different from the shape of a monkey. Phrasing this challenge in mathematical terms, the second task of shape theory lies in making a shape
space into a *metric space*. Comparison of real numbers is simple. We just consider the modulus of their difference. This nicely generalizes to normed vector spaces, where we still have a sensible notion of addition and subtraction. However, how would one add or subtract shapes? Shape spaces do not naturally form a vector space. Therefore, the definition of a metric is not obvious. This will be the primary goal of the present project.

Another challenge in the analysis of shapes is the development of statistical methods for shape spaces. Given a sample picture from a database, how likely is it to belong to a certain class of picture? How can one perform an automated classification of pictures? Taking a collection of images, is there any way to define a *mean shape*? Classical statistical methods are usually defined on Euclidean space and are not directly applicable to the metric spaces defined in shape theory. To generalize these concepts is, arguably, the most challenging task of modern shape analysis.

Being aware of the tasks ahead, what are the benefits and applications of a theory of shapes? This is a question that could fill entire books. Obviously, any situation when we want to automate shape recognition qualifies for an application of (statistical) shape analysis. This may be in medical imaging (such as in the analysis of MRI scans or computer tomography), homeland security (recognition of potentially dangerous objects) or classification of archaeological image databases. For instance, one could imagine an automated classification of bones or other fossils, based on statistical shape theory. More exotic applications include the analysis of leave growth and tree stem data (e.g. [13, 14]). Thus, the applicability of shape analysis is vast and there seems to be a huge demand for technologies suitable for these applications.

Having outlined the major task of an analysis of shapes and its uses, how can we set about addressing these issues? This dissertation will only deal with the first two of the above problems, i.e. defining a shape space along with a metric. It turns out that there is an entire zoo of methods to approach these tasks. Popular methods include the *iterative closest point algorithm* (ICP) [4] as a tool for point cloud analysis, *level set methods* as in [29] to extract boundaries of contours in images, and so-called *medial axis representations* which reduce shapes to skeletons and thereby simplify their representation, see further [32]. Whereas these particular approaches are in a sense very individual and use very different mathematical techniques, there is also a large group of shape representations that arise in a similar situation: as a quotient of some Riemannian manifold under an isometric Lie group action. These representations include Kendall’s shape space (see chapter 2 and [17]), the angle function representation of Klassen *et al.* [18] as well as the elastic shape representations of Mio, Srivastava, Younes and others (chapter 4, and [26, 35]). In these methods, the notion of an isometric Lie group action gives a very precise meaning to what we understand for a shape to be invariant under certain operations. Furthermore, using Riemannian manifolds as general spaces for shapes to live in allows for a very rich and flexible framework.

The structure of this dissertation is as follows. In the first chapter we will introduce the above general situation of a Lie group acting isometrically on a Riemannian manifold. There are many results, well known in other areas of mathematics and theoretical physics, that are summarized here for convenience. In particular, this chapter will deal with the structure of quotient spaces arising under such actions, as well as the famous relation of symmetries and conserved momenta that, surprisingly, also finds applications to shape analysis. Illustrating the use and applications of the general theory, follows a chapter on the classical shape analysis of David G. Kendall. This will deal with shapes represented on finite dimensional manifolds and forms the prototype of the analysis in later chapters.
Chapter 3 then introduces the general setting for an analysis of continuous shapes. Based on work by Mumford and Michor [24], a short survey of Riemannian metrics on spaces of curves is conducted. This motivates the particular choice of Riemannian metric considered in chapter 4. Here we will not only present the current theory, developed by Srivastava et al. [33], but also present new results. These concern a particular problem that arises when dealing with the action of the infinite dimensional diffeomorphism group $\text{Diff}(I)$. Orbits under this action are not closed and therefore prevent a successful construction of a metric. By introducing a larger equivalence relation, we can resolve this problem, at least in special situations, e.g. for smooth curves. The dissertation is concluded by a summary of open questions, outlining desirable, as yet unproven, results and possibilities to approach these proofs. In an appendix, we describe how elastic shape matching, as in chapter 4, can be implemented using a dynamic programming technique. This is supplemented by a set of specific examples, presenting geodesics in shape space as well as the required parameter changes to perform an optimal matching between shapes.

Before we start with the announced discussion, the author would like to take the chance to make a short comment about giving and taking credit. Despite the general declaration on the second page, there may be need for clarification. There are roughly four classes of material contributing to this dissertation. The first class are theorems, assertions and examples that are directly taken from others. These should and will always be highlighted as such. The second class of material consists of new (but trivial) corollaries, applications and conclusions drawn from the first class of material. Usually, it should be clear from the context or mentioned in the surrounding text where these assertions originate from. A third class of material might be best described as ‘general mathematical knowledge’. It is difficult to give or take credit in this case. The final class of material consists of results that are genuinely new. Of course, a clear separation of these four classes is not possible. In light of this classification the author’s declaration ‘this is my own work…’ is to be understood as follows: chapter 1 - chapter 3 belong to the first three classes of material. What is new about these chapters is the selection and application of the cited results. However, most parts of chapter 4 (especially sections 4.3 - 4.6) and the suggestions in the conclusion chapter belong to the final class, unless explicitly stated otherwise.
1 Quotient Spaces of Isometric Lie Group Actions

In all preliminary descriptions of shapes, one has to deal with their invariance under certain symmetry operations such as translation, rotation and scaling. Therefore, we need a mathematical framework which provides us with a tool of ‘quotienting’ out these symmetries. It was David G. Kendall [16] who had the pioneering idea of representing shapes on a nonlinear manifold. Many other authors (e.g. [33], [35]) took up on this idea and we shall see that all these different representations of shapes can be dealt with in a common framework, namely the concept of isometric Lie group actions on Riemannian manifolds. It should be highlighted, however, that this formalism is not limited to shape analysis. In fact, most of the results in this chapter are well known from the study of Lie groups and find stunning applications in mathematical physics. As all the other chapters of this dissertation are concerned with application to shape analysis, and to do some justice to the richness and variety of other applications, the author takes the freedom to illustrate some of the concepts using examples from theoretical physics. It is hoped that this will explain some of the terminology used in later chapters.

The structure of this chapter is as follows. We will start with the very basic notion of a group action in section 1.1, adding the topological, differential and, finally, Riemannian structure bit by bit in sections 1.2-1.4 and discuss some important properties. Sections 1.5 and 1.6 will be devoted to pushing forward these structures to the quotient under the group action. First describing the special case when the quotient is a smooth manifold in 1.5, we shall discuss what can be carried over to the more general case in section 1.6. It should be stressed that all statements 1.1 - 1.6 deal with finite dimensional manifolds and Lie groups, only. Finally, section 1.7 will discuss problems and difficulties that arise when extending these concepts to infinite dimensions. Unfortunately, it is well beyond the scope of this dissertation to deal with these issues in a rigorous way. It is for this reason and to avoid creating false assertions, that we restrict ourselves to finite dimension in almost the entire chapter.

1.1 Group actions

We will now proceed to describe the general situation. Let \((G, \cdot)\) be a Lie group acting smoothly on a manifold \(M\), i.e. let there be a smooth mapping of differentiable manifolds \(\Phi : G \times M \rightarrow M, (g,p) \mapsto \Phi(g,p)\), such that

\[
\Phi(g \cdot h, p) = \Phi(g, \Phi(h, p))
\]

\[
\Phi(e, p) = p
\]

for all \(g, h \in G, p \in M\).\(^1\) We will often write \(\Phi(g, p) = g.p\) if the context doesn’t allow any other action. It is sometimes more convenient to think of such an action in a slightly more abstract way. If \(\Phi\) is an action of \(G\) on \(M\), in the above way, then every \(g \in G\) induces a diffeomorphism \(\Phi_g : M \rightarrow M, p \mapsto g.p\). The above compatibility condition now reads

\[
\Phi_{g \cdot h} = \Phi_g \circ \Phi_h,
\]

where \(\circ\) denotes the composition of maps. Thus, we may think of an action as a group homomorphism

\[
\Phi : G \rightarrow \text{Diff}(M)
\]

---

\(^1\)Here we will only consider left actions. One could also consider right actions \(\Phi : M \times G \rightarrow M\), in which case the compatibility condition reads \(\Phi(\Phi(p, g), h) = \Phi(p, g \cdot h)\). This will occur in the case of the reparametrization action on parametrized curves.
between $G$ and the diffeomorphism group of $M$. Given such an action, there are various standard definitions repeated here for completeness.

**Definition 1.1.** a) For $p \in M$ the orbit of $p$ is defined as 
\[ [p] = \{ g.p \in M \mid g \in G \}. \]
b) The stabilizer or isotropy group of $p \in M$ is defined as 
\[ G_p = \{ g \in G \mid g.p = p \} \subset G. \]
c) An action is called transitive if $M$ consists of only one orbit.
d) An action for which all stabilizers $G_p, p \in M$, are trivial is called a free action.
e) We say an action is effective if for every $g \in G$ there exists some $p \in M$, such that $g.p \neq p$. This is equivalent to demanding that the above group homomorphism is injective.

Throughout this dissertation our main object of interest will be the quotient space 
\[ M/G = \{ [p] \mid p \in M \}. \]
It is easily checked that $p \sim p' :\iff p' \in [p]$ is an equivalence relation. Therefore $M/G$ is a well defined quotient space.

### 1.2 Topological and differential structure of orbits and quotients

So far, all these definitions can be made for a generic group action without demanding any topological properties of the mappings and spaces. Next we will endow $M/G$ with the quotient topology, making the canonical projection $\pi: M \to M/G$ an identification map (i.e. $U \subset M/G$ is open iff $\pi^{-1}(U) \subset M$ is open). The following lemma collects some important properties of orbits and stabilizers, using the additional differential structure of $G$ and $M$ as manifolds.

**Theorem 1.2** ([28], Theorem 2.1). Let $G$ be a finite dimensional Lie group, acting on a manifold $M$ and for $p \in M$ consider the map $\alpha^p: G \to [p] \subset M, g \mapsto g.p$. Then $\alpha^p$ is a map of constant rank $k$, for some $k \leq \dim(G)$, and the following holds:

a) The stabilizer $G_p$ is a normal Lie subgroup of $G$ with Lie algebra $T_{e}G_x = \ker d\alpha^p|_{g=e}$.

b) The orbit $[p]$ is an immersed submanifold in $M$ of dimension $k$ and $[p] \simeq G/G_x$.

We note that $[p]$ as a whole is not always an embedded submanifold of $M$. Generally, the global differential structure of $[p]$ may not be compatible with the one on $M$. To see this, consider the following example:

**Example 1.3.** Let $\Phi$ be the action of $\mathbb{R}$ on the Torus $T = [0,1]^2/\sim$ given by 
\[ (t; [x,y]) \mapsto [x + t, y + \sqrt{2}t]. \]
It is a well known topological result that $\mathbb{R}[x,y]$ is dense in $T$ for any $[x,y] \in T$, wrapping around the torus infinitely often without ever closing the orbit. Thus, the orbit fails to be a submanifold globally. Indeed, we chose any point $p$ in the orbit, take an arbitrary open neighbourhood $U_p$ and intersect it with the orbit. This intersection will never contain only one line segment (see figure 2).
This example is also interesting in another respect: \([p]\) does not generally constitute a closed subset in \(M\). Although this example is of a rather pathological nature, we will encounter the same problem when we deal with the action of the infinite dimensional diffeomorphism groups \(\text{Diff}(S^1)\) and \(\text{Diff}([0, 1])\). This is closely related to the question of whether \(M/G\) is a Hausdorff space.

**Lemma 1.4.** Let \(M/G\) be a Hausdorff space. Then \([p]\) will be closed in \(M\) for all \(p \in M\).

Recalling that any metric space is, in particular, a Hausdorff space, we conclude that \(M/G\) is not metrizable, if not all orbits are closed. We should pause for a moment to reflect on this. At first, it may not be obvious why this is relevant to our applications. However, it is of vital importance for the problem of constructing a metric between shapes. Only if sets of equivalent shapes are closed in our pre-shape space (i.e. \(M\)), do we have any hope of finding a non degenerate metric between shapes. Having understood the importance of this issue, it is now of interest to have criteria for \(M/G\) to be Hausdorff. Of course, a universal remedy is to demand compactness of \(G\), yet it will be useful to have a less restrictive criterion. This is found in the notion of a proper action.

**Definition 1.5.** An action \(\Phi : G \times M \to M\) is called a proper action if for any two sequences \(\{g_n\} \subset G, \{p_n\} \subset M\) with \(p_n \to p, g_n.p_n \to p'\), there exists a converging subsequence of \(\{g_n\}\) with limit point \(g\) and \(g.p = p'\).

The above definition is equivalent to demanding that the map \(G \times M \to M \times M, (g, m) \mapsto (m, g.m)\) is a proper map, i.e. preimages of compact sets are compact. It is easy to see that orbits are closed subsets of \(M\) if \(G\) acts properly on \(M\). Moreover, the subset

\[ R := \{(p, g.p) \mid (p, g) \in M \times G\} \]

is closed in \(M \times M\). As Abraham and Marsden show in [1], proposition 4.1.19, this is sufficient for \(M/G\) to be Hausdorff. Another problem we might encounter is that the dimension of the orbits may vary across the manifold, leaving us with different types of orbits, depending on the corresponding isotropy group. Consider the following extreme example.

**Example 1.6.** \(\mathbb{R}^+ = \{\alpha \in \mathbb{R} \mid \alpha > 0\}\) acts on \(\mathbb{R}^m\) by scaling \((\alpha, x) \mapsto \alpha x\). Orbits are rays originating (but not including) zero, as well as the set \(\{0\}\). Whereas the rays are of dimension one and are not closed in \(\mathbb{R}^m\), \(\{0\}\) constitutes a zero-dimensional closed set. We know that \((\mathbb{R}^m \setminus \{0\})/\mathbb{R}^+\) is a smooth manifold, the \((m-1)\)-sphere \(S^{m-1}\). Including the origin, however, destroys the differential structure. Indeed, the quotient even fails to be Hausdorff. We note that this action is not free.

It turns out that this problem is closely related to the question of whether \(M/G\) can naturally inherit the differential structure of \(M\). As example 1.6 suggests, we need all orbits to be of one type. Since \([p] \simeq G/G_p\), we are lead to the natural demand that
\(G_p = G'_{p'}\) for all \(p, p' \in M\). In this case, however, we may just as well consider the action of \(G/G_p\) instead of \(G\). This leaves us with the criterion that an action should be free (only trivial stabilizers). To exclude topological pathologies, one also requires that the action is a proper action:

**Proposition 1.7 ([1], Prop. 4.1.23).** If \(\Phi : G \times M \to M\) is a proper free smooth action, then \(M/G\) is a smooth manifold and \(\pi : M \to M/G\) is a submersion, i.e. the differential

\[d\pi_p : T_p M \to T_{\pi(p)} M/G\]

is surjective for all \(p \in M\) and \(\ker d\pi_p = T_p[p]\).

The proof amounts to showing that the above set \(R\) is a closed submanifold of \(M \times M\), which is shown to be equivalent to \(M/G\) being a smooth manifold. This is by no means trivial. However, it does not add any insights for problems in shape analysis, and is therefore left out. Note that in this situation we have

\[(T_p M)/(T_p[p]) \simeq T_p[M/G].\]

A more complicated example when \(M/G\) is not a manifold will be Kendall’s shape space in dimension higher than two. This kind of structure is known as an orbifold. Before we proceed to the discussion of Kendall’s Shape space, we will introduce the last remaining part of structure on \(M\), namely the Riemannian metric.

### 1.3 Riemannian Manifolds and Isometric Group Actions

**Definition 1.8.** Let \(G\) be a Group acting on a Riemannian manifold \((M, \eta), (g, p) \mapsto g.p = \Phi_g(p)\). The action is said to be isometric if the differential

\[d\Phi_g : T_p M \mapsto T_{g.p} M\]

is an isometric isomorphism of tangent spaces for every \(p \in M\), i.e. \(d\Phi_g\) is an isomorphism such that for any \(\xi, \zeta \in T_p M\)

\[\eta(\xi, \zeta)_p = \eta(d\Phi_g \xi, d\Phi_g \zeta)_{g.p}.\]

We recall that for every smooth path \(\gamma : [0, 1] \to M\), the Riemannian metric can be used to define the path length as

\[L[\gamma] = \int_0^1 \eta(\dot{\gamma}(t), \dot{\gamma}(t))^{1/2} dt.\]

Furthermore, we may define the geodesic distance on \(M\) by

\[d_M(p, q) = \inf \{L[\gamma] \mid \gamma : [0, 1] \to M, \gamma(0) = p, \gamma(1) = p\}\]

for \(p, q \in M\). The geodesic distance \(d_M\) makes \((M, d_M)\) a well defined metric space. In finite dimensions, the topology induced by this metric is compatible with the original topology. This assertion is absolutely non-trivial, however. Please refer to Lang [20] p. 190, proposition 6.1.

It is not difficult to see, that an isometric group action preserves the distance \(d_M(p, q) = d_M(g.p, g.q)\) for all \(g \in G\). Curves attaining the infimum in (1) are called geodesics, a
concept which is well known from the study of Riemannian geometry. In fact, the notion of a geodesic is slightly more general, allowing that geodesics are local extrema of the length functional. Equivalently, geodesics are characterised via the concept of parallel transport. See further Jost [15]. If $\nabla$ is the Levi-Civita connection on $M$, a curve is geodesic if and only if $\nabla_t \dot{\gamma} = 0$, i.e. if its tangent vector is parallel transported along the curve. To illustrate this, consider the following example.

**Example 1.9** (Geodesics on $S^n$). Geodesics on the $n$-sphere are given by great circles. Let $p \neq \pm q \in S^n \subset \mathbb{R}^{n+1}$ and let $\vartheta = \arccos(p, q) \in [0, \pi]$. Then the segment of the great circle joining $p$ and $q$ is given by

$$\gamma(t) = \frac{1}{\sin \vartheta} (\sin(\vartheta(1 - t)))p + \sin(\vartheta t)q.$$  

Hence, the geodesic distance is given by the length of this arc $d_M(p, q) = \vartheta$. Of course, we could also consider the longer arc joining $p$ and $q$. Being locally length minimizing, this is also a geodesic. If $p = -q$, i.e. if the points are antipodal, there are infinitely many shortest length curves joining $p$ and $q$.

**Example 1.10** (Geodesics in $\mathbb{R}^n \setminus \{0\}$). It is well known that geodesics in $\mathbb{R}^n$ are given by straight lines. Removing the origin, however, leads to the difficulty that two points lying on opposite ends of a straight line through the origin can no longer be joined by a shortest length geodesic. The infimum in (1) is not attained by any path.

The last example directly leads us to the concept of geodesic completeness, which is characterised by the Theorem of Hopf-Rinow.

**Theorem 1.11** (Hopf-Rinow). Let $(M, \eta)$ be a connected finite dimensional Riemannian manifold. Then the following statements are equivalent:

1. The closed and bounded subsets of $M$ are compact;
2. $M$ is a complete metric space;
3. $M$ is geodesically complete; that is, for every $p$ in $M$, the exponential map $\exp_p$ is defined on the entire tangent space $T_pM$.

Furthermore, any one of the above implies that, given any two points $p$ and $q$ in $M$, there exists a length minimizing geodesic connecting these two points.

**Proof.** Please refer to [15] or any other textbook on Riemannian geometry. 

### 1.4 Killing vectors and conserved momenta

Having briefly recalled these basic facts about Riemannian manifolds, we can now proceed to discuss the notion of flows and Killing vector fields. A more detailed discussion may be found in Abraham & Marsden [1], Chapter 4 ‘*Hamiltonian Systems with symmetry*’. Suppose that $G$ acts on $M$ isometrically. For every $\xi \in T_qG$ consider the Lie exponential $t \mapsto \exp_G(t\xi)$. This gives rise to a one-parameter group $\{\phi_t^\xi\}_{t \in \mathbb{R}}$ of isometries on $M$

$$\phi_t^\xi : M \rightarrow M, p \mapsto \exp_G(t\xi) \cdot p.$$
and a corresponding vector field

$$X^\xi : M \rightarrow TM, p \mapsto \frac{d}{dt}_{t=0} \exp_G(t\xi).p$$

In physics, a one-parameter subgroup such as \(\{\phi^\xi_t\}_{t \in \mathbb{R}}\) is often referred to as flow and \(X^\xi\) is called its infinitesimal generator. In the context of Riemannian geometry and General Relativity such vector fields are also called Killing vector fields. By definition (and making use of \(\phi^\xi_{t+s} = \phi^\xi_t \circ \phi^\xi_s\)) we have,

$$\frac{d}{dt} \phi^\xi_t(p) = X^\xi(\phi^\xi_t(p)),$$

i.e. the curves \(t \mapsto \phi^\xi_t(p)\) are integral curves of the vector field \(X^\xi\).

**Example 1.12.** Consider the action of \(G = SO(m)\) on \(M = \mathbb{R}^m\) by left matrix multiplication. As orthogonal transformations preserve the metric \(\langle \xi, \zeta \rangle = \langle O\xi, O\zeta \rangle\) for all \(\xi, \zeta \in T_x\mathbb{R}^m \simeq \mathbb{R}^m, O \in SO(m)\), this is easily seen to be an isometric action. The Lie exponential for a matrix Lie group is given by

$$\exp_{SO(m)}(J) = \sum_{n=0}^{\infty} \frac{1}{n!} J^n.$$ The Lie algebra is characterised by

$$T_{Id}SO(m) = \{ J \in \mathbb{R}^{m \times m} \mid J^T = -J \}$$

the set of all skew symmetric \(m \times m\) matrices. For a given \(J \in T_{Id}SO(m)\) we have its flow \(\phi^J_t(x) = \exp_{SO(m)}(tJ).x\) and the corresponding infinitesimal generator \(X^J(x) = J.x\). For \(m = 2\) the Lie algebra is only one-dimensional, spanned by \(J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\). In this case, integral curves are just concentric circles (see figure 3).

Note that in the above example the integral curves coincide with the orbits of the action. This is no coincidence. By construction, such integral curves will always lie within one orbit. Locally, they will even span the orbit. Yet some care needs to be taken with these assertions when it comes to infinite dimensional Lie algebras (see section 1.7 and Milnor [25] for further discussion). Also note that the notion of exponential map that exists for the manifold \(M\) (also known as Riemannian normal coordinates) is generally independent of the exponential map on \(G\). The curves \(t \mapsto \phi^J_t(p)\) will usually not be geodesics. This is trivially shown in the above example: Geodesics in \(\mathbb{R}^2\) are straight lines, but the orbits of \(SO(2)\) are concentric circles. An exception may occur, for example, when \(M\) is itself a Lie group and \(G\) is a Lie subgroup of \(M\). Riemannian normal coordinates and Lie exponential in this case are identical (at least if the metric on \(M\) is chosen canonically). However, these details need not concern us here, as shape spaces do not usually feature a Lie group structure themselves.

Killing vector fields play an important role in many parts of mathematical physics. The main reason for this is that they provide constants of motion or conserved momenta in the following sense:
Lemma 1.13. Let $X$ be a Killing vector field and $\gamma : [0, 1] \rightarrow M$ be a geodesic. Then $\eta(X(\gamma(t)), \dot{\gamma}(t))_{\gamma(t)}$ is conserved, i.e. for all $t \in [0, 1]$

$$\frac{d}{dt} \eta(X(\gamma(t)), \dot{\gamma}(t))_{\gamma(t)} = 0.$$ 

Going back to the one parameter group of isometries $\phi^t$, we see that this is just another facet of the famous Noether Theorem: ‘Every continuous symmetry implies the existence of a conserved momentum’. To illustrate where this terminology originates from, let us consider the following example from theoretical physics.

Example 1.14. Let $M = \mathbb{R}^{1, 3}$ be the ‘flat space time’, endowed with the Lorentzian metric

$$\eta(\xi, \zeta) = -\xi^0 \zeta^0 + \xi^1 \zeta^1 + \xi^2 \zeta^2 + \xi^3 \zeta^3,$$

where $\xi, \zeta \in T_x M \simeq \mathbb{R}^{1, 3}$. The famous Poincaré group $G = \mathbb{R}^{1, 3} \rtimes SO(1, 3)$ acts isometrically on $M$ via

$$G \times M \rightarrow M, (b, \Lambda; x) \mapsto \Lambda x + b.$$

The $\mathbb{R}^{1, 3}$ part of the semi-direct product $G$ implements the translational symmetry of space time, whereas $SO(1, 3)$ is the group of so-called Lorentz transformations, consisting of rotations in three dimensional space and relativistic boosts. The translational Killing fields for $b \in \mathbb{R}^{1, 3}$ are thus given by the constant vector fields

$$x \mapsto \frac{d}{dt} \exp_G(tb).x = \frac{d}{dt} (x + tb) = b.$$

Correspondingly, we find for a geodesic $\gamma(t)$ the conserved momentum $\eta(b, \dot{\gamma}(t))$. As this must be constant for all $b \in \mathbb{R}^{1, 3}$, we find that $\dot{\gamma}(t) \equiv \text{const}$. This is nothing but the usual linear 4-momentum of a particle travelling on a straight line with constant velocity. The same procedure could be followed for rotational symmetries and boosts. This would yield the well known conservation of angular momentum (or rather, its relativistic equivalent).

In the context of shape analysis, we will later find symmetries associated with scaling or reparametrization of curves. Following the terminology of Mumford and Michor [24], the corresponding momenta will be called scaling momentum or reparametrization momentum, respectively. When we introduce the notion of an orthogonal section in definition 1.20, we will see how conserved momenta can simplify our analysis.

1.5 Horizontal geodesics and the quotient metric

Having summarized how $M/G$ can become a smooth manifold in section 1.2 and after studying isometric Lie group actions in the last section, we are now ready to define a Riemannian structure on $M/G$. We shall assume that $\pi : M \rightarrow M/G$ is a submersion (e.g. the action is proper and free) and, in addition, we demand that the group action be isometric. In this case we can canonically endow $M/G$ with a Riemannian metric. To see how this works we may first decompose the tangent space $T_p M$ at $p \in M$ into a horizontal and a vertical part

$$T_p M = T_p[p] \oplus H_p M,$$  

\(^2\text{At this point we do not worry about the difference between a Lorentzian and a Riemannian metric. Most of the concepts described above can be transferred to the Lorentzian case.}\)
where $H_pM$ is the orthogonal complement of $T_p[p]$ in $T_pM$. Since $\ker d\pi_p = T_p[p]$, this induces the isomorphism
\[
d\pi_p |_{H_pM} : H_pM \longrightarrow T_p[M/G],
\]
which we use to push forward the metric of $M$. For $v, w \in T_p[M/G]$ we define
\[
\bar{\eta}(v,w)_{[p]} = \eta((d\pi_p)^{-1}v, (d\pi_p)^{-1}w)_{[p]}.
\]
(4)

To see that this is well defined, i.e. independent of the point $p \in [p]$, let $p' = g.p \in [p]$. We recall that $g$ gives rise to an isometry $d\Phi g$ of tangent spaces $T_pM$ and $T_{g.p}M$ (definition 1.8). As $d\Phi g$ preserves the metric, we may restrict it to an isometry of the horizontal spaces $H_pM$ and $H_{g.p}M$. Now observe that the following diagram commutes:

\[
\begin{array}{ccc}
H_pM & \xrightarrow{d\Phi_g} & H_{g.p}M \\
d\pi_p & & \downarrow d\pi_{g.p} \\
& T_p[M/G] &
\end{array}
\]

Therefore, (4) is indeed a well defined inner product on $T_p[M/G]$. The smoothness of the metric can be deduced from the fact that it is the pushforward of a smooth metric under the smooth projection $\pi$.

Having constructed a Riemannian metric on $M/G$, we are now interested in the relation between geodesics on $M$ and geodesics on $M/G$. First, we may introduce the notion of horizontal curves and horizontal geodesics.

**Definition 1.15.** a) A curve $\gamma : [a,b] \longrightarrow M$ is called horizontal, if $\dot{\gamma}(t) \in H_{\gamma(t)}M$ for all $t \in [a,b]$.

b) A horizontal geodesic is a geodesic which is also a horizontal curve.

Horizontal curves are a useful concept, as they can be used to describe curves in $M/G$. Figure 4 illustrates the concept of horizontality in a sketch illustration. In some cases, a curve $\alpha : [a,b] \longrightarrow M$ can be lifted for any $p \in \pi^{-1}(\alpha(a))$ to a horizontal curve $\tilde{\alpha}$ in $M$, such that $\alpha = \pi \circ \tilde{\alpha}$ and $\tilde{\alpha}(a) = p$. If this actually holds for any curve in $M/G$, the Riemannian submersion is called *Ehresmann-complete*. If $M$ is complete, then the Riemannian submersion $\pi : M \longrightarrow M/G$ is Ehresmann-complete (c.f. Falcitelli et al. [10], p. 34). As we will see later, horizontal geodesics play a crucial role in our analysis of shapes. We summarize a few useful properties.

**Proposition 1.16.** i) A geodesic, horizontal at one point, is horizontal everywhere.

ii) Projections of horizontal geodesics are geodesics in $M/G$.

iii) If $M$ is complete and connected the reverse is also true, i.e. a horizontal lift of a geodesic in $M/G$ is a geodesic in $M$. Furthermore, $M/G$ will be complete.

\footnote{With the Hopf-Rinow theorem, we know that complete as a metric space and geodesically complete is the same. However, for Ehresmann-completeness we only need the former notion. It is good to keep that in mind for the infinite dimensional case, where Hopf-Rinow is generally false.}
Figure 4: Illustration of a horizontal and a non horizontal curve

**Proof.** For i) we may observe that $T_p[p]$ is spanned by Killing vector fields $X^\xi, \xi \in T_eG$, as in section 1.4. If $\gamma$ is a geodesic and horizontal at some point $t_0$ we find $\eta(X^\xi, \dot{\gamma}(t_0))_{\gamma(t_0)} = 0$ for all $\xi \in T_eG$. However, using Lemma 1.13, we find that this is a conserved quantity. Therefore, $\eta(X^\xi, \dot{\gamma}(t))_{\gamma(t)} = 0$ for all $t$ and $\gamma$ is horizontal everywhere.

ii) and iii) employ certain relations between the Levi-Civita connection $\nabla$ on $M$ and the induced Levi-Civita connection $\nabla'$ on $M/G$. See Falcitelli et al. [10] p.25 and p. 37. If $\gamma$ is horizontal, $\nabla \dot{\gamma}$ will also be horizontal and for $\gamma' = \pi \circ \gamma$ we have $d\pi(\nabla \dot{\gamma}) = \nabla' \dot{\gamma}'$. This proves ii) and the first part of iii).

For iii) note that the Hopf-Rinow theorem shows completeness of $M$ to be equivalent to the property that every geodesic can be extended to the entire real line. Given a geodesic $\gamma': [a,b] \rightarrow M/G$, we may lift it to a geodesic $\gamma \in M$ (using Ehresmann completeness) and extend it in $M$ to the entire real line. By i) the extended $\gamma$ will still be horizontal. By ii) its projection is a geodesic.

Assuming Ehresmann-completeness, the proof of ii) may also be done in another way. Using the isometry of $d\pi |_{H_pM}$, the length of a horizontal curve $\gamma$ coincides with the length of its projection $\pi \circ \gamma$.

$$L_M[\gamma] = \int \|\dot{\gamma}(\tau)\|d\tau = \int \|d\pi \dot{\gamma}(\tau)\|d\tau = \int \|d\tau (\pi \circ \gamma(\tau))\|d\tau = L_{M/G}[\pi \circ \gamma]$$

If $\gamma$ is length minimizing in $M$ it will still be length minimizing if we restrict to horizontal curves. As we may lift any curve in $M/G$ to a horizontal curve in $M$, $\pi \circ \gamma$ will be length minimizing in $M/G$ and is therefore a geodesic.

As in section 1.3, $M/G$ is naturally endowed with a metric

$$d_{M/G}([p], [q]) = \inf \{L_{M/G}[\gamma] \mid \gamma : [0,1] \rightarrow M/G, \gamma(0) = [p], \gamma(1) = [q]\}. \quad (5)$$

A crucial step to generalizing this metric to the case where $M/G$ is no longer a manifold is the following equivalent definition of this metric:

$$\tilde{d}_{M/G}([p], [q]) = \inf_{g,h \in G} d_M(g.p, h.q) = \inf_{g \in G} d_M(p, g.q) \quad (6)$$

Here, the last inequality follows directly from the isometry of the group action.\footnote{In chapter 4 we will have equivalence classes that are not the orbits of an isometric group action. In this situation, the last equality is absolutely non-trivial and may generally be false.}

**Proposition 1.17.** Let $M$ be a complete connected manifold. For any $[p], [q] \in M/G$ the definitions (5) and (6) coincide.
Although the infimum in (6) is not always attained, it is useful to make the following

we say that

The form (6) is desirable as it makes the abstract quotient metric more explicit and

To establish equality, let \( \{g_n\}_{n \in \mathbb{N}} \subset G \) be a sequence, such that \(|\tilde{d}_{M/G}([p],[q]) - d_M(p,g,q)| \leq 1/n\). Furthermore, let \( \alpha_n \) be a sequence of (not necessarily horizontal) geodesics with \( d_M(p,g_n,p) = L_M[\alpha_n] \). We then find

\[
\tilde{d}_{M/G}([p],[q]) \geq d_M(p,g_n,q) - \frac{1}{n} = L_M[\alpha_n] - \frac{1}{n} \geq L_{M/G}[\pi \circ \alpha_n] - \frac{1}{n} \geq d_{M/G}([p],[q]) - \frac{1}{n}.
\]

Passing the limit \( n \to \infty \), we establish the desired result. Note, even though the \( \alpha_n \) may not be horizontal, we can still project them to a curve \( \pi \circ \alpha_n \) joining \( [p] \) and \( [q] \). Generally, we have \( L_{M/G}[\pi \circ \alpha_n] \leq L_M[\alpha_n] \), as the projection cuts off the vertical part. \( \square \)

Although the infimum in (6) is not always attained, it is useful to make the following definition.

**Definition 1.18.** Let \([p],[q] \in M/G\). If there exists a \( g \in G \), such that

\[
d_{M/G}([p],[q]) = d_M(p,g,q),
\]

we say that \( p \) and \( g.q \) are in **optimal position** or **optimally registered**.

The form (6) is desirable as it makes the abstract quotient metric more explicit and enables us to reduce its calculation to an optimization problem in \( M \). In our applications, \( M \) is usually the simpler space, e.g. a sphere or a flat vector space. Optimally registered points have the following useful property.

**Proposition 1.19.** ([14], theorem 2.4) Let \( p \) and \( g.p \) be optimally registered, joined by a geodesic \( \alpha \) with \( L[\alpha] = d_M(p,g.p) \). Then \( \alpha \) is horizontal.

Note, however, that two points on a horizontal geodesic segment are not necessarily in optimal position. An obvious example where this may happen is when two points on the sphere are connected by the longer arc of the great circle (if it happens to be horizontal for some action). A more elaborate example with arbitrarily close points is shown in [14] theorem 5.4b, in the context of Kendall’s shape space. We now continue with the description of another tool for making quotients more explicit, the notion of an orthogonal section.

**Definition 1.20.** A submanifold \( S \subset M \) is called an **orthogonal section** of \( M \) if

1. \( S \cap [p] \) contains not more than one point for all \( p \in M \)
2. \( T_p[p] \oplus T_pS = T_pM \) for all \( p \in S \)
3. \( T_p[p] \perp T_pS \) for all \( p \in S \)

If \( S \cap [p] \) contains exactly one point for all \( p \in M \), we say \( S \) is a global orthogonal section.

Note that orthogonal sections do not always exist. Let us now try to see how this works and study a very basic example.
Example 1.21. Let $M = \mathbb{R}^m \setminus \{0\}$ and let the metric at $x \in M$ be

$$\eta(\xi, \zeta)_x = \langle \xi, \zeta \rangle_{\mathbb{R}^n} / \langle x, x \rangle_{\mathbb{R}^n}.$$ 

Consider the action of $\mathbb{R}^+, (\alpha, x) \mapsto \alpha x$, as discussed in example 1.6. The scaling factor in the denominator of the metric makes this action isometric. A qualified guess might already suggest that

$$S = \{ x \in M \mid \langle x, x \rangle_{\mathbb{R}^n} = 1 \} = S^{n-1}$$

is an orthogonal section. In fact, it is easy to see that $T_p S$ is orthogonal to rays originating from 0 and we see that all conditions 1.-3. are met (recalling that these rays are the orbits).

Having promised that these orthogonal sections constitute a simplification to our analysis, the following lemma summarizes why this is the case.

Lemma 1.22. Let $S$ be a global orthogonal section for a free proper isometric group action $G \times M \rightarrow M$. Then $\pi |_S : S \rightarrow M/G$ is an isometric diffeomorphism of Riemannian manifolds.

Proof. As $S \cap [p]$ contains exactly one representative of the orbit $[p]$, the smooth map $\pi |_S$ is bijective. Conditions 2. and 3. of definition 1.20 imply $T_p S = H_p[p]$ for the horizontal part $H_p[p]$ of the tangent space (c.f. eq. (3) above). However, we have already seen that $H_p[p] \simeq T_{[p]}(M/G)$ are isometrically isomorphic. 

Thus, we may work on the submanifold $S$ instead of using the abstract quotient. We will often use this technique to remove the scaling and translation actions from our considerations. Furthermore, this finally explains why we are so interested in conserved momenta. These help us find orthogonal sections! Indeed, whenever the momentum vanishes for all curves in some submanifold $S \subset M$, the criterion $T_p[p] \perp S$ is automatically satisfied. This technique is illustrated in much detail in section 2.2, when we discuss Kendall’s shape space.

1.6 Generalized geodesics for non smooth Quotients

If $M/G$ is not a manifold, there will not be a Riemannian metric to define a useful notion of a geodesic. We do not even have a notion of when a curve is smooth, as the quotient does not everywhere look like $\mathbb{R}^n$ locally and there may be some ‘bumps’. Kendall describes in [17] how one can still have a useful notion of a geodesic in $M/G$ for the special case of Kendall’s shape space. Huckemann et al. generalize this idea in [14]. We shall closely follow their description in this section. Firstly, we note that one may still think of $H_p M$ as the tangent space to $M/G$ at $[p]$, only keeping in mind that their disjoint union is no longer a smooth tangent bundle. Secondly, we assert that the alternative definition of the quotient metric

$$d_{M/G}([p], [q]) = \inf_{g \in G} d_M(p, g.q)$$

is still possible, as it does not rely on the existence of a Riemannian metric on $M/G$. Moreover, the proposition 1.19 still holds. Its proof only uses properties of $M$ and $G$ and does not require any particular structure in the quotient (c.f. [14]). This suggests that horizontal geodesics are still a useful concept in the general case and gives rise to the following definition.
**Definition 1.23.** A curve \( \gamma \) in \( M/G \) is called a generalized geodesic, if it is the projection of a horizontal geodesic \( \alpha \) in \( M \).

Just as before, we call \( \alpha \) a horizontal lift of \( \gamma \). Furthermore, we can find a natural notion of the length of a generalized geodesic:

\[
L[\gamma] = L_M[\alpha] = \int \|\dot{\alpha}(\tau)\|\,d\tau.
\]

Of course, we need to check that this is independent of the lift \( \alpha \). To that account, let \( \beta \) be another horizontal lift \( \pi \circ \beta = \gamma \). We may find a smooth curve \( \delta \subset G \), such that \( \gamma(t) = \delta(t).\beta(t) \). Taking the derivative with respect to \( t \) yields,

\[
\dot{\gamma}(t_0) = \frac{d}{dt}(\delta(t).\beta(t_0)) \bigg|_{t=t_0} + \frac{d}{dt}(\delta(t_0).\beta(t)) \bigg|_{t=t_0}
\]

for all \( t_0 \). But the first of the two terms is tangent to the orbit (as the curve \( \delta(t).\beta(t_0) \) lies within). Furthermore, the differential \( d\Phi_{\delta(t_0)} \) preserves the metric and \( \beta \) is horizontal, so that the second term is entirely horizontal. Thus, for \( \gamma \) to be horizontal, the first term must vanish identically, implying that \( \delta(t) \) be constant. Therefore, there exists a \( g \in G \), such that \( \gamma = g.\beta \) and the length of the curve is preserved, by isometry of the group action.

Having generalised the concept of geodesics, it is of interest whether we can prove something similar to the Hopf-Rinow theorem. At least in the case of a compact Lie group \( G \) we have a very pleasing result.

**Corollary 1.24** ([14], corollary 2.5). Let \( M \) be a finite dimensional complete Riemannian manifold, and \( G \) a compact Lie group acting isometrically on \( M \). Then any \( q_1, q_2 \in M/G \) are joined by a generalized geodesic of length \( d_{M/G}(q_1, q_2) \).

**Proof.** As \( G \) is compact the infimum in (6) is attained and there are \( p_1 \in q_1 \) and \( p_2 \in q_2 \) optimally registered. Using the completeness of \( M \) and proposition 1.19 we find a horizontal geodesic of minimal length joining the two points. Its projection to \( M/G \) is the desired generalized geodesic.

The metric and generalized geodesics are almost all we need for our purposes in shape analysis, as our main practical interest is the construction of a metric between shapes. However, it turns out that one can accomplish quite a lot more. Let us define\(^6\) \( M^* := \{p \in M \mid G_p = \{e\} \} \) and \( M^0 := M \setminus M^* \). We find the following theorem:

**Theorem 1.25** ([14], theorem 2.7). Let \( G \) be a compact Lie group acting isometrically and effectively on a finite-dimensional Riemannian manifold \( M \). Then

a) \( M^* \) and \( \pi(M^*) \) are open and dense in \( M, M/G \), respectively.

b) Any geodesic on \( M \) that meets \( M^* \) has at most isolated points in \( M^0 \).

---

\(^5\)One may wonder why this kind of product rule holds. This can be made rigorous by treating the action as a map \( \Phi : G \times M \to M \) and using the usual rules for the differential.

\(^6\)Huckemann *et al.* use the terminology **regular** and **singular** space for these sets.
Note that, in particular, \( \pi(M^*) \) is a smooth manifold. Furthermore, the projection map \( \pi|_{M^*} : M^* \rightarrow \pi(M^*) \) is a Riemannian submersion, which follows from applying the results of sections 1.2 and 1.5. As done for Kendall’s shape space in [17], Chapter 6, one may carry this result even further and break up the manifold \( M \) into submanifolds of points according to their orbit type (i.e. their stabilizer). Although this looks like a very nice result, there are, unfortunately, some issues with sectional curvatures tending to infinity as we approach points in \( M^0 \). This is discussed in detail in [14]. For the applications in this dissertation, we are content with the established theory.

1.7 Obstacles in infinite dimensions

It was mentioned at the beginning of this chapter that we eventually intend to apply the theory developed here to infinite dimensional manifolds. First, we should make clear what kind of spaces we are talking about. Classically, a manifold is a topological space, which is a patchwork of sets that look like open sets of \( \mathbb{R}^n \) (the base space). These are well understood and widely used objects. The main technique in proving results about finite dimensional differential manifolds is to use charts and apply results that are known to hold in \( \mathbb{R}^n \). Naturally, the question arises as to whether we can use more general spaces than \( \mathbb{R}^n \) as base spaces. These should allow basic differential calculus, such as derivatives and smooth maps. This leads to the concept of Banach-, Hilbert- and Fréchet-manifolds, which are essentially patchworks of sets that look like an open subset of a Banach-, Hilbert- and Fréchet-spaces, respectively. Lang develops in [19] almost all the classical theorems about differential manifolds, such as inverse and implicit function theorems, in a language that naturally allows Banach spaces as a base space. One difficulty is to free the theory of notions like constant rank, and to avoid proofs that are solely based on dimensionality arguments (e.g. proving surjectivity by showing the dimension of the image of some linear map to be equal to the dimension of the target space). Another aspect to pay attention to is the closedness of subspaces. Whenever we decompose a Banach space into a subspace and its complement, we have to make sure that the spaces are closed.\(^7\) This adds restraints to what kind of maps between manifolds we consider (i.e. kernel and image of the differential have to be closed). Taking into account these difficulties, the theory of Banach manifolds is very well developed.

As we mostly deal with Riemannian manifolds, we will also want to allow the base space to be a Hilbert space. If we endow the tangent bundle with a family of inner products and if this family meets some kind of smoothness conditions, we have found the infinite dimensional generalization of a Riemannian structure. Unfortunately, here we encounter the first major problem for our theory of quotient manifolds. The theorem of Hopf-Rinow (theorem 1.11) is generally false in infinite dimensions, as is proven by Atkin in [2]. Two points on a connected complete Hilbert-manifold need not be connected by a geodesic. Thus, whenever we deal with complete manifolds we have to carefully distinguish between the notion of a complete metric space and the property that any two points may be joined by geodesics. For instance, corollary 1.24 is no longer true.

The third generalization of manifolds we will need for shape analysis, is the notion of a Fréchet-manifold. A Fréchet-space is a locally convex vector space \( V \) carrying a translation invariant metric\(^8\) which makes \( V \) a complete metric space. This is slightly more general than a Banach space, as the metric is not necessarily induced by a norm. However, Fréchet

---

\(^7\)Imagine, for instance, a dense subspace of a Banach space.

\(^8\)I.e. there exists a metric \( d : V \times V \rightarrow \mathbb{R}_{\geq 0} \) such that \( d(x + a, y + a) = d(x, y) \) for all \( x, y, a \in V \)
spaces still allow many important notions of differential calculus, enabling us to develop a
theory of Fréchet-manifolds. An example of a Fréchet space is $C^\infty(S^1, \mathbb{R}^2)$, some of its
subsets forming the starting point for our analysis of continuous shapes. Fréchet manifolds
are in a sense very general, as they allow many generalizations of finite dimensional
concepts, one such being the notion of a Fréchet-Lie group, i.e. a topological group which
is a Fréchet manifold. A very general introduction to infinite dimensional Lie groups is
found in Omori [27]. The only such groups we are interested in are the diffeomorphism
groups $\text{Diff}([0, 1])$ and $\text{Diff}(S^1)$. Unfortunately, a detailed discussion of these spaces is well
beyond the scope of this project. However, we should point out some of the notorious
difficulties in infinite dimensions.

Our first warning concerns the Lie group structure of diffeomorphism groups $G = \text{Diff}(M)$. Every such group comes with an exponential map $\exp_G : U \rightarrow G$, where
$U \subset T_eG$ is an open neighbourhood of the origin. In finite dimensions, we are used to
think of this as a local chart, being invertible and mapping the Lie algebra $T_eG$ (or at
least an open neighbourhood of 0) diffeomorphically to $G$. Unfortunately, this is no longer
true. In any neighbourhood of the identity, there are elements that cannot be expressed as
the exponential of an infinitesimal generator $\xi \in T_eG$. Thus, $\exp_G$ is not locally invertible.
An example of this is given in Milnor [25] p. 1017.

A second warning is to the ‘properness’ of our actions. We shall encounter examples
where the orbits of our action are no longer closed. As pointed out in section 1.2 this is
essential for the construction of a quotient metric. See also proposition 3.3 for a concrete
example. The solution will be to widen the equivalence relation (and hence the orbits) in
such a way that they are closed. This will be explicitly described in chapter 4.

Our final warning is, related to the last point, that the quotient metric, defined as the
infimum of geodesic distance taken over the entire orbit (c.f. section 1.5 eq. (6), may
be degenerate. As Mumford and Michor show in [23, 22], there are metrics on spaces of
curves that produce a vanishing distance between elements of the quotient. This does
not only happen when orbits are unclosed, but also in ‘proper’ situations. This is due
to the ‘weak nature’ of Riemannian metrics we are dealing with. In fact, there are two
different notions of Riemannian metric in infinite dimension. A Riemannian metric is a
family of inner products on the tangent spaces. In finite dimensions, the tangent spaces
are always complete with respect to the norm induced by these inner products. If this
still holds in infinite dimensions, we call a Riemannian metric strong. Otherwise, the
metric is called a weak Riemannian metric. Unfortunately, strong metrics do not exist for
the class of Fréchet manifolds we are considering. This means that quite a lot of results,
which we usually take for granted, are no longer true. One such problem is that the metric
topology induced by the geodesic distance (section 1.3, eq. (1)) is no longer the same
as the topology the manifold naturally carries. The study of weak Riemannian metrics
is still an active area of research. Please refer to Clarke [7] for a detailed discussion of
Fréchet manifolds with weak Riemannian metric.
2 Kendall’s Shape Space

Having introduced some theoretical background material in chapter 1, we are now ready to see a first example of its applications: Kendall’s shape space. As was already mentioned, Kendall was the first to propose the idea of shape analysis in a Riemannian framework (c.f. [16]). His work was originally inspired by problems in archaeology and astronomy, studying random alignments of polygons. Apart from rigorously defining the shape spaces, Kendall also proposed some tools for the statistical analysis on these non linear spaces, which lead to the foundation of an entirely new mathematical area: statistical shape analysis and statistics of non linear manifolds. As we are mainly interested in the study of continuous shapes, we will focus on shape modelling and only describe the most basic of Kendall’s ideas. This will be the prototype of our analysis in later chapters, introducing the concept of pre-shape and shape spaces.

At the beginning of the chapter we shall look at the action of the scaling, translation and rotation groups on the space of \(k\)-ads in \(\mathbb{R}^m\). This will lead us to a Riemannian metric, that is invariant under these actions and which takes a very convenient form in a specific set of coordinates. Section 2.2 treats translation and scaling, using the language of conserved momenta and orthogonal sections. It turns out that scaling and translation can be very conveniently removed from our shape representation. Both actions are free proper actions and the resulting quotient is still a manifold. However, the action of the rotational group \(SO(m)\) is not as simple. Section 2.3 discusses how the quotienting is performed in this case and points out the failure of the action to be free in dimension \(m \geq 3\). The following section 2.4 shows how the optimal rotational alignment between two pre-shapes can be found using a technique called Procrustes analysis. This holds for all \(m\). The case \(m = 2\) is special, as the rotational action is free. Furthermore, we can identify \(\mathbb{R}^2 \simeq \mathbb{C}\), allowing a simpler treatment of planar \(k\)-ads. This will be the content of section 2.5 along with several explicit examples of geodesics in shape space. Finally, the last section of this chapter will discuss some issues and limitations of this so-called landmark based shape analysis and gives some motivation as to why we should consider continuous shapes.

2.1 Preliminaries

The starting point of our analysis is the space of not totally degenerate ordered \(k\)-ads in \(\mathbb{R}^m\), \(X = (x(1), \ldots, x(k))\), which we identify with the space

\[ M = \mathbb{R}^{m \times k} \setminus \{0\}. \]

We shall always assume that \(m < k\) to avoid spelling out too many special cases.\(^9\) As an open subset of \(\mathbb{R}^{m \times k} \simeq \mathbb{R}^{m \cdot k}\) this is an \(m \cdot k\) dimensional differential manifold that comes with a natural Riemannian metric

\[ \tilde{\eta}(W, V)_X = \text{tr}(WV^t) = \sum_{j=1}^{k} \langle w(j), v(j) \rangle_{\mathbb{R}^m}, \quad (7) \]

for \(X \in M\) and \(W, V \in T_X M \simeq \mathbb{R}^{m \times k}\). We now wish to identify shapes that are related to each other by similarity transformations, i.e. scaling, rigid translation and rotation.\(^10\) The

\(^9\)Excluding these cases is not a real limitation. For instance, shapes consisting of only one or two vertices are not particularly interesting.

\(^10\)We do not include reflections, meaning, for instance, that we still distinguish between a right hand and a left hand in our images.
corresponding groups are $G_T = \mathbb{R}^m$ (translation), $G_S = \mathbb{R}^+$ (scaling) and $G_R = SO(m)$ (rotation). It should come as no surprise that these act as

$$(b, X) \mapsto (x_{(1)} + b, \ldots, x_{(k)} + b) = X + b.1_k \quad \text{(translation)}$$

$$(\alpha, X) \mapsto (\alpha \cdot x_{(1)}, \ldots, \alpha \cdot x_{(k)}) = \alpha \cdot X \quad \text{(scaling)}$$

$$(O, X) \mapsto (O.x_{(1)}, \ldots, O.x_{(k)}) = O.X \quad \text{(rotation)}$$

where $1_k = (1, \ldots, 1) \in \mathbb{R}^{1 \times k}$. We can immediately check that the metric (7) is invariant under rotation and translation. Indeed, orthogonal transformations preserve the inner product on $\mathbb{R}^m$ and the translational action is not felt in the tangent space. However, the metric is not invariant to scaling, as the action of $\alpha \in \mathbb{R}^+$ will introduce a factor of $\alpha^2$ multiplying the metric. This leads us to studying the conformally related metric

$$\eta(W, V)_X = \frac{\tilde{\eta}(W, V)_X}{s(X)^2} = \frac{\text{tr}(WV^t)}{s(X)^2}, \quad (9)$$

where we defined the size of the shape $X = (x_{(1)}, \ldots, x_{(k)})$ as

$$s(X) = \left(\sum_{j=1}^k ||x_{(j)} - \bar{x}||_{\mathbb{R}^m}^2\right)^{1/2}.$$ 

Here, $\bar{x}$ denotes the centroid of $X$, $\bar{x} = \frac{1}{k} \sum_{j=1}^k x_{(j)}$. The shape size is also invariant under translation and rotation and reacts homogeneously to scaling, $s(\alpha X) = \alpha s(X)$ for all $\alpha \in \mathbb{R}^+, X \in M$. We note that, whereas scaling and rotation commute in their action, both actions do not commute with translation, e.g. $\alpha (X + b.1_k) \neq \alpha X + b.1_k$. Therefore, we do not have a well defined action of the product $G_T \times G_S \times G_R$ and we have to take some care in which order we proceed with the quotienting. However, we do have a well defined action of the semi-direct product $G = \mathbb{R}^m \times (\mathbb{R}^+ \ltimes SO(m))$, given by

$$(b, \alpha, O; X) \mapsto \alpha O.X + b.1_k.$$ 

Therefore, we might directly apply the theory of chapter 1 to this action. It will be more convenient, however, to break this up into several steps. Before we proceed, we will introduce another set of coordinates for $M$. Let us define

$$Z_{(j)} = \frac{1}{\sqrt{j^2 + j}} \left(jx_{(j+1)} - (X_{(1)} + \ldots + X_{(j)})\right) \quad (j = 1, \ldots, k - 1)$$

$$Z_{(k)} = \frac{1}{\sqrt{k}} \left((X_{(1)} + \ldots + X_{(k)})\right)$$

In the first k-1 slots this transformation captures the relative coordinates of $X_{(j+1)}$ with respect to the previous $X_{(i)}$, $i = 1, \ldots, j$. The last column vector is proportional to the centroid and represents the absolute position of the k-ad. The transformation is equivalent to right multiplication by the $k \times k$ orthogonal matrix

$$Q_k = \begin{pmatrix}
-\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{4}} & \cdots & \cdots & -\frac{1}{\sqrt{(k-1)^2 + (k-1)}} \\
\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{4}} & \cdots & \cdots & -\frac{1}{\sqrt{(k-1)^2 + (k-1)}} \\
0 & \frac{2}{\sqrt{3}} & -\frac{1}{\sqrt{4}} & \cdots & \cdots & -\frac{1}{\sqrt{(k-1)^2 + (k-1)}} \\
0 & 0 & \frac{3}{\sqrt{4}} & \cdots & \cdots & -\frac{1}{\sqrt{(k-1)^2 + (k-1)}} \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \cdots & \frac{1}{\sqrt{(k-1)^2 + (k-1)}} \\
0 & 0 & 0 & \cdots & 0 & -\frac{(k-1)}{\sqrt{(k-1)^2 + (k-1)}} \\
\end{pmatrix}$$
and its inverse is given by the transpose $Q_k^t$. Going back to the action (8), we see that the action of scaling and rotation does not change and is still given by (left-)multiplication. The only change is in the translational part of the action. This is entirely restricted to the last column

$$(b; Z_{(k)}) \mapsto Z_{(k)} + \sqrt{k}b,$$

as all other slots only represent relative coordinates. The benefit of these coordinates will soon become clear. Let us first see how the metric is expressed in these coordinates.

Noting that $Z \mapsto \sqrt{Q_k^t Z Q_k}$, indeed, expressing the shape size as

$$s(Z)^2 = \sum_{j=1}^{k-1} \|Z_{(j)}\|_{R^m}^2,$$  \tag{10}

Indeed, expressing the shape size as $s(X) = \text{tr}((\dot{X} \dot{X})^t)$ with $X = X - \bar{x}1_k$ results in

$$s(X)^2 = \text{tr}((\dot{X} \dot{X})^t) = \text{tr}((\ddot{X} Q_k)(\dot{X} Q_k)^t) = \sum_{j=1}^{k} \|((\ddot{X} Q_k)_{(j)})\|_{R^m}^2.$$  

Noting that $(\ddot{X} Q_k)_{(j)} = Z_{(j)}$ for $j = 1, \ldots, k - 1$, and $(\ddot{X} Q_k)_{(k)} = 0$, proves (10).

### 2.2 Removing Scaling and Translation

It is now time to employ some of the theory of section 1.4 and study Killing vectors and conserved quantities on this manifold. Working in Kendall coordinates\footnote{This term is nowhere to be found in the literature. It seems appropriate to have a name for these coordinates, though.} the translational Killing vector field $\xi^b$ for $b \in \mathbb{R}^m \simeq T_0 \mathbb{R}^m$ is given by

$$Z \mapsto \frac{d}{dt} (\exp_{\mathbb{R}^m}(tb).Z) = \frac{d}{dt} \left( Z_{(1)}, \ldots, Z_{(k-1)}, Z_{(k)} + \sqrt{k}tb \right) = \left( 0, \ldots, 0, \sqrt{k}b \right)$$

The corresponding linear momentum $p_b$ of a curve $\gamma = (\gamma_{(1)}, \ldots, \gamma_{(k)})$ in $M$ is given by

$$p_b = \eta(\xi^b, \dot{\gamma}(t))_{\gamma(t)} = \frac{1}{{s(\gamma(t))}^2} \text{tr}(\xi^b \dot{\gamma}(t)^t) = \frac{\sqrt{k}}{s(\gamma(t))^2} (b, \dot{\gamma}(t)_{(k)})_{\mathbb{R}^m}.$$  

Thus, for a curve to be horizontal to translations we need $\dot{\gamma}(t)_{(k)} = 0$, i.e. the centroid has to be constant. Next, we consider the scaling action. For $\alpha \in \mathbb{R} \simeq T_1 \mathbb{R}^+$ we have the Killing vector field $\xi^\alpha$

$$Z \mapsto \frac{d}{dt} \left( e^{t\alpha} Z \right)_{t=0} = \alpha Z,$$

which results in the scaling momentum

$$p_\alpha = \eta(\alpha \gamma(t), \dot{\gamma}(t))_{\gamma(t)}.$$  

It turns out that we can find a very convenient expression for this:

$$p_\alpha = \alpha \left( \frac{d}{dt} \ln s(\gamma(t)) + \frac{\langle \gamma(t)_{(k)}, \dot{\gamma}(t)_{(k)} \rangle_{\mathbb{R}^m}}{s(\gamma(t))^2} \right). \tag{11}$$
This is easily checked. In fact, differentiating
\[ \frac{d}{dt} \frac{1}{2} \ln s(\gamma(t))^2 = \frac{1}{s(\gamma(t))^2} \frac{d}{dt} \sum_{j=1}^{k-1} \|\gamma(t)_{(j)}\|_{\mathbb{R}^m}^2 = \frac{1}{s(\gamma(t))^2} \sum_{j=1}^{k-1} \langle \gamma(t)_{(j)}, \dot{\gamma}(t)_{(j)} \rangle_{\mathbb{R}^m} \]
immediately proves (11). We can employ this to find the condition that a curve be horizontal to both translation and scaling. \( \gamma(t) \) must have a constant centroid and a constant shape size. In the language of orthogonal sections (c.f. definition 1.20) we have proven that
\[ S = \{ Z \in M \mid s(Z) = 1, Z_{(k)} = 0 \} \cong \{ Z \in \mathbb{R}^{m \times (k-1)} \mid \sum_{j=1}^{k-1} \|Z_{(j)}\|_{\mathbb{R}^m}^2 = 1 \} = S^{m(k-1)-1} \]
is an orthogonal section for the combined action of scaling and translation (i.e. \( G = \mathbb{R}^m \times \mathbb{R}^+ \)). Indeed, every shape has exactly one representative with centroid 0 and shape size 1. Furthermore, the tangent space \( T_Z S \) is orthogonal to the orbits (i.e. momenta of curves in \( S \) vanish). Finally, \( T_Z S \) has dimension \( m(k-1)-1 \) and therefore fully complements the tangent space of the orbit (with dimension \( m+1 \)) in \( T_Z M \). Note that, restricted on \( S \), the metric becomes the standard metric of \( S^{m(k-1)-1} \) and thus we have proven
\[ M/(\mathbb{R}^m \times \mathbb{R}^+) \cong S^{m(k-1)-1}. \quad (12) \]

Remark 2.1. Arguably, this is a lot of work for a step that may be summarized as: “Without loss of generality, we assume that \( X \) is centred at 0 and normalized to size one.” It is good, however, to give some theoretical justification of why this is equivalent to considering the quotient of our original space. Of course, we would have used the induced euclidean metric on \( S^{m(k-1)-1} \) anyway. However, we shall later see that the choice of metric for shape spaces is not always straightforward. Here we have given some justification to the metric (9), as it naturally arises as the quotient metric with respect to an isometric group action. Actually, on \( \mathbb{R}^m \) the standard metric is uniquely singled out as the metric which is invariant under the action of \( SO(m) \). One might argue that this makes the chosen metric ‘even more canonical’. At some point, this discussion becomes pointless, as there is no ‘one canonical metric’ and certainly no measure of ‘canonicity’. In fact, only based on the demand that our action be isometric, there is nothing to stop us using
\[ \eta(W, V)_X = \frac{1}{s(X)^2} \sum_{i,j=1}^{k} A_{ij} \langle W_{(i)}, V_{(j)} \rangle_{\mathbb{R}^m} \]
for some positive definite \( (A_{ij}) \in \mathbb{R}^{k \times k} \). This would put some unusual weights on the different vertices of our \( k \)-ad. If we identify our vertices with certain landmarks in a picture, such a metric might be desirable to impose different degrees of importance on various features. However, this would render the computation of geodesics into a real nightmare and maybe lead us to reconsider our choice of metric. After all, computational convenience is a criterion just as important as ‘canonicity’, even more so in the case of continuous shapes, as we shall see.

2.3 Pre-shape and Shape space

We are now ready to deal with the rotational action of \( SO(m) \). Restricted to the orthogonal section \( S^k_m := S^{m(k-1)-1} \) from the previous section, this is given by
\[ (O; (Z_{(1)}, \ldots, Z_{(k-1)})) \mapsto (O.Z_{(1)}, \ldots, O.Z_{(k-1)}). \]
In accordance with Kendall, we use the following terminology. The quotient
\[ \Sigma^k_m = S^k_m / SO(m) \]
is called the shape space, as every element uniquely determines one shape. The sphere \( S^k_m \) shall be called pre-shape space, as there are still elements representing the same shape. Correspondingly, we refer to elements of these spaces as shapes and pre-shapes. As one might still be concerned about the order in which to perform the quotienting, let us note that the following diagram commutes,

\[ \begin{array}{cccc}
M & \xrightarrow{T_b} & M \\
\pi & & \pi \\
S^k_m / SO(m) & \xrightarrow{\text{Id}} & S^k_m / SO(m)
\end{array} \]

where \( \pi : M \rightarrow S^k_m / SO(m) \) is the canonical quotient map and \( T_b \) denotes translation by \( b \in \mathbb{R}^m \). In other words, a \( k \)-ad \( Z \) is mapped to the same shape before and after the translation. This works because we have removed the non commuting part in discarding the last component of \( Z \) in Kendall coordinates.

Let us now apply our framework of chapter 1 to the new situation of \( G = SO(m) \) acting on \( M = S^k_m \). It is well known that \( SO(m) \) is a compact group.\(^{12}\) Thus, the action is a proper action and all orbits will be closed. The next point on our imaginary check-list is whether the action is free. This puts us in a situation where we need to distinguish between different cases. For \( m = 2 \), i.e. planar \( k \)-ads, we find that the action is indeed free. For higher dimensions\(^{13}\) \( m \geq 3 \), there are actually pre-shapes that have a non trivial isotropy group. To see why this is the case let us consider, for concreteness, \( m = 3 \).

Let \( Z \) be a pre-shape of rank 1, i.e. all columns are scalar multiples of a single unit vector \( \hat{n}_Z \). Acting with an arbitrary rotation with axis \( \hat{n}_Z \) will preserve the pre-shape. Therefore the isotropy group of \( Z \) is non trivial and the action cannot be free. In the case \( m = 2 \), however, this construction does not work. \( O.Z(j) = Z(j) ; j = 1, \ldots, k - 1 \) immediately implies \( O = \text{Id} \). Non-trivial elements of \( SO(2) \) do not have eigenvectors with eigenvalue 1. Hence, the action of \( SO(2) \) on \( S^2_2 \) is free and we find that the quotient can canonically be endowed with a Riemannian metric, as discussed in section 1.5. More generally, it can be shown that the isotropy group of a pre-shape \( Z \in S^k_m \) with rank \( l \leq m - 1 \) is homeomorphic to \( SO(m)/SO(m - l) \). This is known as the Stiefel manifold of orthonormal \( l \)-frames in \( \mathbb{R}^m \). An analogous, infinite dimensional manifold is also found in Younes et al. [35], where it arises in the study of closed continuous curves.

### 2.4 Procrustes analysis

As \( \Sigma^k_m \) is not generally a manifold, we will first treat the more general case, where \( \Sigma^k_m \) does not carry the structure of a Riemannian manifold. The quotient metric (c.f. eq. (6) in section 1.5) for \( \pi(Z_1) , \pi(Z_2) \in \Sigma^k_m \) is given by

\[ d(\pi(Z_1) , \pi(Z_2)) = \min_{O \in SO(m)} d_{S^k_m}(Z_1 , O.Z_2) . \]

\(^{12}\)\( SO(m) \) is a bounded and closed subset of \( \mathbb{R}^{m \times m} \).  
\(^{13}\)The action of \( SO(1) = \{1\} \) is trivial and hardly worth mentioning at all.
Since $S^k_m$ is just a $(m(k - 1) - 1)$-sphere, on which geodesics are given by segments of great circles, $d_{S^k_m}(Z, W)$ is found to be the angle enclosed between $Z, W \in S^k_m$, that is
\[
d_{S^k_m}(Z, W) = \arccos(WZ^t).
\] (13)

Using the fact that $\arccos$ is monotonically decreasing, we find
\[
d(\pi(Z_1), \pi(Z_2)) = \arccos \max_{O \in SO(m)} \text{tr}(OZ_2Z_1^t). \tag{14}
\]

This optimization problem can be solved explicitly by using a technique known as Procrustes analysis (see also Kendall’s original treatment in [16]). To do this, one needs the pseudo singular value decomposition of the $m \times m$ matrix $Z_2Z_1^t$. This is a decomposition of the form
\[
Z_2Z_1^t = U\Lambda V^t,
\]
where $U, V \in SO(m)$ and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_m)$ with the pseudo singular values\footnote{The usual singular value decomposition needs $U, V \in O(m)$ and requires that all $\lambda_i$ be non negative.}

\[
\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{m - 1} \geq |\lambda_m|.
\]

The sign of $\lambda_m$ is determined by $\text{sign}(\det(Z_2Z_1^t))$. Employing this and the cyclic property of the trace, we recast eq. (14) to
\[
d(\pi(Z_1), \pi(Z_2)) = \arccos \max_{O \in SO(m)} \text{tr}(VOU\Lambda) = \arccos \max_{R \in SO(m)} \text{tr}(RA\Lambda).
\]

We now claim that $\max_{R \in SO(m)} \text{tr}(RA\Lambda) = \sum_{i=1, \ldots, m} r_{ii}\lambda_i$ is attained for $R = \text{Id}$, i.e. $O = (UV)^t$. To see this, we need one further result about the diagonal entries of rotational matrices $R \in SO(m)$.

**Theorem 2.2** ([12], theorem 8). A vector $(r_1, \ldots, r_m)$ is the diagonal of a matrix $R \in SO(m)$ if and only if it lies in the convex hull of
\[
\mathcal{E} = \{(\pm 1, \ldots, \pm 1) \mid \text{even number of minus signs}\}.
\]

This reduces our problem to a standard linear optimization problem on a convex set. It is known that, if a solution exists, it is attained at one of the extreme points of its domain. This is the basis for the famous simplex algorithm by Dantzig (c.f. [8]). Note that $\mathcal{E}$ is the set of extreme points in the convex compact set of diagonal entries of rotational matrices. Taking into account the signs of the $\lambda_i$, it is now clear (without employing the simplex algorithm) that the maximum is attained for $r_{ii} = 1, i = 1, \ldots, m$. We repeat and summarize our result in the following theorem.

**Theorem 2.3.** Let $Z_1, Z_2 \in S^k_m, 2 \leq m \leq k - 1$, be pre-shapes and let $Z_2Z_1^t = U\Lambda V$ be a pseudo singular value decomposition with $U, V \in SO(m)$ and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_m), \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{m - 1} \geq |\lambda_m| \geq 0$. Then $Z_1$ and $Z_2$ are optimally registered for $O^* = (UV)^t$ and the generalized geodesic distance between the shapes $\pi(Z_1)$ and $\pi(Z_2)$ is given by
\[
d(\pi(Z_1), \pi(Z_2)) = d_{S^k_m}(Z_1, O^*Z_2) = \arccos(\text{tr}\Lambda), \tag{15}
\]
with $\text{tr}\Lambda \geq 0$. In particular, $d(\pi(Z_1), \pi(Z_2))$ is bounded by $\pi/2$.

Having found the optimal alignment between two pre-shapes, we can also compute the generalized geodesic connecting the corresponding shapes. According to proposition 1.19, the length minimizing arc between $Z_1$ and $O^*Z_2$ is a horizontal geodesic. Therefore, its projection to $\Sigma^k_m$ is a generalized geodesic, connecting $\pi(Z_1)$ and $\pi(Z_2)$.
2.5 Planar shapes: Analysis of $\Sigma^k_2$

In later chapters we will discuss shape representations which mainly focus on planar shapes. This brings the advantage that one may identify $\mathbb{R}^2 \simeq \mathbb{C}$, simplifying many calculations and allowing operations that do not exist in higher dimensions, such as taking a square root of a vector. This is also advantageous for Kendall’s shape space. Furthermore, the two dimensional case is interesting as it allows a Riemannian structure on the quotient (we have already seen this at the end of section 2.3).

We start with $S^k_2$, as in section 2.3, but we identify each column $Z_{(j)} \in \mathbb{R}^2$ with a single complex number $z_j$. This way $S^k_2$ becomes a complex sphere

$$S^k_2 \simeq \{ z = (z_1, \ldots, z_{k-1}) \in \mathbb{C}^{k-1} \mid \sum_{j=1}^{k-1} |z_j|^2 = 1 \},$$

where $|z_j|$ denotes the absolute value of the complex entry $z_j$. The Riemannian metric (9) turns into the real part of the standard Hermitian inner product

$$\eta(x, y)_z = \Re \langle x, y \rangle_{C^{k-1}} = \Re \left( \sum_{j=1}^{k-1} x_j \bar{y}_j \right)$$

and geodesics are still given by segments of great circles. In particular, the geodesic distance between $z_1, z_2 \in S^k_2$ is the enclosed angle

$$d_{S^k_2}(z_1, z_2) = \arccos \Re (\langle z_1, z_2 \rangle).$$

Finally, the action of the rotation group $SO(2) = U(1) = S^1$ is just a (complex) scalar multiplication $(e^{i\phi}, z) \mapsto e^{i\phi} z$. It is now easy to see that the quotient space $\Sigma^k_2$ is nothing but the complex projective space $^{15} \mathbb{CP}^{k-2}$. The metric on $\mathbb{CP}^{k-2}$ that arises from our quotient construction is called the Fubini study metric. In complex differential geometry this is a well known metric. We will not derive this, however, as the actual form of the metric is not needed for the computation of geodesics. Of course, if we were interested in further geometric properties such as curvature, a detailed study of the metric would be inevitable. Please refer to Kendall’s work in [17] and [16] for details. Returning to the rotational action, we see an immense simplification, as we can perform the optimization over $SO(2)$ without a pseudo singular value decomposition. Indeed, we compute the generalized geodesic distance between two shapes $\pi(z_1), \pi(z_2)$ as

$$d(\pi(z_1), \pi(z_2)) = \min_{\phi \in S^1} d_{S^k_2}(z_1, e^{i\phi} z_2)$$

$$= \min_{\phi \in S^1} \arccos \Re \left( e^{-i\phi} \langle z_1, z_2 \rangle \right)$$

$$= \arccos \max_{\phi \in S^1} \Re \left( e^{-i\phi} \langle z_1, z_2 \rangle \right)$$

$$= \arccos |\langle z_1, z_2 \rangle|,$$

where the minimum is attained for $\phi^* = \arg \langle z_1, z_2 \rangle$. This is straightforward to implement. We present some examples of geodesics obtained this way in figure 5. The first four rows show some examples of how polygons are deformed into one another. The calculations

$^{15}$i.e. the space of $\mathbb{C}$-one dimensional subspaces of $\mathbb{C}^{k-1}$. 

24
were performed using the vertices only. Connecting lines were added afterwards to make the changes more visible. The last two rows show continuous shapes. To make Kendall’s shape analysis applicable to these, we used a finite number of sample points (the same for each pair) and again connected the sample points afterwards.

It is also interesting to make a comparison to geodesics between pre-shapes that are not optimally aligned, i.e. non horizontal geodesics in the pre-shape space. This can be seen in figure 6. The first row shows the horizontal geodesic between two almost identical triangles.\textsuperscript{16} The second row takes the pre-shapes without rotational alignment and shows a geodesic between them. Once more, this illustrates the quotienting process in our mathematical framework.

\subsection*{2.6 Limitations of landmark based shape analysis}

It is hard to judge the performance of Kendall’s shape analysis, as there is no general objective to do so. Depending on the application, the performance may be very good,

\textsuperscript{16}‘Almost’, as the upper right vertex is marginally displaced. The distance is actually not zero.
especially when the shapes naturally arise as $k$-ads. For instance, one could imagine Kendall’s shape analysis to be useful for classification of stellar constellations. Kendall himself studied the non accidental alignment of triangles, arising in archaeological surveys of stone monuments (c.f. [16]). One may even leave these restricted applications and attribute vertices to the position of different features or landmarks in a picture or higher dimensional geometric object. Conversely, applying the theory to continuous shapes may yield results that are not very intuitive, e.g. the last two geodesics in figure 5 show deformations that do not seem very ‘natural’. The human eye immediately recognizes both shapes in the last row as comic elephants. The geodesic distance, however, is not much smaller than the geodesic distance between the shapes in the row above.

Regardless of the application, there is one particularly crucial point in landmark based shape analysis (such as Kendall’s theory): the selection and registration of landmarks. In our construction of $\Sigma^k_m$ we started with the set of ordered and not totally degenerate $k$-ads $\mathbb{R}^{m \times k} \setminus \{0\}$. Therefore, we need to select an order among vertices of the $k$-ads. If our shape is the contour of some image, this is not difficult, as we only need to follow the path around the shape. However, we need to select a first vertex, if our shape is closed. In Kendall’s shape representation there is no way to tell whether a shape is closed or not. In applications, however, the shape may arise as a closed contour and there is no natural first vertex. We illustrate this in figure 7. Both pictures show geodesics between identical shapes, but as we select different first vertices on the shapes, we draw a mathematical distinction between them. A similar problem one has to address is how to obtain samples from a given continuous contour. One way to deal with this is to select equally spaced points on an arc-length parametrized curve, i.e. if $\alpha_1 : [0, L_1] \rightarrow C$ and $\alpha_2 : [0, L_2] \rightarrow C$ are arc-length parametrized, we select the vertices $\alpha_1(jL_1/(k-1)), \alpha_2(jL_2/(k-1)), j = 0, \ldots, k-1$, to obtain elements in $\mathbb{R}^{2 \times k} \setminus \{0\}$ for further processing. However, similarly to the selection of the first vertex, this will generally result in a non optimal matching of vertices between shapes. Of course, we have no definition of what an optimal registration of vertices should be. Nevertheless, it seems desirable to match as many common features between two shapes, while preserving the order of sampling points on the contour. If we fix the parametrization, this cannot be achieved. The sketches in figure 8 illustrate this further. 8(a) and 8(b) show how vertices are selected equally spaced. As the actual shapes only differ in the extension at the far right, a vertex selection as in 8(c) seems more appropriate, matching more common features between the two shapes. A similar problem can be observed, when dealing with continuous shape representations. Fixing the parametrization

Figure 7: Examples of bad vertex registration

\begin{table}
\centering
\begin{tabular}{|c|ccccc|}
\hline
Geodesic & B & B & B & B & B \\
\hline
Distance & 0.4478 \\
\hline
\end{tabular}
\end{table}
Figure 8: Illustration why fixing the parametrization to arc-length is undesirable

to arc-length has the same undesirable effect. The solution to this will be to perform an additional quotienting step, which removes the curve parametrization without fixing it to arc-length. Similarly, one might perform an additional optimization for $k$-ads, by searching for the optimal registration of vertices before applying Kendall’s analysis. Even though this does not solve the problem of vertex selection on a contour, it could avoid bad vertex registration as in figure 7. This is especially so for closed contours with fixed parametrization, when we only need to find the optimal first vertex, this is easily carried out and avoids problems as these.
3 Metrics on the Space of Continuous Curves

We have seen in the previous chapter how the theoretical apparatus developed in chapter 1 can be applied to problems in shape analysis. We shall now leave Kendall’s landmark based approach and turn to continuous shape representations. This means that in addition to the non linearity of the relevant spaces, we also add the difficulty of infinite dimensionality. But we should not only see this as a difficulty, but also as a chance to overcome the problems and limitations of landmark based analysis. Passing to continuous curves will allow us to perform what is called an elastic shape analysis (c.f. [26, 33]). This will exploit the freedom to reparametrize continuous curves at will, and help improve the feature matching between shapes. As we pass to continuous curves we also reach a point where the complexity of the underlying theory, sadly, goes far beyond the level of this dissertation. The author tries to keep the analysis as rigorous as possible, adopting and reproducing many ideas that where recently published by P. Michor, D. Mumford and others, c.f. [6, 22, 23, 24, 35]. However, it will be unfeasible to work through the proofs. The guiding principle will be to present the theory, as if what developed in chapter 1 were still applicable, and cite the correct results found in the literature. The objective for this chapter is to introduce a convenient setting for an analysis of continuous shapes and to conduct a short survey on possible choices of Riemannian metrics. This will also motivate the choice of metric considered in chapter 4.

The rough structure goes as follows. Section 3.1 will be devoted to setting up the appropriate spaces of curves, i.e. spaces of embeddings and immersions of the unit interval or unit circle. In this context, we shall also investigate the action of the reparametrization group and the differential structure of the corresponding orbit spaces. The following section then introduces the Riemannian structure by discussing a simple class of metrics, so-called almost local metrics. Once we get used to Riemannian manifolds of curves, we proceed to discuss more advanced choices of metric, so called Sobolev type metrics, introducing higher derivatives into the metric, much as in the theory of Sobolev spaces that arose in the study of PDEs. In particular, we shall introduce Riemannian metrics that only involve first order derivatives. This will finally lead us to the so called elastic metrics recently used in shape analysis (c.f. [35, 26]).

3.1 Manifolds of curves and the reparametrization action

Our first task in developing a theory of continuous shape analysis is to set up the appropriate spaces in which our shapes shall eventually live. Much as the space of $k$-ads in $\mathbb{R}^m$ was a convenient place to start the construction of Kendall’s shape space, we shall now take the space of embeddings of $S^1$ into $\mathbb{R}^2$

$$\text{Emb}(S^1, \mathbb{R}^2) = \{c : S^1 \to \mathbb{R}^2 \mid c \text{ is an embedding}\}$$

as our starting point. Equivalently, this is the space of all simple closed parametrized curves in $\mathbb{R}^2$. This space is naturally contained in the space of immersions of $S^1$ into $\mathbb{R}^2$, denoted as $\text{Imm}(S^1, \mathbb{R}^2)$. This also allows the crossing of curves, e.g. a figure eight graph. Both spaces are smooth submanifolds of the Fréchet manifold$^{19} C^\infty(S^1, \mathbb{R}^2)$. It is natural to ask why we should only consider closed curves. We shall, indeed, only deal with open curves in chapter 4, as it will turn out to be much easier for the specific situation considered there. To keep things simple in this chapter, we will restrict to closed curves.

$^{19}$Please refer to Omori [27] for a detailed description of such spaces and their differential structure.
The case of open curves can be treated by either allowing for one possible discontinuity on the curve or, equivalently, replacing $S^1$ by the unit interval $[0,1]$. This is just to avoid spelling out every statement for open and closed curves. One might also ask why we should only consider smooth $C^\infty$ curves. Indeed, we might consider the whole hierarchy $C^0 \supset C^1 \supset C^2 \supset \ldots \supset C^\infty$ or maybe even more exotic spaces, such as $L^2$ or absolutely continuous functions. All these spaces contain the core space of $C^\infty$ functions. Much as Sobolev spaces in the theory of partial differential equations, Michor and Mumford speculate in [23] that the ‘most natural shape spaces’ arise as a completion of $C^\infty$ curves, with respect to a chosen metric. We shall follow their example and focus on the study of smooth curves. As an example of such a ‘natural’ shape space, we will explicitly carry out the completion in the particularly easy situation of chapter 4 (section 4.1).

Like the preliminary space of $k$-ads in chapter 2, the spaces $\text{Emb}(S^1, \mathbb{R}^2)$ and $\text{Imm}(S^1, \mathbb{R}^2)$ have some superficial degrees of freedom for the purpose of shape description. By now, we are already familiar with the actions of translations, rotations and scaling, and it is obvious how they act. A new shape preserving transformation is the reparametrization of curves. Thus, we have an action

$$\text{Emb}(S^1, \mathbb{R}^2) \times \text{Diff}(S^1) \to \text{Emb}(S^1, \mathbb{R}^2); (c, \gamma) \mapsto c \circ \gamma$$

and a similar one for $\text{Imm}(S^1, \mathbb{R}^2)$ and $C^\infty(S^1, \mathbb{R}^2)$. Following our usual procedure, we would now like to pass to the quotient spaces

$$B_c := \text{Emb}(S^1, \mathbb{R}^2)/\text{Diff}(S^1)$$

$$B_\iota := \text{Imm}(S^1, \mathbb{R}^2)/\text{Diff}(S^1)$$

Naturally, the question arises what kind of structure these spaces carry. The most basic structure we can equip them with is the quotient topology, descending from the larger spaces $\text{Emb}(S^1, \mathbb{R}^2)$ and $\text{Imm}(S^1, \mathbb{R}^2)$. Are these spaces Hausdorff? Are orbits of the reparametrization action closed?

**Remark 3.1.** All results in this chapter refer to the strong topology, which $\text{Emb}(S^1, \mathbb{R}^2)$ and $\text{Imm}(S^1, \mathbb{R}^2)$ naturally inherit as open subsets of the Fréchet space $C^\infty(S^1, \mathbb{R}^2)$. This topology has a very restrictive notion of convergence. A sequence of curves converges to a limit curve if and only if the curves and all their derivatives converge uniformly to the limit curve and its derivatives, respectively. In the following sections we will define weak Riemannian metrics on these spaces of curves. They induce a new metric topology which is not equivalent to the Fréchet topology. Thus, it is wise to take all the topological results about our spaces of curves with a pinch of salt, as they may not be true for weak metrics.

Consider the following example, taken from Younes et al. [35], section 3.5.

**Example 3.2.** Let $\psi : [0,1] \to [0,1]$ be a non decreasing smooth map $\psi(0) = 0, \psi(1) = 1$. Furthermore, let $\psi$ be constant on some small interval $I \subset [0,1]$. $\psi$ is not a diffeomorphism, but we can build a sequence of diffeomorphisms of $[0,1]$ 

$$\psi_n(\theta) = \left(1 - \frac{1}{n}\right) \psi(\theta) + \frac{\theta}{n},$$

which converges to $\psi$ as $n \to \infty$.\(^\text{20}\) Now, take any immersion $c \in \text{Imm}([0,1], \mathbb{R}^2)$ and consider the reparametrizations $c_n := c \circ \psi_n$. We have $c_n \to c \circ \psi$ as $n \to \infty$ in the topology

\(^{20}\)To be precise, $\psi_n$ converges to $\psi$ uniformly in all derivatives. As a Fréchet-Lie group $\text{Diff}([0,1])$ carries a very similar topology as $C^\infty(I, \mathbb{R}^2)$

29
of $C^\infty([0, 1], \mathbb{R}^2)$, i.e. $c_n$ and all derivatives converge uniformly to $c \circ \psi$ and its derivatives, respectively. By construction, $c \circ \psi$ is constant on the subinterval $I$ and therefore is no longer an immersion. Hence, $c \circ \psi$ and $c$ are not related by a diffeomorphism.

This proves the following proposition:

**Proposition 3.3.** Let $c$ be an immersion of $[0, 1]$ into $\mathbb{R}^2$ and let $[c] = \{c \circ \gamma \mid \gamma \in \text{Diff}([0, 1])\}$ be the orbit under the action of the reparametrization group. Then $[c]$ is not closed in $C^\infty([0, 1], \mathbb{R}^2)$.

A similar result holds, of course, for immersions of $\mathbb{S}^1$ as $\text{Diff}([0, 1])$ is naturally contained in $\text{Diff}(\mathbb{S}^1)$ as a subgroup. What does this mean for our quotient? Looking back to lemma 1.4 and the subsequent remark about the connection of closed orbits with the Hausdorff property of the quotient, we might fear that any hope of constructing a metric is futile. However, browsing the literature we find a seemingly contradicting result.

**Theorem 3.4** ([6], theorem 2.1). Let $M, N$ be connected finite dimensional manifolds $\dim(M) \leq \dim(N)$ and let $\text{Imm}_{\text{prop}}(M, N)$ denote the space of all proper\(^{21}\) immersions of $M$ into $N$. Then the orbit space of $\text{Imm}_{\text{prop}}(M, N)$ under the action of $\text{Diff}(M)$ is Hausdorff in the quotient topology.

Note that the technical condition of a proper immersion has no relevance for our considerations, as $\mathbb{S}^1$ and $[0, 1]$ are compact and therefore any immersion is proper. So how do proposition 3.3, lemma 1.4 and theorem 3.4 go together?

Splendidly! There is no contradiction. Proposition 3.3 refers to the larger space $C^\infty([0, 1], \mathbb{R}^2)$, not to the space of immersions. Thus, care has to be taken, whenever we wish to extend our considerations to the space $C^\infty([0, 1], \mathbb{R}^2)$. For example, this is done in Younes et al. [35]. The reason why one might want to do that is that the space of immersions can be geodesically incomplete (depending on the metric). This will also be the case in chapter 4 where we will undertake a considerable effort to deal with this problem.

We are next interested in the question of whether $B_c$ and $B_\hat{c}$ have a differential structure compatible with the one on $\text{Emb}(\mathbb{S}^1, \mathbb{R}^2)$ and $\text{Imm}(\mathbb{S}^1, \mathbb{R}^2)$. We would like to use something like proposition 1.7, i.e. we would like to check whether $\text{Diff}(\mathbb{S}^1)$ acts freely and properly. Indeed, it turns out that we can establish both properties. In [23] Michor and Mumford assert that the map

$$\text{Imm}(\mathbb{S}^1, \mathbb{R}^2) \times \text{Diff}(\mathbb{S}^1) \longrightarrow \text{Imm}(\mathbb{S}^1, \mathbb{R}^2) \times \text{Imm}(\mathbb{S}^1, \mathbb{R}^2), (c, \gamma) \mapsto (c, c \circ \gamma)$$

is a proper map. Furthermore, they present the following handy criterion to investigate whether $\gamma \in \text{Diff}(\mathbb{S}^1)$ acts trivially on a given immersion.

**Lemma 3.5** ([23], section 2.4). If $\gamma \in \text{Diff}(\mathbb{S}^1)$ has a fixed point and if $c \circ \gamma = c$ for some $c \in \text{Imm}(\mathbb{S}^1, \mathbb{R}^2)$, then $\gamma = \text{Id}$.

Now suppose that $c$ is an immersion whose image has a point $p \in c(\mathbb{S}^1)$ with only one pre image. Moreover, let $\gamma$ be a trivial acting diffeomorphism $c \circ \gamma = c$. Then $x_0 = c^{-1}(p) \in \mathbb{S}^1$ is a fixed point of $\gamma$ and lemma 3.5 implies $\gamma = \text{Id}$, i.e. $\text{Diff}(\mathbb{S}^1)$ acts freely on $c$. In particular, the action of $\text{Diff}(\mathbb{S}^1)$ on $\text{Emb}(\mathbb{S}^1, \mathbb{R}^2)$ is free, as all embeddings are injective. It is still a subject of current research whether $\text{Diff}(\mathbb{S}^1)$ acts freely on any immersion. So

\(^{21}\)A function between topological spaces is a proper map if every pre image of a compact set is compact.
far, no counterexample is known. Furthermore, the singularities that can occur are of a ‘rather mild’ nature, in the sense that the isotropy groups that can occur are finite. For further information on this, consult Cervera et al. [6]. In any case, Imm$(S^1, \mathbb{R}^2)$ qualifies for the ‘generalized treatment’ of section 1.6. For the space Emb$(S^1, \mathbb{R}^2)$ we have the very pleasing result.\footnote{The statement uses the slightly different language of principal fibre bundles, structure groups etc. To avoid introducing these concepts in full generality, some rephrasing was done, limiting the assertion to what we are interested in.}

**Theorem 3.6** (Michor & Mumford [23], section 2.3(A)). The action of Diff$(S^1)$ on Emb$(S^1, \mathbb{R}^2)$ is proper and free. The quotient space $B_e = \text{Emb}(S^1, \mathbb{R}^2)/\text{Diff}(S^1)$ is a smooth Fréchet manifold and the quotient map $\pi : \text{Emb}(S^1, \mathbb{R}^2) \to B_e$ is a submersion.

### 3.2 The natural splitting and almost local metrics

To introduce Riemannian metrics on Imm$(S^1, \mathbb{R}^2)$, Emb$(S^1, \mathbb{R}^2)$, $C^\infty(S^1, \mathbb{R}^2)$ and $B_e$, we first need to talk about the tangent spaces of these manifolds, i.e. their infinitesimal versions. To that account, let $c(\cdot, \cdot) : (-\epsilon, \epsilon) \times S^1 \to \mathbb{R}^2$, $(t, \theta) \mapsto c(t, \theta)$ be a path of curves\footnote{It seems convenient to say path of curves, rather than curve of curves. Therefore, let us reserve the word curve for elements in Emb$(S^1, \mathbb{R}^2)$, Imm$(S^1, \mathbb{R}^2)$, etc.} with $c(0, \cdot) = c_0 \in C^\infty(S^1, \mathbb{R}^2)$. Tangent vectors to $c_0$ arise as derivatives of such curves, i.e. $\theta \mapsto \partial_\theta c(0, \theta)$ is a tangent vector. However, this is nothing but another smooth map $S^1 \to \mathbb{R}^2$, an element of $C^\infty(S^1, \mathbb{R}^2)$. Hence we find $T_{c_0}C^\infty(S^1, \mathbb{R}^2) \simeq C^\infty(S^1, \mathbb{R}^2)$. As Imm$(S^1, \mathbb{R}^2)$ and Emb$(S^1, \mathbb{R}^2)$ are open subsets of $C^\infty(S^1, \mathbb{R}^2)$, they have the same tangent spaces. We can do the same for Diff$(S^1)$ to find that its tangent algebra $T_{\text{id}}\text{Diff}(S^1)$ is given by the space $\mathfrak{X}(S^1)$ of vector fields on $S^1$. Indeed, if we let $\gamma(t, \theta)$ be a path of diffeomorphisms with $\gamma(0, \cdot) = \text{Id}$ then for each fixed $\theta$, $t \mapsto \gamma(t, \theta)$ is a smooth path through $\theta$. Thus, $\partial_\theta \gamma(0, \theta) \in T_\theta S^1$ for all $\theta$ and we see that $\theta \mapsto \partial_\theta \gamma(0, \theta)$ is a smooth vector field on $S^1$. As $S^1$ is of dimension one, we can further identify $\mathfrak{X}(S^1) \simeq C^\infty(S^1, \mathbb{R})$, i.e. we identify infinitesimal changes to diffeomorphisms of $S^1$ with smooth functions $\Delta \gamma : S^1 \to \mathbb{R}$.

In order to get an idea what the tangent space $T_{[c]}B_e$ looks like, we recall that proposition 1.7 and the analogous application to Emb$(S^1, \mathbb{R}^2)$, theorem 3.6, establish the identification

$$
(T_c \text{Emb}(S^1, \mathbb{R}^2))/(T_c[c]) \simeq T_{[c]}B_e. \tag{16}
$$

Let us therefore investigate tangent spaces of orbits under the diffeomorphism group $T_{[c]}$ for $c \in \text{Emb}(S^1, \mathbb{R}^2)$. As above, let $\gamma(t, \theta)$ be a path of diffeomorphisms of $S^1$ with $\gamma(0, \cdot) = \text{Id}$. Then $t \mapsto (\theta \mapsto c(\gamma(t, \theta)))$ is a path through $c$ within the orbit $[c]$. Differentiation by $t$ at $t = 0$ yields

$$
\partial_t c(\gamma(t, \theta))|_{t=0} = c'(\gamma(0, \theta)) \cdot \partial_\theta \gamma(t, \theta)|_{t=0} = c'(\theta) \cdot \partial_\theta \gamma(t, \theta)|_{t=0}
$$

where $c'(\theta) = \partial_\theta c(\theta)$. We see that the mapping $\theta \mapsto c'(\theta) \cdot \partial_\theta \gamma(t, \theta)|_{t=0}$ is everywhere tangent to the original curve $c$ itself. Turning back to (16), we can identify $T_{[c]}B_e$ with the space of vector fields $X : S^1 \to \mathbb{R}^2$ modulo vector fields that are tangent to $c$. Of course, there is nothing to stop us from choosing a representative for each such class of vector fields. The most ‘natural’ choice for such a representative seems to be the vector field $X$ that is everywhere perpendicular to the curve $c$, i.e. $\langle X(\theta), \partial_\theta c(\theta) \rangle_{\mathbb{R}^2} = 0$ for all $\theta \in S^1$. This provides us with a very natural notion when an element of $T_c \text{Emb}(S^1, \mathbb{R}^2)$ is
horizontal to the orbit \([c]\). In the literature this is referred to as the natural splitting, c.f. [24].

**Example 3.7.** Let \(c(t, \theta) = r(t)(\cos(\theta), \sin(\theta))\) be a path of concentric circles with a smooth radius function \(r(t) > 0\). On the one hand, the tangent vector field to each curve \(c(t, \cdot)\) is given by \(\theta \mapsto \partial_\theta c(t, \theta) = r(t)(-\sin(\theta), \cos(\theta))\). On the other hand, we have the tangent to the path \(t \mapsto c(t, \cdot)\). For every \(t\) this is given by \(\theta \mapsto \partial_t c(t, \theta) = r'(t)(\cos(\theta), \sin(\theta))\) and constitutes a vector field along \(c\), which is normal to \(c\). Therefore, we have \(\langle \partial_t c(t, \theta), \partial_\theta c(t, \theta) \rangle_{\mathbb{R}^2} = 0\). In the natural splitting, this is what we mean by a horizontal curve (with respect to reparametrization).

Previously, we defined the notion of horizontality using a Riemannian metric on our manifold \(M\), in this case \(\text{Emb}(\mathbb{S}^1, \mathbb{R}^2)\). This enabled us to decide if a tangent vector was orthogonal to an orbit or not. It is now natural to ask how these two definitions of horizontality relate to each other. Are there any metrics that support this natural splitting, i.e. for which both concepts of horizontality coincide? Indeed, there is a whole family of such metrics.

Arguably, the simplest metric we know for spaces of functions is the constant \(L^2\) metric

\[
G(h, k)_c = \int_{\mathbb{S}^1} \langle h(\theta), k(\theta) \rangle d\theta
\]

for \(h, k \in T_{c} \text{Emb}(\mathbb{S}^1, \mathbb{R}^2) \simeq C^\infty(\mathbb{S}^1, \mathbb{R}^2)\). The translational action of \(\mathbb{R}^2\) is not felt in the tangent space, so the metric is invariant under it. Furthermore, the \(\mathbb{R}^2\) inner product provides invariance under the \(SO(2)\) rotational action. Unfortunately, this metric is not invariant under the action of scaling and reparametrization. However, we can establish reparametrization invariance by introducing the invariant volume form \(|c'(\theta)|d\theta\).

This leads us to consider

\[
G^0(h, k)_c = \int_{\mathbb{S}^1} \langle h(\theta), k(\theta) \rangle |\partial_\theta c(\theta)|d\theta. \tag{17}
\]

To get used to such metrics on spaces of curves, let us check that this metric indeed provides the same notion of horizontality as the natural splitting.

**Example 3.8 (Reparametrization momentum in \(G^0\) metric).** We apply the framework of section 1.4 and look for the momentum associated with the action of reparametrization. Let \(\Delta \gamma \in \mathfrak{X}(\mathbb{S}^1) \simeq C^\infty(\mathbb{S}^1, \mathbb{R})\) be a small change\(^{24}\) of \(\text{Id}_{\mathbb{S}^1}\). Its corresponding Killing vector field on \(\text{Emb}(\mathbb{S}^1, \mathbb{R}^2)\) is formally given by

\[
X^{\Delta \gamma} : \text{Emb}(\mathbb{S}^1, \mathbb{R}^2) \longrightarrow T\text{Emb}(\mathbb{S}^1, \mathbb{R}^2); c \mapsto \left. \frac{d}{d\tau} \right|_{\tau = 0} (c \circ \exp_{\text{Diff}(\mathbb{S}^1)}(\tau \Delta \gamma))
\]

This somewhat stiff expression can be simplified by noting that \(\exp_{\text{Diff}(\mathbb{S}^1)}(\tau \Delta \gamma)(\theta)\) for small \(\tau\) is nothing other than \(\theta + \tau \Delta \gamma(\theta)\). Hence application of the chain rule yields

\[
X^{\Delta \gamma}(c)(\theta) = c'(\theta) \cdot \Delta \gamma(\theta),
\]

which is the tangent vector field to \(c\), as a curve in \(\mathbb{R}^2\), times a scalar function. For a path of curves \(t \mapsto c(t, \cdot)\), we find the reparametrization momentum

\[
G^0(\partial_t c, X^{\Delta \gamma}(c)) = \int_{\mathbb{S}^1} \langle \partial_t c(t, \theta), \partial_\theta c(t, \theta) \rangle \Delta \gamma(\theta) |\partial_\theta c(t, \theta)| d\theta.
\]

\(^{24}\)In the general set-up of section 1.4 we denoted such elements of the tangent algebra as \(\xi \in T_{c} G\).
Thus, for a curve to be horizontal to reparametrization orbits, this must vanish identically for all \( \Delta \gamma \in C^\infty(S^1, \mathbb{R}) \). We finally find the condition \( \langle \partial_t c(t, \theta), \partial c(t, \theta) \rangle = 0 \) for all \( t, \theta \), which is already familiar from example 3.7.

Now, having appreciated the simplicity of this metric for a moment, is \( G^0(\cdot, \cdot) \) the best and end-all of Riemannian metrics for \( \text{Emb}(S^1, \mathbb{R}^2) \)? It is invariant under rotation, translation and reparametrization and therefore would naturally suit the space \( B_e \) and quotients of it. However, as we pass this metric onto our quotient \( B_e \) there arises a major difficulty which we have not encountered so far. Recall that the metric distance between two curves \( c_0, c_1 \in \text{Emb}(S^1, \mathbb{R}^2) \) is defined as in section 1.3, eq. (1), as the infimum path length between them

\[
d_{\text{Emb}}(c_0, c_1) = \inf \{ L[\alpha] = \int_0^1 G^0(\partial_t \alpha, \partial_t \alpha)^{1/2} dt \mid \alpha \text{ a connecting path in } \text{Emb}(S^1, \mathbb{R}^2) \}
\]

and that the quotient metric on \( B_e \) is given similarly by

\[
d_{B_e}([c_0], [c_1]) = \inf \{ L[\pi \circ \alpha] = \int_0^1 G^0(\partial_t \alpha^\perp, \partial_t \alpha^\perp)^{1/2} dt \mid \alpha \text{ a connecting path in } \text{Emb}(S^1, \mathbb{R}^2) \}
\]

where we consider only the horizontal contribution \( \partial_t \alpha^\perp \) of a path. Quite surprisingly, this infimum is found to be zero! We have the following theorem.

**Theorem 3.9** ([23], section 3.10). For \( c_0, c_1 \in \text{Imm}(S^1, \mathbb{R}^2) \) there always exists a path \( t \mapsto c(t, \cdot) \) with \( c(0, \cdot) = c_0 \) and \( \pi(c(1, \cdot)) = \pi(c_1) \) such that \( L[\pi \circ \alpha] \) is arbitrarily small.

Note that we asserted in theorem 3.6 that the action of the diffeomorphism group is quite well behaved. The reason for this ‘strange’ behaviour now, is that we are only considering weak Riemannian metrics. As mentioned in section 1.7, the inner products on the tangent spaces defined by these metrics are incomplete. As a consequence, the induced geodesic distance features all sorts of ‘unusual’ behaviour. One such is the vanishing geodesic distance in special cases. As such a quotient metric is useless in shape analysis, we need a way to ‘strengthen’ the metric, i.e. prevent it from becoming degenerate when passing to the quotient. Strong Riemannian metrics, that are entirely compatible with the topological structure of our Fréchet manifold do not exist, c.f. Clarke [7]. However, we can modify \( G^0 \) to overcome the degeneracy and define an actual metric distance on our spaces of curves. One way to achieve this is to introduce a function \( \Phi(c) \)

\[
G^\Phi(h, k)_c = \int_{S^1} \langle h, k \rangle \Phi(c)|c'(\theta)| d\theta.
\]

By \( \Phi(c) \) we mean that it may involve very general expressions which are derived from \( c \). Such derived expressions could be the length \( l(c) \) or a curvature term \( \kappa_c = \frac{\langle c''(\theta), n(\theta) \rangle}{|c'(\theta)|^2} \), where \( n(\theta) \) is the normal vector to \( c \) at \( \theta \). All these metrics have the property that they provide the same horizontal splitting as \( G^0 \). The argument in example 3.8 works just as well for \( G^\Phi \). The family of such metrics is referred to as almost local metrics, c.f. [24]. Particularly appealing are choices that depend only on \( l(c) \), as the \( \Phi \) factor can be taken out of the integral and makes the metric conformally related to \( G^0 \). Such choices are studied by J. Shah in [31]. It seems that one can remove the degeneracy this way, at least in certain cases. Especially interesting would be a scale invariant choice \( \Phi(c) = l(c)^{-3} \). To the knowledge of the author, there are currently no results dealing with this particular
choice of metric. Another choice for $\Phi$ is $\Phi(c) = 1 + A\kappa_c(\theta)^2$, for some dimensional constant $A > 0$. This has intensively been studied by Michor and Mumford in [23]. It successfully produces a non-degenerate metric on the quotient $B_c$. However, it is not invariant to scaling. This could be cured by a choice $\Phi(c) = l(c)^{-3} + A|\kappa|l(c)^{-1}$, which was proposed in [24].

3.3 Sobolev type metrics

Another way to overcome the difficulty of vanishing geodesic distance is found by introducing higher derivatives of $h, k \in T_c\text{Emb}(S^1, \mathbb{R}^2)$ into the metric. As this idea closely relates to the inner products of Sobolev spaces that arise in PDEs (c.f. L.C. Evans [9]), this kind of metrics are referred to as Sobolev type metrics. Under action of reparametrization such derivatives transform as

$$\partial_\theta (h \circ \gamma)(\theta) = h'(\gamma(\theta)) \cdot \partial_\theta \gamma(\theta).$$

Thus, simply introducing $\partial_\theta$ factors into the metric will not give an invariant action. However, if we divide by $|\partial_\theta c|$ this will cancel the Jacobian factor. This is the same as if we differentiate with respect to the arc length parameter. We will write this derivative as

$$D_s h := \frac{\partial_\theta h}{|\partial_\theta c|}.$$ 

Note that, whereas partial derivatives of smooth functions always commute, we have to be careful with the arc length derivative and derivatives to other parameters. For instance, we often differentiate with respect to the parameter $t$ of a path $c(t, \cdot)$ in $\text{Emb}(S^1, \mathbb{R}^2)$. In this case, $D_s$ and $\partial_t$ do not generally commute. On the other hand, if we also adopt the notation $ds = |\partial_\theta c|d\theta$, we have a very nice form of ‘partial integration’:

$$\int_{S^1} \langle D_s h, k \rangle ds = -\int_{S^1} \langle h, D_s k \rangle ds.$$ 

We can now write down a whole family of reparametrization invariant metrics

$$G^n(h, k)_c = \int_{S^1} \sum_{i=0}^n A_i \langle D^i_s h, D^i_s k \rangle ds,$$

with dimensional constants $A_i > 0$. In general, these metrics give an entirely different notion of when a path of curves is horizontal to reparametrization orbits. Indeed, going back to example 3.8, we see that the higher derivative terms also contribute to the reparametrization momentum. We shall only be interested in one particular special case of this metric. We take $n = 1$ and let $A_1 \to \infty$, while keeping $A_0$ fixed. This is equivalent to discarding the $\langle h, k \rangle$ term and yields

$$G^{1,\infty}(h, k)_c = \int_{S^1} \langle D_s h, D_s k \rangle ds.$$ 

---

25 More carefully one should write $D_{s,c}$, as the derivative depends on the particular curve $c$. However, it does not usually occur that we differentiate with respect to different curves at the same time. The particular $c$ should be clear from the context.

26 In the language of functional analysis: The operator $D_s$ is anti-selfadjoint with respect to the $G^0$ metric.
The first thing to note about $G^{1,\infty}$ is that it is not a genuine metric on $\text{Emb}(S^1, \mathbb{R}^2)$. It vanishes whenever $h$ or $k$ is constant in $\theta$. However, such vector fields lie tangent to the orbits of rigid translations $[c]|_T = \{c(\cdot) + b \mid b \in \mathbb{R}^2\}$. Eventually, we will quotient out the translational action anyway. On the quotient $\text{Emb}(S^1, \mathbb{R}^2)/(\text{transl.})$ this will be a genuine metric. Therefore, this kind of degeneracy does not bother us. Another point to note is that this metric is not scale invariant. Whereas the $|\partial_\theta c|$ denominator of the $D_\nu$-operation ensures that $D_\nu h$ and $D_\nu k$ are scale invariant, the line element $ds$ still produces a factor of $\alpha$ if we rescale the curve by $\alpha$. This leads us to introduce the conformal factor $1/l(c)$ in front of the metric

$$G^{1,\infty}(h, k)|_c = \frac{1}{l(c)} \int_{S^1} \langle D_\nu h, D_\nu k \rangle ds,$$

ensuring scale invariance. In the literature, the scaling freedom is often dealt with by fixing the curve length to one. In chapter 2 we spent quite a lot of time to justify that this is actually the right thing to do. It does not suffice to have an action that is scale invariant. We also need to ensure that paths of curves with constant curve length are horizontal to the scaling orbits. Is this true for this metric? In analogy to eq. (11) in section 2.2 we find the following lemma.

**Lemma 3.10.** Let $t \mapsto c(t, \cdot)$ be a path of curves in $\text{Emb}(S^1, \mathbb{R}^2)$. The scaling momentum corresponding to $\lambda \in \mathbb{R} \simeq T_1 \mathbb{R}^+$ in the $G^{1,\infty}_{\text{scl}}$ metric is given by

$$G^{1,\infty}_{\text{scl}}(\partial_t c, \lambda c) = \lambda \cdot \partial_t \log l(c).$$

**Proof.** The Killing vector field on $\text{Emb}(S^1, \mathbb{R}^2)$ corresponding to $\lambda$ is given by $X^\lambda(c) = \lambda c$. Hence, the scaling momentum along the path $t \mapsto c(t, \cdot)$ is given by $G^{1,\infty}_{\text{scl}}(\partial_t c, \lambda c)$. We now calculate

$$\partial_t \log l(c) = \frac{1}{l(c)} \partial_t \int_{S^1} \langle \partial_\theta c, \partial_\theta \rangle^{1/2} d\theta$$

$$= \frac{1}{l(c)} \int_{S^1} \frac{1}{|\partial_\theta c|} \langle \partial_\theta c, \partial_\theta \rangle d\theta$$

$$= \frac{1}{l(c)} \int_{S^1} \frac{1}{|\partial_\theta c|^2} \langle \partial_\theta c, \partial_\theta c \rangle|\partial_\theta c| d\theta = G^{1,\infty}_{\text{scl}}(\partial_t c, c) \quad \Box$$

For paths horizontal to the action of $\mathbb{R}^+$ this scaling momentum has to be zero. This is the case if and only if $l(c)$ is constant along the path. We find that this metric has all the properties we desire a shape metric to have. It is discussed in much detail in [35], where it is also shown that this metric allows, unlike many other metrics, a very explicit treatment. Even in the case of closed curves one finds that the geometry of the pre-shape space is relatively simple, allowing geodesics to be calculated analytically.

### 3.4 Bending and Stretching coefficients

Let us now conclude our survey of metrics with a generalization of $G^{1,\infty}_{\text{scl}}$, proposed by Mio et al. in [26]. Let $v = D_\nu c = \frac{\partial c}{|\partial c|}$ be the unit tangent vector along $c$ and $n = (\frac{\nu}{\nu}, -1).v$ the normal. For $h \in T_c\text{Emb}(S^1, \mathbb{R}^2)$ we can decompose $D_\nu h$ pointwise as

$$D_\nu h = \langle v, D_\nu h \rangle v + \langle n, D_\nu h \rangle n.$$
Inserting this into the metric (18) yields

\[ G_{scl}^{1,\infty}(h, k)_c = \frac{1}{l(c)} \left( \int_{S^1} \langle v, D_s h \rangle \langle v, D_s k \rangle ds + \int_{S^1} \langle n, D_s h \rangle \langle n, D_s k \rangle ds \right). \]

The interpretation of the two terms is as follows. The \( v \) component describes deformation parallel to \( c \), i.e. stretching the curve like a rubber band. Conversely, the second term accounts for changes normal to the curve. This could be described as a bending of the curve, probably best compared with the deformation of a bicycle chain, where there is (ideally) no stretching. Having attributed these geometrical interpretations to both terms, Mio et al. suggested to introduce weight factors \( a, b > 0 \) and found the family of metrics

\[ G^{a,b}(h, k)_c = \frac{1}{l(c)} \left( a \int_{S^1} \langle v, D_s h \rangle \langle v, D_s k \rangle ds + b \int_{S^1} \langle n, D_s h \rangle \langle n, D_s k \rangle ds \right). \]

The conformal \( l(c)^{-1} \) factor is, actually, never found in their paper. They pass to curves of length 1 beforehand. In fact, it is worth checking that this is still compatible with the notion of horizontality to scaling if \( a \neq b \). Let us compute the scaling momentum:

\[
G^{a,b}(\partial_t c, c)_c = \frac{1}{l(c)} \left( \int_{S^1} a \cdot \langle v, D_s \partial_t c \rangle \langle v, D_s c \rangle + b \cdot \langle n, D_s \partial_t c \rangle \langle n, D_s c \rangle ds \right) = \frac{a}{l(c)} \int_{S^1} \langle D_s c, D_s \partial_t c \rangle ds = a \cdot G_{scl}^{1,\infty}(c, \partial_t c) \text{lem. 3.10} = a \cdot \partial_t \log l(c).
\]

Thus, the scaling momentum still vanishes for paths of constant curve length. Moreover, there is no bending contribution \( \sim b \) to the scaling momentum. At least from a superficial perspective, this makes sense, as bending does not change the length of a curve. The factors \( a, b \) can now be chosen at will. Depending on the application, one might wish to work with a metric which is more stiff (\( a/b \) small) or a metric that allows more stretching (\( a/b \) large). Apart from \( a = b = 1 \) which is the subject of Younes et al. in [35], there is also the particularly interesting case when \( a = 1/4 \) and \( b = 1 \). This choice may seem arbitrary now, but we will see that it allows us to establish an isometry to a far simpler space, for which the metric is just the constant \( L^2 \) metric. This was proposed by Srivastava et al. in [33] and will be the subject of the last chapter. Finally, the bending only choice \( a = 0 \) and \( b = 1 \), has been studied in [18]. Again, this allows some nice simplifications. Here, the basic idea is to fix the parametrization to arc length and represent a curve in \( \mathbb{R}^2 \simeq \mathbb{C} \) by the complex argument of \( D_s c \). This is known as the angle function representation. In this particular case the metric again looks like the \( L^2 \) metric, this time for one-dimensional functions. Though appealing by its simplicity, an application to shape clustering in [26] demonstrates its weaknesses compared to elastic metrics, i.e. metrics with \( a \neq 0 \). We will not include this special case in the further discussions. This concludes our short survey of metrics on spaces of curves and we now proceed to the explicit treatment of the above case where \( a = 1/4, b = 1 \).
4 Elastic Shape Analysis

Being finally at a stage where all the necessary theory is at our command, we can now start to look at a concrete example of a shape space for continuous curves. This chapter will be devoted to shape analysis using the square root velocity (SRV) representation developed by Srivastava et al. in [33]. The basic idea is to represent each shape, given as a curve \( c \in \text{Emb}(I, \mathbb{R}^2) \), using

\[
q(\theta) := \frac{\partial_{\theta} c}{\sqrt{|\partial_{\theta} c|}},
\]

the square root velocity function (SRVF) of \( c \). Then, we will equip the space of SRVFs with the constant \( L^2 \) metric, which makes the geometry particularly simple. As announced, this corresponds to a special case of elastic metric, as in section 3.4, with \( a = 1/4 \) and \( b = 1 \). The benefit of this representation is twofold. Firstly, we can write down geodesics of open curves quite explicitly. As the computation of geodesics in a general metric is only possible using advanced numerical methods, this constitutes a major simplification. This simplification is lost if we try to work with closed curves. It is true that the same representation may be used in this case. However, the additional constraint on the SRVF destroys the simplicity one has for open curves. This difficulty is addressed in [33]. As we have an interest to keep the treatment as explicit as possible, we will not deal with this case. The second benefit of the SRVF representation lies in the fact that it readily generalizes to curves in higher dimensions. The main objective of this chapter, however, will not be to discuss the benefits of this metric, but rather to take a close look at a theoretical difficulty that arises with the action of the infinite dimensional, non compact, diffeomorphism group.

In section 4.1 we will briefly discuss the set-up of the spaces we are working with, and the connection of the chosen metric to the general family of elastic metrics, introduced in 3.4. In a way, the SRV representation is nothing but a convenient set of coordinates to work with. Section 4.2 will show how the already familiar actions of scaling, rotation and reparametrization translate to these coordinates. In section 4.3 we will see that orbits of the diffeomorphism group are not closed, making it impossible to define the usual quotient metric we introduced in chapter 1. This will motivate us to define a wider equivalence relation on our pre-shape space, called Fréchet equivalence. Sections 4.4 and 4.5 will show that the ‘optimal reparametrization’ of a pre-shape may generally be found within this equivalence relation. Building up on this, section 4.6 shall be devoted to the question of how to define a metric on Fréchet equivalence classes. Although we will succeed with this, there are still many open questions. These questions, along with suggestions on how to resolve them, are outlined in the conclusion.

4.1 Preliminaries

We begin by looking at the exact spaces we wish to work on. We will restrict to open curves \( c : I = [0, 2\pi] \to \mathbb{R}^2 \). At first, it seems desirable to consider embeddings of \( I \). However, when we work with the square root velocity representation it is quite difficult to tell whether it actually represents an embedding or an immersion. It does not translate directly to \( q \) whether a curve is injective or not. Therefore, we might as well pass to immersions of \( I \), and allow for crossings etc. As we are not interested in translations, we can also directly quotient these out and deal with the space \( \text{Imm}(I, \mathbb{R}^2)/\text{transl} \). We then
have a bijective map

\[ \Psi : \text{Imm}(I, \mathbb{R}^2) / \text{transl} \rightarrow \{ q \in C^\infty(I, \mathbb{R}^2) \mid q(\theta) \neq 0 \}; \quad c \mapsto \frac{\partial_c}{\sqrt{|\partial_c|}} \]  \quad (20)

with inverse given by

\[ \Psi^{-1}(q)(\theta) = \int_0^\theta q(\tau)|q(\tau)|\,d\tau. \]  \quad (21)

Note that the function

\[ F : \mathbb{R}^2 \rightarrow \mathbb{R}^2; \quad x \mapsto \begin{cases} 0 & x = 0 \\ \frac{x}{|x|} & \text{otherwise} \end{cases} \]

is continuous. Thus, we can extend the SRV representation even further to the entire space \( C^\infty(I, \mathbb{R}^2) / \text{transl} \). We denote the obvious extension of \( \Psi \) by \( \bar{\Psi} \), which is still a bijection. This gives us the following hierarchy of spaces:

\[
\begin{array}{ccc}
\text{Emb}(I, \mathbb{R}^2) / \text{transl} & \subset & \text{Imm}(I, \mathbb{R}^2) / \text{transl} \\
\downarrow & & \downarrow \Psi \\
\{ q \in C^\infty \mid q(\theta) \neq 0 \} & \subset & C^\infty(I, \mathbb{R}^2)
\end{array}
\]

Using the continuity of \( F \), we will from now on implicitly set expressions like \( a/\sqrt{a} \) to zero, should \( a = 0 \). Let us proceed by equipping the space of SRVFs with the constant \( L^2 \) metric

\[ G(f, g)_q = \int_0^{2\pi} \langle f(\tau), g(\tau) \rangle_{\mathbb{R}^2} d\tau, \]

where \( f, g \in C^\infty(I, \mathbb{R}^2) \simeq T_qC^\infty(I, \mathbb{R}^2) \). So as not to confuse \( C^\infty(I, \mathbb{R}^2) \) with its different metrics, we shall denote \( V := (C^\infty(I, \mathbb{R}^2), G) \), the space of SRVFs with the \( L^2 \) metric, and correspondingly its open subset \( V^0 := \{ q \in C^\infty(I, \mathbb{R}^2) \mid q(\theta) \neq 0 \} \). Let us see what this metric looks like on our original spaces of curves. For this purpose, we need the differential of \( \Psi \).

**Lemma 4.1.** The differential \( d\Psi : T_c(\text{Imm}(I, \mathbb{R}^2) / \text{transl}) \rightarrow T_{\Psi(c)}V^0 \) is given by

\[ d\Psi(h) = \frac{\partial_c}{|\partial_c|^{1/2}} - \frac{1}{2}\frac{\langle \partial_c, \partial_c \rangle}{|\partial_c|^{3/2}} \partial_c \]

**Proof.** Let \( t \mapsto c(t, \cdot) \) be a path of curves such that \( h = \partial_t c(0, \cdot) \). We then find

\[ d\Psi(h) = \partial_t \Psi \circ c|_{t=0} = \partial_t \left( \frac{\partial_c}{\sqrt{|\partial_c|}} \right) = \frac{\partial_c}{|\partial_c|^{1/2}} - \frac{1}{2}\frac{\langle \partial_c, \partial_c \rangle}{|\partial_c|^{3/2}} \partial_c \]

and substituting \( \partial_c = \partial_{tac} = \partial_{ac} \) yields the desired result. \( \square \)

For \( h, k \in T_c(\text{Imm}(I, \mathbb{R}^2) / \text{transl}) \) we can now readily compute

\[ G(d\Psi(h), d\Psi(k)) = \int_0^{2\pi} \left( \frac{\langle \partial_c h, \partial_c k \rangle}{|\partial_c|} - \frac{3}{4}\frac{1}{|\partial_c|^3} \langle \partial_c, \partial_c \rangle \langle \partial_c, \partial_c \rangle \langle \partial_c, \partial_c \rangle \right) d\theta. \]
Introducing the same notation as in 3.4, this becomes
\[ = \int_{0}^{2\pi} \left( \langle D_s h, D_s k \rangle - \frac{3}{4} \langle v, D_s h \rangle \langle v, D_s k \rangle \right) ds. \]

Once again using the orthogonal decomposition of $h, k$ into normal and tangent components to $c$, we finally find
\[ = \int_{0}^{2\pi} \left( \langle n, D_s h \rangle \langle n, D_s k \rangle + \frac{1}{4} \langle v, D_s h \rangle \langle v, D_s k \rangle \right) ds = G^{1/4,1}(h, k)_c. \]

As promised, this is just a special case of the elastic metric with coefficients $a = 1/4, b = 1$. In fact, what we have established is that $\Psi$ and $\bar{\Psi}$ are isometric diffeomorphisms between
\[(\text{Imm}(I, \mathbb{R}^2) / \text{transl}, G^{1/4,1}) \simeq (V^0, G)\]
and
\[(C^\infty(I, \mathbb{R}^2) / \text{transl}, G^{1/4,1}) \simeq (V, G),\]
respectively. If we like, we can think of this as a convenient set of coordinates, just as we introduced Kendall’s coordinates in section 2.1.

Apart from the sets of functions we are working with, we should also take some care about the topology we endow them with. Essentially, there are two choices. $C^\infty(I, \mathbb{R}^2)$ as a Fréchet space comes with the strong topology of uniform convergence in all derivatives. Imm$(I, \mathbb{R}^2)$ and Emb$(I, \mathbb{R}^2)$ inherit this topology as open subsets. Most topological results found in the literature (e.g. [23, 6]) refer to this topology. As the Riemannian metric we have put on $C^\infty(I, \mathbb{R}^2)$ and its subspaces is only a weak Riemannian metric (c.f. section 1.7 and remark 3.1), the topology induced by the geodesic distance may differ significantly. However, our special choice of metric enables us to discuss this topology quite explicitly.

Working in the SRV representation, our metric takes the form of a constant $L^2$ metric. This means that our space is flat (at least before we consider scaling). The geodesic distance on $V$ and $V^0$ is thus given by
\[ d(q_1, q_2) = \|q_1 - q_2\|_{L^2}. \]

We know this topology quite well. In particular, we know that it is strictly weaker than the $C^\infty$ topology, $\tau_{L^2} \subsetneq \tau_{C^\infty}$. Uniform convergence of all derivatives implies $L^2$ convergence of SRV functions.\(^{27}\) The converse is generally false. In shape analysis, we are interested in employing the geodesic distance as a metric to compare shapes. Thus, if we use this metric in applications, it makes sense to work with the corresponding topology. We should be aware, however, that $V$ and $V^0$ are not complete in this topology. Naturally, we may ask what the completion with respect to the chosen metric looks like.

In the SRV representation this is straightforward to answer: $C^\infty(I, \mathbb{R}^2)$ lies dense in the space $L^2(I, \mathbb{R}^2)$ of all square-integrable functions endowed with the $L^2$ norm. Thus, the completion of $V$ is just $L^2(I, \mathbb{R}^2)$. How does this translate to our original space of curves? Suppose that $q \in L^2$ is a generic function. Then $q \cdot |q|$ will be integrable ($L^1$). Thus, the integral
\[ c(\theta) = \int_{0}^{\theta} q(\tau) |q(\tau)| d\tau \]

\(^{27}\)Taking one step at a time, $C^\infty$ convergence first implies uniform convergence of the SRVF, which in turn implies $L^2$ convergence.

39
exists a.e. Functions like $c$, which arise as the integral of some $L^1$ function, are known as absolutely continuous functions. The corresponding integrand (in our case $q(\tau)|q(\tau)|$) is a form of generalized derivative. Absolutely continuous functions are, in particular, continuous. We denote the space of all absolutely continuous functions $c : I \to \mathbb{R}^2$ as $C_{\text{abs}}$. As we have lost all information about the absolute placement of our curves, we are really just working with $C_{\text{abs}}/\text{transl}$. In a sense, we can think of this space as the ‘most natural’ pre-shape space, given the $G^{1/4,1}$ metric. It is the smallest complete space that contains $V$ and $V^0$. Once again, we summarize the set of spaces in a diagram.

$$\text{Emb}(I, \mathbb{R}^2)/\text{transl} \subset \text{Imm}(I, \mathbb{R}^2)/\text{transl} \subset C^\infty(I, \mathbb{R}^2)/\text{transl} \subset C_{\text{abs}}(I, \mathbb{R}^2)/\text{transl}$$

Completion in weak topology

$$\{q \in C^\infty | q(\theta) \neq 0\} \subset C^\infty(I, \mathbb{R}^2) \subset L^2(I, \mathbb{R}^2)$$

There are good reasons why one might wish to use the complete space instead of $C^\infty$ curves. Unfortunately, some results of the next sections a priori require more smoothness properties, which are not guaranteed for a generic $L^2$ SRVF. Therefore, we will continue to work with the ‘core space’ of smooth functions.

### 4.2 Shape preserving transformations on $V$ and $V^0$

Having closed the connection to $G^{1/4,1}$ and being aware of the two different topologies, let us now proceed with the discussion of shape preserving transformations. In section 3.2 we already suggested the $L^2$ metric and discarded it, as it was not invariant under reparametrization of curves. However, we have to keep in mind that the actions we are used to working with on $\text{Imm}(I, \mathbb{R}^2)$ may translate into different actions in our ‘new set of coordinates’ $V^0$. Indeed, the action of a diffeomorphism $\gamma \in \text{Diff}(I)$ now becomes

$$(\Psi(c), \gamma) \mapsto \Psi(c \circ \gamma) = \frac{\partial_b(c \circ \gamma)}{\sqrt{|\partial_b(c \circ \gamma)|}} = \sqrt{\gamma} \cdot \frac{c'(\gamma(\theta))}{\sqrt{|c'(\gamma(\theta))|}},$$

where $c'(y) := \partial_y c$ and $\gamma(t) = \partial_t \gamma$. In terms of $q \in V^0$ this translates to

$$(q, \gamma) \mapsto \sqrt{\gamma} \cdot (q \circ \gamma).$$

By a change of variables, this preserves the metric, in contrast to the earlier situation where we considered a constant $L^2$ metric. The other action that changes its appearance in the SRV representation is the scaling action. The square root in the denominator causes the change. We now have the scaling action

$$(\alpha, q) \mapsto \sqrt{\alpha} q,$$

for $\alpha \in \mathbb{R}^+$. We already discussed that we can remove scaling in $\text{Imm}(I, \mathbb{R}^2)$ by considering the orthogonal section of curves with length $l(c) = 1$, as the scaling momentum in $\text{Imm}(I, \mathbb{R}^2)$ with the $G^{1/4,1}$ metric was just $\frac{1}{4} \partial_t \log l(c)$. This translates very nicely to $V^0$. We find

$$l(c) = \int_0^{2\pi} |\partial_\theta c| d\theta = \int_0^{2\pi} |\Psi(c)|^2 d\theta = G(\Psi(c), \Psi(c)), \quad (22)$$
i.e. curves of length one have unit $L^2$ norm. All this, of course, is equally true for $C^\infty(I, \mathbb{R}^2)$ and $V$. We have therefore established correspondences

\[
\text{Imm}(I, \mathbb{R}^2)/\text{transl.scl} \cong \{ q \in V^0 \mid G(q, q) = 1 \} \;=: \; S(V^0)
\]

\[
C^\infty(I, \mathbb{R}^2)/\text{transl.scl} \cong \{ q \in V \mid G(q, q) = 1 \} \;=: \; S(V).
\]

The only action that has not changed is the rotational action of $SO(2)$, which is still given as $(O, q) \mapsto O.q$, by left matrix multiplication. One might be worried about the order in which the quotienting with respect to the different transformations is performed. However, as we have removed rigid translation from our representation entirely, we are only left with scaling, rotation and reparametrization. These remaining actions all commute. In analogy to chapter 2, we shall refer to $S(V)$ and $S(V^0)$ as pre-shape spaces. As they are both spheres, they feature a very nice geometry, e.g. geodesics between pre-shapes $q_1, q_2$ are given as great circles

\[
q(t) = \frac{1}{\sin(s)}(\sin(ts)q_1 + \sin((1-t)s)q_2),
\]

where $s = \arccos(G(q_1, q_2))$. Thus, the geodesic distance on $S(V)$ and $S(V^0)$ is given by

\[
d(q_1, q_2) = \arccos(G(q_1, q_2)).
\]

However, there is one point to note about great circles in $S(V^0)$. If $q_1, q_2 \in S(V^0)$ are SRVFs of immersions, i.e. $q_1(\theta) \neq 0$ and $q_2(\theta) \neq 0$ for all $\theta \in I$, then this does not automatically hold for all intermediate $q(t)$.

**Example 4.2.** Let $q_1, q_2 \in S(V^0)$ be SRVFs, such that there is a $\theta_0 \in I$ with $q_1(\theta_0) = -\alpha q_2(\theta_0)$ for some $\alpha > 0$. Then

\[
t \mapsto \sin(ts)q_1(\theta_0) + \sin((1-t)s)q_2(\theta_0)
\]

is a straight line in $\mathbb{R}^2$ between $q_1(\theta_0)$ and $q_2(\theta_0)$, and thus passes the origin at some intermediate $t \in (0, 1)$.

It is actually not difficult to see that paths like that in example 4.2 are the only situations in which great circles between $q_1, q_2 \in S(V^0)$ leave $S(V^0)$. We have found that $S(V^0)$ is \textit{geodesically incomplete}. The length minimizing great circles may leave $S(V^0)$, much like our very first example of incompleteness $\mathbb{R}^n \setminus \{0\}$ in example 1.3. Thus, if we are interested in a space that holds the entire geodesic between two pre-shapes $q_1$ and $q_2$, we have to work in the larger spaces $S(V)$ and $C^\infty(I, \mathbb{R}^2)/\text{transl.scl}$.

After computing geodesics in the pre-shape spaces $S(V^0)$ and $S(V)$, the next step is to take the quotient with respect to reparametrization and rotation. Similar to the discussion in chapter 2, rotation may be dealt with by performing an optimization step

\[
d([q_1]_{\text{rot}}, [q_2]_{\text{rot}}) = \inf_{O \in SO(2)} d(q_1, 0, q_2).
\]

Here, $[q_1]_{\text{rot}}$ denotes the orbit of the action under the rotation group $SO(2)$. Using an identification of $\mathbb{R}^2 \simeq \mathbb{C}$ one may use almost exactly the same formula as in section 2.5 to compute an optimal rotation, given two pre-shapes $q_1, q_2 \in S(V)$. In the end, we have to combine this rotational alignment with the optimization of the curve’s parametrization. Having said that, let us ignore rotation from now on and focus on the difficulties arising
with the action of $\text{Diff}(I)$. Following our usual procedure, we would like to use the following quotient metric

$$d_1([c_1]_{\text{Diff}}, [c_2]_{\text{Diff}}) = \inf_{\gamma \in \text{Diff}(I)} d(c_1, c_2 \circ \gamma),$$

for $c_1, c_2 \in C^\infty(I, \mathbb{R}^2)/\langle\text{transl, scl}\rangle$, where $d(c_1, c_2 \circ \gamma)$ denotes the geodesic distance in the pre-shape space $S(V)$. Naturally, two questions about this definition arise. Firstly, is (26) a well defined metric, i.e. does it satisfy the three axioms we require for a metric? Secondly, is this infimum attained? In other words, is there an optimal registration $\gamma \in \text{Diff}(I)$, such that $d(c_1, c_2 \circ \gamma) = d_1([c_1]_{\text{Diff}}, [c_2]_{\text{Diff}})$? Related to this, do we always find a generalized geodesic between $[c_1]_{\text{Diff}}$ and $[c_2]_{\text{Diff}}$, as in corollary 1.24? Of course, if the infimum were attained for any $[c_1]_{\text{Diff}}, [c_2]_{\text{Diff}}$, then (26) would certainly be well defined. All axioms would pass from the pre-shape metric to the quotient. Unfortunately, the answer is not quite as nice and we will indeed encounter some difficulties, addressed in the following sections.

### 4.3 Unclosed reparametrization orbits and Fréchet equivalence

Let us consider the question of whether (26) is a well defined metric on $S(V)/\text{Diff}(I)$ with the weak quotient topology. The answer is actually quite simple: It is not. The reason for this was already discovered in example 3.2. Reparametrization orbits in $C^\infty(I, \mathbb{R}^2)$ are not closed in the strong topology. We can use the same example to find a weakly converging sequence that leaves the orbit. For instance, the closure of some generic reparametrization orbit contains ‘reparametrizations’ that remain stationary on some interval. This implies that $S(V)/\text{Diff}(I)$ is not Hausdorff (in both the strong and weak topologies). Therefore, the quotient cannot be endowed with a compatible metric and the definition of $d_1$ in (26) fails. Let us make this failure more explicit, at the level of an example which closely resembles the above description, but explicitly uses the weak topology.

**Example 4.3.** Let $c$ be some pre-shape with SRVF $q \in S(V)$. Define $\gamma : I \to I$ by

$$\gamma(\theta) = \begin{cases} 
\frac{3}{2} \theta, & \theta \in \left[0, \frac{1}{3}\right] \\
\frac{1}{2}, & \theta \in \left[\frac{1}{3}, \frac{2}{3}\right] \\
\frac{3}{2} \theta - \frac{1}{2}, & \theta \in \left[\frac{2}{3}, 1\right]
\end{cases}$$

and let $\gamma_n = \frac{1}{n} \text{id} + \left(1 - \frac{1}{n}\right) \gamma$.\(^{28}\) Obviously, $c \circ \gamma$ is not in the reparametrization orbit of $c$, i.e. $[c]_{\text{Diff}} \neq [c \circ \gamma]_{\text{Diff}}$. Let us check that $d(c \circ \gamma_n, \gamma) \to 0$ as $n \to \infty$. We show that $G(\sqrt{\gamma_n} q \circ \gamma_n, \sqrt{\gamma} q \circ \gamma)_{L^2}$ converges to 1, which is equivalent to the distance converging to zero, as $d(c_1, c_2) = \arccos(G(q_1, q_2)_{L^2})$ is given as in eq. (25). To that account, note that $\gamma_n$ is differentiable a.e. and the derivative is bounded by $\dot{\gamma}_n = \frac{3}{2}$. By the Cauchy-Schwarz inequality for $\mathbb{R}^2$ we find

$$|\langle \sqrt{\gamma_n} q \circ \gamma_n, \sqrt{\gamma} q \circ \gamma \rangle| \leq \frac{3}{2} \|q \circ \gamma_n\|_2 \|q \circ \gamma\|_2 \leq \frac{3}{2} \|q\|_\infty.$$ 

Thus, the integrand is dominated by the integrable function $f \equiv \frac{3}{2} \|q\|_\infty$. We can therefore apply Lebesgue’s theorem of dominated convergence and interchange the limits

$$\lim_{n \to \infty} \int_I \langle \sqrt{\gamma_n} q \circ \gamma_n, \sqrt{\gamma} q \circ \gamma \rangle d\theta = \int_I \langle \sqrt{\gamma} q \circ \gamma, \sqrt{\gamma} q \circ \gamma \rangle d\theta = \|q\|_2^2 = 1.$$

\(^{28}\)The reader may notice that $\gamma_n$ are not $C^\infty$ and, therefore, can’t be in the orbit of $\text{Diff}(I)$. To make this more rigorous, but less explicit, one could use a monotonicity preserving mollification to smooth the $\gamma_n$. **42**
Remark 4.4. Note that the last interchange of limits requires pointwise convergence of the integrand. This is ensured if \( q \) is continuous. We will work with similar limit procedures later on. These all require some regularity of the \( q \)'s, which is one reason why we continue to work with \( C^\infty \) curves instead of absolutely continuous curves \( c \in C_{\text{abs}} \). We also note that \( \gamma_n \) converges pointwise a.e. in this example. Unfortunately, this is not guaranteed in a generic situation and will render exchanges of limits similar to this one more difficult, not to say impossible. Example 4.12 in section 4.4 will consider a situation where this does not work.

Having recognized the failure of the metric in eq. (26), what can we do to cure it? The previous example showed that this metric treats more curves as equivalent than are actually related by diffeomorphisms. Therefore, we need to widen our definition of when two pre-shapes \( c_1, c_2 \in C^\infty(I, \mathbb{R}^2)/(\text{trans}, \text{scl}) \) are equivalent. One way to do this would be to close up the orbits in the weak topology and define shapes to be equivalent, whenever they share the same closed orbit. By definition, this would make the metric definite. Unfortunately, we do not have any more explicit ways of describing this kind of equivalence. For instance, it is difficult to relate limit points of such orbits to other equivalent curves via a change of parameter. \( L^2 \) convergence is very little to conclude any such relation. Another approach to widen the equivalence relation has been taken by Younes et al. in [35]. They suggest to use the concept of Fréchet equivalence. There are various definitions of Fréchet equivalence in the literature (c.f. [21] p. 131, [11] p. 211). It is not obvious whether these concepts are all equivalent. However, we will not address this issue here, but simply take the following definition, which amounts to the same as in [11] and [35].

**Definition 4.5.** Let \( c_1, c_2 : I \to \mathbb{R}^2 \) be curves. We say that a function \( \gamma : I \to I \) is an admissible change of parameter for \( c_2 \), if the following holds

i) \( \gamma \) is non-decreasing, left continuous, with \( \gamma(0) = 0 \) and \( \gamma(1) \leq 1 \).

ii) Every discontinuity \([\gamma(\theta^-), \gamma(\theta^+)]\) is contained in an interval of constancy of \( c_2 \). Should \( \gamma(1) < 1 \), then \([\gamma(1), 1]\) must also be an interval of constancy for \( c_2 \).

Furthermore, we call \( c_1 \) and \( c_2 \) Fréchet equivalent, in symbols \( c_1 \sim_F c_2 \), if there is an admissible change of parameter \( \gamma \) for \( c_2 \), such that \( c_1 = c_2 \circ \gamma \).

Although this definition looks rather asymmetric, this is a well defined equivalence relation. Symmetry can be established by making use of the pseudo inverse

\[
\gamma^-(y) := \sup\{x \in I \mid \gamma(x) \leq y\};
\]

(27)
defined for the admissible change of parameter \( \gamma \). Note that condition ii) is vital for this to work. Otherwise, we would not have a well defined equivalence relation.

**Example 4.6.** Let \( c_1 \) be a straight line in the plane and let \( c_2 \) be the same line, continuously concatenated with a loop in its middle (c.f. figure 9). As suggested by the figure, we can relate \( c_1 \) and \( c_2 \) by a discontinuous non-decreasing function \( \gamma \) in such a way that \( c_1 = c_2 \circ \gamma \). The discontinuity of \( \gamma \) cuts out the entire circle. However, \( c_1 \) and \( c_2 \) are not Fréchet equivalent, as the requirement ii) in the definition is not met. \( c_2 \) is not constant along the interval \([\gamma(\frac{1}{2}^-), \gamma(\frac{1}{2}^+)]\). Furthermore, applying \( \gamma^- \) to \( c_1 \) would not work, i.e. \( c_1 \circ \gamma^- \neq c_2 \).
Before we proceed to discuss the quotient $C^\infty(I, \mathbb{R}^2)/(\text{transl}, \text{scl}, \sim_F)$, let us collect a few properties of Fréchet equivalent curves. The first thing to note is that this equivalence relation is not induced by a group action, in contrast to all other equivalence relations considered so far. We could try to define an action $(\gamma, c) \mapsto c \circ \gamma$ using admissible changes of parameter. However, the set of admissible $\gamma$’s is different for each curve. We cannot apply an arbitrary non-decreasing function to a curve without, generally, cutting out pieces of it. This is, of course, a drawback, as most of our theory in chapter 1 was tailored to the specific situation of an isometric group action. Let us, nevertheless, carry on and deal with the difficulties as they occur. Consider the following lemma, an encouragement that Fréchet equivalence is a ‘good extension’ of reparametrization orbits.

Lemma 4.7. Let $c_1, c_2$ be Fréchet equivalent curves. Then $c_1$ and $c_2$ have the same image in $\mathbb{R}^2$.

Proof. Let $x \in \text{Im}(c_1)$, i.e. $x = c_1(s)$ for some $s \in I$. As $c_1$ and $c_2$ are Fréchet equivalent, there is an admissible change of parameter $\gamma$, such that $c_2(\gamma(s)) = c_1(s) = x$. Therefore, $x \in \text{Im}(c_2)$. 

The previous lemma holds for any two Fréchet equivalent curves. As we are currently working with $C^\infty(I, \mathbb{R}^2)$ curves, our equivalence class is slightly smaller. For instance, there may be non smooth admissible changes of parameter like a piecewise linear function. These would still create a Fréchet equivalent curve, however, there may be some places where the resulting curves are no longer differentiable. This will be important when we deal with the question of minimizing distance between Fréchet equivalence classes. We will see that the ‘optimal registration’ may indeed leave the class of $C^\infty$ curves. This suggests to allow for weaker conditions, e.g. the absolutely continuous curves introduced in section 4.1. However, we will soon encounter difficulties that will force us into working with curves that have at least a continuous SRVF. This is not guaranteed for absolutely continuous curves.

Having defined Fréchet equivalence directly on parametrized curves, we still need to establish how Fréchet equivalence translates to the SRV representation. Non-decreasing functions $\gamma$, like the admissible changes of parameter in Fréchet equivalence, are differentiable almost everywhere. Let us denote this derivative by $\dot{\gamma}$ and set it to zero where it does not exist. With this definition, the ‘action’ $(\gamma, c) \mapsto c \circ \gamma$ still translates to $(\gamma, q) \mapsto \sqrt{\gamma}q \circ \gamma$. Condition ii) in definition 4.5 corresponds to $q \equiv 0$ on the interval $[\gamma(t^-), \gamma(t^+)]$ whenever there is a discontinuity $t$. We therefore find the equivalence class.
of \( q \in S(V) \) to be

\[
[q]_F = \{ \sqrt{\gamma} q \circ \gamma \mid \gamma \text{ non-decreasing, s.t. } q \equiv 0 \text{ on every interval } [\gamma(t^-), \gamma(t^+)] \}\.
\]

The reader may have noted that we included the peculiar demand of left-continuity in the conditions for an admissible change of parameter. A non-decreasing function at some point \( t \in I \) is either left continuous or right continuous. Technically, we do not need either of these conditions for Fréchet equivalence to work. This condition has been introduced for convenience, to relate Fréchet equivalence more easily to the existence of optimal matchings between shapes. This will be addressed in the following section.

### 4.4 Existence of Optimal Matchings

Having seen that \( d_1([c_1], [c_2]) = \inf_{\gamma \in \text{Diff}(I)} d(c_1, c_2 \circ \gamma) \) is a priori not a good metric (but a pseudo metric) for orbits of the diffeomorphism group, let us now turn to the second question raised at the end of section 4.2. Is the infimum in this definition attained? In other words, is there an optimal reparametrization of the curve \( c_2 \) which realizes this minimal ‘distance’ between reparametrization orbits? Again, we already know the answer. In example 4.3 we have shown that the infimum is not attained, by pointing out the degeneracy of this pseudo metric. However, we can generally prove that an optimal matching exists in the wider class of Fréchet equivalent functions. The general problem of finding optimal diffeomorphisms for a certain class of matching problems has been addressed in [34] by Trouvé and Younes. We will heavily draw on their results here and apply them to our specific shape matching problem. To that account, let us introduce the following set

\[
D^* := \{ \varphi : I \rightarrow I \mid \varphi \text{ non-decreasing, left-continuous, } \varphi(0) = 0, \varphi(1) \leq 1 \}.
\]

There is a well known connection between functions \( \varphi \in D^* \) and the space of probability measures on the interval \( I \). Given some probability measure \( \mu \) on the Borel \( \sigma \)-algebra \( B([0,1]) \), we may consider its cumulative distribution function (cdf) \( \varphi(s) := \mu([0,s)) \). This is exactly a left-continuous, non-decreasing function with \( \varphi(0) = 0 \) and \( \varphi(1) \leq 1 \).

In fact, there is a one-to-one correspondence between such measures and non-decreasing, left continuous functions. This can be found in any book on measure theory (e.g. Bauer [3]). The space \( D^* \) carries the topology of pointwise convergence, also referred to as the weak* topology (c.f. [34]). This corresponds to the concept of convergence in distribution, familiar from elementary probability theory. The reason for introducing the set \( D^* \) is an important property of this topology. We have the following lemma.

**Lemma 4.8** (Parthasarathy,[30] theorem 6.4.). \( D^* \) is compact in the weak* topology.

**Remark** 4.9. Note that we are making the identification of \( D^* \) with probability measures only to prove this topological result. We will employ this connection once more in the proof of 4.17. However, a great difference between these measures and the set \( D^* \) is our definition of derivative. For \( \varphi \in D^* \) we define the derivative \( \dot{\varphi} \) as usual, where it exists, and set it to zero otherwise (at points of discontinuity). This agrees with our convention for parameter changes in Fréchet equivalence. Generally, a probability measure \( \mu_\varphi \) that corresponds to \( \varphi \in D^* \) has no derivative. It may be decomposed in a part that is absolutely continuous with respect to the Lebesgue measure \( dx \) and a singular part \( d\nu \)

\[
d\mu_\varphi = f \cdot dx + d\nu.
\]
The function $f$ is defined up to a set of Lebesgue measure zero and agrees with $\dot{\varphi}$ almost everywhere.\footnote{Usually, $f$ is referred to as the Radon-Nikodym derivative in measure theory. It formalizes the concept of a probability density.} When we write $\dot{\varphi}$ we are not working with any kind of $\delta$-functions or distributional derivatives.

Having introduced the set $D^*$ and its most important property, let us now state the central theorem of this section. Its proof will form the rest of this section and will be the most technical part of this dissertation. In a first reading the reader may want to skip the proof and continue with section 4.5 for how theorem 4.10 may be interpreted for the shape matching problem.

**Theorem 4.10.** Let $c_1, c_2 \in C^\infty(I, \mathbb{R}^2)$ be smooth curves of unit length with SRV representations $q_1, q_2 \in S(V)$ and define $E_{q_1, q_2} : D^* \to \mathbb{R}$ as

$$E_{q_1, q_2}[\gamma] := \langle q_1, \sqrt{\gamma} q_2 \circ \gamma \rangle_{L^2}. $$

Then $E$ attains its supremum. Furthermore, we have

$$E_{q_1, q_2}[\gamma] = E_{q_2, q_1}[\gamma^{-}]$$

for all $\gamma \in D^*$, with the involution $\gamma^{-}$ defined as in eq. (27).

**Remark 4.11.** Note that maximizing $E_{q_1, q_2}$ is equivalent to minimizing $\arccos(E_{q_1, q_2})$, being the geodesic distance in $S(V)$.

**Example 4.12.** The obvious way to prove the above theorem would be to show that $E_{q_1, q_2}$ is continuous and exploit the compactness of $D^*$. However, this is not true! Consider the following counterexample. Define $\gamma_n(t) := t - \frac{1}{2\pi n} \sin(2\pi nt)$ and take $q_1(t) = q_2(t) = (1, 0)^T$ to be the SRVF of a straight line. $E_{q_1, q_2}$ takes now the form

$$E_{q_1, q_2}[\gamma_n] = \int_0^1 \sqrt{1 - \cos(2\pi nt)} \, dt. $$

This is a complete elliptic integral and can be performed analytically, resulting in $E_{q_1, q_2}[\gamma_n] = \frac{2\sqrt{2}}{\pi} < 1$ for all $n$. On the other hand, we have $\lim_{n \to \infty} \gamma_n(t) = t$ for all $t \in [0, 1]$, i.e. $\gamma_n \to \text{Id}$ in $D^*$. Thus we find

$$\lim_{n \to \infty} E_{q_1, q_2}[\gamma_n] = \frac{2\sqrt{2}}{\pi} < 1 = E_{q_1, q_2}[\text{Id}] = 1. $$

This proves that $E_{q_1, q_2}$ is generally not continuous.

All hope is not lost, though. There is a slightly weaker notion of continuity called semi-continuity. This replaces the usual condition $\lim_{x \to x_0} f(x) = f(x_0)$ that a function $f$ is continuous, by the weaker condition

$$\limsup_{x \to x_0} f(x) \leq f(x_0)$$

for upper semi-continuity and

$$\liminf_{x \to x_0} f(x) \geq f(x_0)$$

for lower semi-continuity. If a function $f$ is both, lower and upper semi-continuous, then it is continuous. However, we will not dwell on too many details here. Please refer to some general introductions to topology, e.g. Bourbaki [5] p. 360. The key property of semi-continuity we need is the following.

\footnote{29}
Lemma 4.13. Let $X$ be a compact topological space and $f : X \to \mathbb{R}$ be upper semi-continuous. Then $f$ attains a global maximum. Conversely, if $g : X \to \mathbb{R}$ is lower semi-continuous, then $g$ attains its global minimum.

Having an application of this lemma in mind, upper semi-continuity is what we need to establish for $E_{q_1, q_2}$. This is not a straightforward proof to carry out. Fortunately, Trouvé and Younes [34] serve us a fitting result on a silver plate. It goes as follows.

Proposition 4.14 ([34], Proposition 5.1.). Let $f : [0, 1]^2 \to \mathbb{R}_{\geq 0}$ be continuous and non-negative. Define $U_f : D^* \to \mathbb{R}$ by

$$U_f[\varphi] = \int_0^1 \sqrt{\varphi} f(x, \varphi(x))dx.$$ 

Then $U_f$ is upper semi-continuous.

But this is almost exactly the situation we are dealing with! Setting $f(x, y) = \langle q_1(x), q_2(y) \rangle_{\mathbb{R}^2}$ does not quite work, as this may be negative, depending on $q_1$ and $q_2$. We could, however, apply proposition 4.14 to $\tilde{f}(x, y) = \max(\langle q_1(x), q_2(y) \rangle_{\mathbb{R}^2}, 0)$. If $q_1, q_2$ are continuous, so will $\tilde{f}$, and we are almost there. All we need is the following lemma.

Lemma 4.15. Let $U_f : D^* \to \mathbb{R}$ be given by $U_f[\varphi] = \int_0^1 \sqrt{\varphi} f(x, \varphi(x))dx$ for some measurable $f : [0, 1]^2 \to \mathbb{R}$. Then maximizing $U_f$ is equivalent to maximizing $U_{f^+}[\varphi] := \int_0^1 \sqrt{\varphi} \max(f(x, \varphi(x)), 0)dx$.

Proof. Consider the mapping $x \mapsto f(x, \varphi(x))$. Whenever this becomes negative, we may replace $\varphi$ by a constant piece (and possibly introduce a discontinuity into $\varphi$). This will make $\tilde{\varphi} = 0$, and we might as well use $\max(0, f)$ instead. \hfill $\square$

We can, finally, complete the proof of 4.10. According to lemma 4.15, maximizing $E_{q_1, q_2}$ is equivalent to maximizing

$$\tilde{E}_{q_1, q_2}[\varphi] := \int_0^1 \sqrt{\varphi} \max(\langle q_1(t), q_2(\varphi(t)) \rangle_{\mathbb{R}^2}, 0)dt.$$ 

If $q_1, q_2$ are continuous, we may apply proposition 4.14. Thus, $\tilde{E}_{q_1, q_2}$ is upper semi-continuous. Employing lemma 4.13 and 4.8, we get the existence of $\gamma^* \in D^*$, such that

$$\max_{\gamma \in D^*} E_{q_1, q_2}[\gamma] = \max_{\gamma \in D^*} \tilde{E}_{q_1, q_2}[\gamma] = \tilde{E}_{q_1, q_2}[\gamma^*].$$

Finally, we may have to alter $\tilde{\gamma}^*$ as in the proof of 4.15, which yields $\gamma^*$, such that

$$E_{q_1, q_2}[\gamma^*] = \max_{\gamma \in D^*} E_{q_1, q_2}[\gamma].$$

This proves the first part of theorem 4.10. The second part is a direct application of another result by Trouvé and Younes. In the same situation as proposition 4.14, only allowing for negative $f$ as well, we have:

Proposition 4.16 ([34], Proposition 5.7.). Let $U_f$ be defined as above and define $\tilde{f}(x, y) = f(y, x)$. Then

$$U_f[\varphi] = U_f[\varphi^+]$$

for all $\varphi \in D^*$.

This completes the proof of theorem 4.10.
4.5 Interpretation of Singular Matchings

Having found a positive answer to the matching problem within the set $D^*$, how can we interpret this? Generally, the optimal $\gamma$ may not be an admissible change of parameter for $q_2$. Trouvé and Younes present criteria, when the optimal matching is actually found within the set of homeomorphisms of the interval. Being useful for applications presented in their paper, these criteria are, however, too strong to fit into our shape matching problems. Their idea is to demand that the function $f$ must not vanish on horizontal and vertical segments in some neighbourhood of the main diagonal in $[0,1]^2$. For most $q$’s however, this neighbourhood will be too large to demand max($\langle q_1, q_2 \rangle R_2, 0$) to be non-vanishing. This is supported by numerical evidence. In the appendix we describe how the optimization over $\gamma \in D^*$ can be carried out. This algorithm was applied to various shapes. An example is given in figure 10. Here figure 10(a) and 10(b) feature the shapes the algorithm was applied to. We show the optimal $\gamma$ in 10(c), which is clearly discontinuous.

![Figure 10](image)

Figure 10: Shape matching where the optimal $\gamma \in D^*$ is not continuous.

Similar results are found for other shapes. The phenomenon that $\gamma$ is not an admissible reparametrization (in the sense of definition 4.5) seems to be quite generic and does not only happen for pathological cases (c.f. Figure 15 in the appendix). The obvious problem with this is that $c \circ \gamma$ produces a discontinuous curve, if $\gamma$ is not admissible for $c$. What does this mean for the corresponding SRVF $\sqrt{\gamma}q \circ \gamma$? It seems clear that it may be discontinuous even if $c$ was chosen to be $C^\infty$. What space does $\sqrt{\gamma}q \circ \gamma$ lie in then? In section 4.1 we established that every $\tilde{q} \in L^2(I, R^2)$ can be integrated to an absolutely continuous curve (c.f. eq. (21))

$$\Psi^{-1}(\tilde{q}) = \int_0^\theta \tilde{q}(\tau) |\tilde{q}(\tau)| d\tau.$$  

This seems to suggest that $\sqrt{\gamma}q \circ \gamma$ cannot lie in $L^2(I, R^2)$, as the reparametrization of the original curve $c \circ \gamma$ is clearly discontinuous. Let us prove a seemingly contradicting result:

**Lemma 4.17.** Let $q_1, q_2 \in V$ be SRVFs with $q_1 = \sqrt{\gamma}q_2 \circ \gamma$ for some $\gamma \in D^*$. Then $\|q_1\|_{L^2} \leq \|q_2\|_{L^2}$. Furthermore, we have $q_1 \sim_F q_2$ if and only if $\|q_2\|_{L^2} = \|q_1\|_{L^2}$.

**Proof of lemma 4.17.** This is an application of a generalized change of variable rule, valid for any $\gamma \in D^*$. If $\mu$ is the measure on $\mathcal{B}([0,1])$ corresponding to $\gamma \in D^*$, it may be uniquely decomposed such that $\mu = \gamma dx + \nu_\gamma$. Here, $dx$ is the Lebesgue measure and $\nu_\gamma$ denotes the singular part of $\mu$, capturing the discontinuities of $\gamma$. For any $\mu$-integrable $g$
we then have (c.f. lemma 5.11. in [34])
\[
\int_0^1 g \circ \gamma \circ (\nu) d\nu = \int_0^1 g(u)\gamma(u) du + \int \chi_{[0,t]}(u) g(u) d\nu(u),
\tag{30}
\]
where \(\chi_{[0,t]}(u)\) is the usual characteristic function of \([0, t]\). Setting \(g = |q_2 \circ \gamma|^2\), we have \(g \circ \gamma^{-1}(v) = |q_2(v)|^2\), apart from the places \(t = \gamma^{-1}(v)\) where \(\gamma(t)\) is discontinuous. This allows the rewriting of \(\|q_2\|_{L^2}\) as
\[
\int_0^1 |q_2(v)|^2 d\nu
= \int_0^1 |q_2(\gamma(\nu^{-1}(v)))|^2 d\nu + \sum_{t \in J} \left( \int_{\gamma(t^-)}^{\gamma(t^+)} |q_2(v)|^2 d\nu - |q_2(\gamma(t^-))|^2 (\gamma(t^+) - \gamma(t^-)) \right).
\]
Here, \(J\) is the discontinuity set of \(\gamma\) and the additional terms amend the errors made by replacing \(q(v) \leftrightarrow q(\gamma(\nu^{-1}(v)))\). Indeed, every discontinuity \(t \in J\) leads to an interval of constancy \(\gamma \circ \gamma^{-1}(v) = \gamma(t^-), v \in [\gamma(t^-), \gamma(t^+)]\). Therefore, we subtract the contribution \(|q_2(\gamma(t^-))|^2 (\gamma(t^+) - \gamma(t^-))\) and add the missing part \(\int_{\gamma(t^-)}^{\gamma(t^+)} |q_2(v)|^2 d\nu\). Introducing the singular measure \(\nu\) and removing the contribution of a possible discontinuity\(^{30}\) \(t = 1\), we get
\[
\|q_2\|_{L^2} = \int_0^1 |q_2(\gamma(\nu^{-1}(v)))|^2 d\nu - \int |q_2(\gamma(u))|^2 d\nu(u) + \sum_{t \in J} \int_{\gamma(t^-)}^{\gamma(t^+)} |q_2(v)|^2 d\nu\nu(u)
\geq \int_0^{\gamma(1)} |q_2(\gamma(\nu^{-1}(v)))|^2 d\nu - \int \chi_{[0,1]}(u) |q_2(\gamma(u))|^2 d\nu(u) + \sum_{t \in J \setminus \{1\}} \int_{\gamma(t^-)}^{\gamma(t^+)} |q_2(v)|^2 d\nu.
\]
We may now apply eq. (30) to the first term in the last row, and conclude
\[
\|q_2\|_{L^2} \geq \int_0^{\gamma(1)} |q_2 \circ \gamma(\nu)|^2 d\nu + \sum_{t \in J \setminus \{1\}} \int_{\gamma(t^-)}^{\gamma(t^+)} |q_2(v)|^2 d\nu \geq \|q_1\|_{L^2}^2.
\]
Equality holds if \(|q_2(v)|^2 = 0\) a.e. on all sets \([\gamma(t^-), \gamma(t^+)\], \(t \in J\). As we are working with continuous \(q\)'s, this is exactly the condition an admissible change of parameter needs to satisfy (translated to the SRV representation). Thus, equality holds if and only if \(q_1 \sim_F q_2\).

\(30\)To be more accurate one should say: ‘remove the contribution of \([\gamma(1), 1]\)’, as \(t = 1\) is not a discontinuity in our convention. In abuse of notation we will write \(\gamma(1^+) = 1\) and include \(t = 1\) in \(J\) should \(\gamma(1) < 1\).

Remark 4.18. Recall that \(\|g\|_{L^2}^2\) for an SRVF is just the length of the curve that it represents (c.f. eq. (22) in section 4.2). Thus, translating this lemma to our original curves, it says that \(\gamma\) is an admissible change of parameter if and only if it preserves the length of the curve. Furthermore, the fact that the norm is not increasing ensures that expressions such as \(\arccos(⟨q_1, \sqrt{2}q_2 \circ \gamma⟩_{L^2})\) are still well defined, by virtue of the \(L^2\) Cauchy-Schwarz inequality.

Turning back to non-admissible changes of parameter, how can it both be true that \(\sqrt{2}q \circ \gamma \in L^2\) and that \(c \circ \gamma\) is discontinuous? This is best expressed in terms of
commutativity of maps. Using the familiar map $\Psi(c)(\theta) = \frac{\partial c}{\sqrt{\gamma q}}$, taking a curve $c$ to its SRVF, we have the following commuting diagram:

$$
\begin{array}{c}
C_{\text{abs}}(I, \mathbb{R}^2) / \text{transl} \xrightarrow{\Psi} \gamma \in D^* \rightarrow \mathcal{F} / \text{transl} \\
\downarrow \Psi \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \q
between one point and an entire segment in the limit case. Correspondingly, one might call this behaviour ‘overstretching’. Gaps in the target shape \( q_2 \) are segments that are matched to a single point on the initial shape \( q_1 \). This may seem asymmetric, but it is not. Theorem 4.10 asserts that \( E_{q_1,q_2}[\gamma] = E_{q_2,q_1}[\gamma^-] \), using the pseudo inverse of \( \gamma \). We can further resolve this apparent asymmetry by considering the following lemma.

**Lemma 4.19.** Let \( \gamma \in D^* \). Then there exist continuous \( \gamma_1, \gamma_2 \in D^* \), such that \( \gamma = \gamma_2 \circ \gamma_1^- \).

Heuristically, we take \( \gamma \) and insert additional intervals at points of discontinuity to bridge the gaps. At the same time we start with \( \tilde{\gamma}_1 = \text{Id} \) and insert intervals of constancy at the same locations where we have bypassed a gap. This is illustrated in figure 12. However, we have to deal with the possibility of countably many jumps, which makes the actual proof a bit cumbersome.

![Figure 12: Illustration of the proof of lemma 4.19.](image)

**Proof.** Let \( J = \{ t \mid \gamma(t^+) - \gamma(t^-) > 0 \} \) be the set of discontinuities.\(^{31} \) For \( t \in [0, 1] \) let us define

\[
A_t := \{ \hat{t} \in J \mid \hat{t} \leq t \} \text{ jumps that happen before } t
\]

\[
B_t := \{ \hat{t} \in J \mid \hat{t} > t \} \text{ jumps that happen after } t
\]

\[
L_t := \sum_{t \in A_t} l_t \text{ the accumulated jump height}
\]

\[
a_t := \text{sup} \, A_t \text{ the location of the previous jump}
\]

\[
b_t := \text{inf} \, B_t \text{ the location of the following jump}
\]

\[
l_t := \gamma(a_t^-) - \gamma(a_t^-) \text{, the height of the previous jump.}
\]

using the convention \( \text{sup} \, \emptyset = 0 \) and \( \text{inf} \, \emptyset = 1 \). As \( \gamma(0) = 0 \) and \( \gamma(1) \leq 1 \), we have the upper bound \( L_t \leq 1 \) for all \( t \). We can now define

\[
\tilde{\gamma}_2(t) := \begin{cases} 
\gamma(a_t) + t - (a_t + L_t) & , t \in [a_t + L_t, a_t + L_t + l_t] \\
\gamma(t - (L_t + l_t)) & , t \in (a_t + L_t + l_t, b_t + L_t + l_t]
\end{cases}
\]

\[
\tilde{\gamma}_1(t) := \begin{cases} 
a_t & , t \in [a_t + L_t, a_t + L_t + l_t] \\
t - (L_t + l_t) & , t \in (a_t + L_t + l_t, b_t + L_t + l_t]
\end{cases}
\]

\(^{31}\)Again, one should include \( t = 1 \) in the case \( \gamma(1) < 1 \).
By definition, these functions are continuous. As the functions have been stretched by an overall length of $L_1$, we need to perform a rescaling. This leads to $\gamma_2(t):=\tilde{\gamma}_2(L_1t)$ and $\gamma_1(t):=\tilde{\gamma}_1(L_1t)$. $\gamma_1$ now has discontinuities at exactly the same places as $\gamma$. Everywhere else, its slope is given by $L_1^{-1}$, balancing the rescaling performed on $\tilde{\gamma}_2$. We finally find that $\gamma = \gamma_2 \circ \gamma_1$.

This lemma further emphasizes that a $\gamma \in D^*$ should be considered a matching prescription, rather than a reparametrization. Equivalently, we might decompose $\gamma = \gamma_2 \circ \gamma_1$ and reparametrize the shapes as $\sqrt{\gamma_1}q_1 \circ \gamma_1$ and $\sqrt{\gamma_2}q_2 \circ \gamma_2$ with the continuous $\gamma_1$ and $\gamma_2$. By the symmetry property in theorem 4.10, these admissible reparametrizations do have the same geodesic distance as $q_1$ and $\sqrt{\gamma}q_2 \circ \gamma$. This means that we may actually find an optimal matching constellation within the Fréchet equivalence classes of $q_1$ and $q_2$. After the discussion in section 4.3, this is what we were already prepared to admit. Furthermore, if we perform the optimization over both equivalence classes, we do not need to include any discontinuous changes of parameter (not even the admissible ones). Any discontinuity that arises on one side may be moved to the other side by introducing an additional interval of constancy. Continuing the example of figures 10 and 11 we show in figure 13 how the discontinuous $\gamma$ is decomposed (13(b) and 13(c)) and how a geodesic between the optimally aligned shapes looks like (13(d)). This makes clear that there are no discontinuities at all. Further examples are found in appendix A.3. There is still one drawback, however. The optimal matching may generally be non-smooth on a set of measure zero. Thus, the optimal reparametrization may leave the space of $C^\infty$ curves, i.e. lose differentiability properties. This may be attributed to the incompleteness of $S(V)$ in the weak topology. Again, this would be a good reason to work with the completion of $C^\infty$, i.e. with absolutely continuous curves. However, it is not clear if the existence proof can be extended to such a broad class of functions. So far, we do require more continuity properties than a generic $L^2$ SRVF might have.

![Graphs showing decomposition and geodesic](image-url)

Figure 13: a) - c) decomposition according to lemma 4.19; d) geodesic connecting the corresponding shapes of figure 10 with optimal reparametrization.

**Remark 4.20.** Lemma 4.19 seems to suggest that discontinuous changes of parameter are redundant at all. In particular, one might think that admissible changes of parameter are
not necessary and one might work with continuous $\gamma$’s from the start on. However, even though these discontinuous (but admissible) changes of parameter are not required to find an optimal matching between Fréchet equivalence classes, they are of vital importance to make Fréchet equivalence a well defined concept. Only defining equivalence with continuous $\gamma$’s does not work, as this does not constitute a symmetric equivalence relation.

### 4.6 A quotient metric on Fréchet equivalence classes

Let us now turn back to our original task, defining a ‘decent’ metric between shapes. We have already identified $C^\infty(I, \mathbb{R}^2)/(\text{transl,scl}) \simeq S(V)$ as a suitable pre-shape space and would now like to pass to a shape space, by quotienting out Fréchet equivalence.\(^{32}\) Previously, we investigated the Hausdorff property of these quotients (in their quotient topology) to enquire about the general possibility of defining a metric. Let us skip this step now, and directly define a metric. Afterwards we can investigate how this relates to the different topologies. The key result will be the following theorem. As we have emphasized the equivalence of $C^\infty(I, \mathbb{R}^2)/(\text{transl,scl})$ with the space of normed SRVFs, we will continue to work in the SRV representation only.

**Theorem 4.21.** Let $Q := S(V)/\sim_F$ be the space of normed SRVFs modulo Fréchet equivalence. We define $d_Q : Q \times Q \to \mathbb{R}_{\geq 0}$ as

$$d_Q([q_1], [q_2]) := \inf_{\tilde{q}_1 \in [q_1], \tilde{q}_2 \in [q_2]} d(\tilde{q}_1, \tilde{q}_2),$$

where $d(\tilde{q}_1, \tilde{q}_2) = \arccos(\langle \tilde{q}_1, \tilde{q}_2 \rangle_{L^2})$ is the geodesic distance on $S(V)$. Then $d_Q$ is a well defined metric.

**Proof.** To prove this we will need to rewrite this two-sided infimum definition as a one-sided infimum over $\gamma \in D^*$

$$d_Q([q_1], [q_2]) := \inf_{\tilde{q}_1 \in [q_1], \tilde{q}_2 \in [q_2]} d(\tilde{q}_1, \tilde{q}_2) = \inf_{\gamma \in D^*} d(q_1, \sqrt{\gamma} q_2 \circ \gamma). \quad (31)$$

This was already established in the previous section. Indeed, with lemma 4.19 we can decompose every $\gamma \in D^*$ as $\gamma = \gamma_2 \circ \gamma_1^\circ$, such that $\gamma_1, \gamma_2$ are continuous. Employing the symmetry property of theorem 4.10 we get $d(q_1, \sqrt{\gamma} q_2 \circ \gamma) = d(\sqrt{\gamma_1} q_1 \circ \gamma_1, \sqrt{\gamma_2} q_2 \circ \gamma_2)$. This proves eq. (31) and will help us to establish definiteness and the triangle inequality.

**Symmetry:** as the geodesic distance $d$ is a well defined metric on $S(V)$, the symmetry of $d_Q$ directly descends from $d$ and the two-sided infimum definition.

**Triangle inequality:** let $[q_1], [q_2], [q_3] \in Q$ and $\epsilon > 0$. Chose $\tilde{q}_1 \in [q_1]$ and $\tilde{q}_3 \in [q_3]$ such that $d(\tilde{q}_1, \tilde{q}_3) \leq d_Q([q_1], [q_3]) + \epsilon / 2$ and find $\gamma \in D^*$ such that $d(\tilde{q}_3, \sqrt{\gamma} q_2 \circ \gamma) \leq d_Q([q_3], [q_2]) + \epsilon / 2$. Furthermore, we decompose $\gamma = \gamma_2 \circ \gamma_1$ as in lemma 4.19. Applying the triangle inequality for the geodesic distance $d$ yields\(^{33}\)

$$d_Q([q_1], [q_2]) \leq d(\tilde{q}_1, \sqrt{\gamma} q_2 \circ \gamma) = d(\sqrt{\gamma_1} q_1 \circ \gamma_1, \sqrt{\gamma_2} q_2 \circ \gamma_2)$$

$$\leq d(\sqrt{\gamma_1} q_1 \circ \gamma_1, \sqrt{\gamma_1} q_3 \circ \gamma_1) + d(\sqrt{\gamma_1} q_3 \circ \gamma_1, \sqrt{\gamma_2} q_2 \circ \gamma_2)$$

$$\leq d(\tilde{q}_1, \tilde{q}_3) + d(\tilde{q}_3, \sqrt{\gamma} q_2 \circ \gamma)$$

$$\leq d_Q([q_1], [q_3]) + d_Q([q_3], [q_2]) + \epsilon$$

\(^{32}\)In the end, we still have to account for the rotational action of $SO(2)$. To leave this as simple as possible, we will continue to ignore this.

\(^{33}\)Note that we have to be careful not to apply the triangle inequality of $S(V)$ to curves that are not of unit length, which might arise under composition of $\gamma \in D^*$. Although the expression $d(q_3, \sqrt{\gamma} q_2 \circ \gamma)$ is well defined by lemma 4.17, the triangle inequality might not hold in this case.
To conclude \( d(\bar{q}_1, \bar{q}_3) = d(\sqrt{\gamma_1} \bar{q}_1 \circ \gamma_1, \sqrt{\gamma_1} \bar{q}_3 \circ \gamma_1) \) we used the specific structure of \( \gamma_1 \) in the decomposition of lemma 4.19. \( \gamma_1 \) is just the rescaled identity map with some intervals of constancy. These do not contribute to the integral as \( \dot{\gamma} = 0 \) on these intervals. This proves the triangle inequality, as the above holds for all \( \epsilon > 0 \).

**Definiteness:** clearly, if \( \{q_1\} = \{q_2\} \), then \( d([q_1], [q_2]) = 0 \), as we only need to chose the same representatives. Conversely, let \( d([q_1], [q_2]) = 0 \). This is equivalent to

\[
\sup_{\gamma \in D^*} \langle q_1, \sqrt{\gamma} q_2 \circ \gamma \rangle_{L^2} = 1.
\]

With theorem 4.10 we know that there exists \( \gamma^* \in D^* \) such that \( \langle q_1, \sqrt{\gamma^*} q_2 \circ \gamma^* \rangle_{L^2} = 1 \). Employing the Cauchy-Schwarz inequality for \( \langle \cdot, \cdot \rangle_{L^2} \) we have

\[
1 \leq \langle q_1, \sqrt{\gamma^*} q_2 \circ \gamma^* \rangle_{L^2} \leq \|q_1\|^2_{L^2} \cdot \|\sqrt{\gamma^*} q_2 \circ \gamma^*\|^2_{L^2} \overset{\text{lem. 4.17}}{=} \|q_1\|^2_{L^2} \cdot \|q_2\|^2_{L^2} = 1,
\]

where the last inequality was concluded using the first property of lemma 4.17. This sequence of inequalities establishes \( \|\sqrt{\gamma^*} q_2 \circ \gamma^*\|^2_{L^2} = \|q_2\|^2_{L^2} = 1 \), as well as

\[
\langle q_1, \sqrt{\gamma^*} q_2 \circ \gamma^* \rangle_{L^2} = \|q_1\|^2_{L^2} \cdot \|\sqrt{\gamma^*} q_2 \circ \gamma^*\|^2_{L^2}.
\]

But equality in the Cauchy-Schwarz inequality holds only if \( q_1 \propto \sqrt{\gamma^*} q_2 \circ \gamma^* \). Thus we find \( q_1 = \sqrt{\gamma^*} q_2 \circ \gamma^* \) a.e. (once again making use of the length constraint). As \( q_1, q_2 \) are continuous, we find \( q_1(t) = \sqrt{\gamma^*(t)} q_2 \circ \gamma(t) \) for all \( t \in I \). Finally, we use the second property of lemma 4.17 to conclude \( q_1 \sim_F q_2 \).

We have now equipped \( S(V) \) with a reasonable equivalence relation and a well defined quotient metric. This has been one of our primary goals. However, there is another theoretical subtlety concerning the topology of this space. On the one hand, we have endowed \( S(V) \) with the weak topology induced by the geodesic distance metric. When we pass to the quotient \( Q = S(V)/ \sim_F \), this induces a quotient topology \( \tau_Q \) on \( Q \). On the other hand, we have defined the metric \( d_Q \) as in theorem 4.21. This in turn induces its own metric topology \( \tau_D \), which leaves us with yet another topology on \( Q \) ! It would be a huge disappointment if these weren’t related. However, so far it hasn’t been established that \( \tau_Q = \tau_D \). All we can show is the following.

**Lemma 4.22.** As above, let \( \tau_Q \) be the quotient topology of \( Q = S(V)/ \sim_F \) and let \( \tau_D \) be the metric topology induced by the metric \( d_Q \). Then \( \tau_D \subseteq \tau_Q \).

**Proof.** Let \( B^Q([q]) = \{ [p] \in Q \mid d_Q([p], [q]) < \epsilon \} \subset Q \) be a standard epsilon ball. It suffices to show that every such ball is open in the quotient topology. This can be done by noting that

\[
\pi \left( \bigcup_{\tilde{q} \in [q]} B_\epsilon(\tilde{q}) \right) = B^Q([q]),
\]

where \( \pi \) is the canonical projection. Indeed, for every \( \tilde{q} \in [q] \) we have

\[
\pi \left( B_\epsilon(\tilde{q}) \right) = \{ [p] \mid d(\tilde{q}, p) < \epsilon \} \subset B^Q([q]).
\]

Conversely, if \( [p] \in B^Q([q]) \) we will find \( \tilde{p} \in [p] \) and \( \tilde{q} \in [q] \) with \( d(\tilde{q}, \tilde{p}) < \epsilon \). This implies \( [p] \in \pi \left( B_\epsilon(\tilde{q}) \right) \). Thus, \( B^Q([q]) \) is the projection of an open set and, therefore, open in the quotient topology. \( \square \)
The difficulty of showing the missing inclusion $\tau_Q \subset \tau_D$ lies in the two-fold infimum definition of our metric. Generally, it does not suffice to align only one element and keep the other one fixed. It is true that we may rewrite the metric as one infimum over $D^*$ using eq. (31). However, this generally takes us outside the Fréchet equivalence class (e.g. by cutting out segments). Unfortunately, an equality like

$$d_Q([q_1], [q_2]) = \inf_{\gamma \in D^*} d(q_1, \sqrt{\gamma}q_1 \circ \gamma) \equiv \inf_{\tilde{q}_2 \in [q_2]} d(q_1, \tilde{q}_2)$$

has not been established or disproved, so far. We will come back to this in the conclusion chapter, when we discuss related open questions. For now, we have to be content with the result $\tau_D \subset \tau_Q$. For certain assertions this is already helpful. For instance, $\tau_D$, being induced by a metric, is Hausdorff. Thus, $\tau_Q$ must be Hausdorff itself.
Conclusion and Open Questions

Having succeeded in the construction of a shape space $Q$ and a well defined metric $d_Q$ for continuous shapes, we have achieved the primary goal of this dissertation. We now conclude this discussion with open questions related to this metric and space. We have seen the unsatisfactory result about the topology induced by $d_Q$ in relation to the natural quotient topology and how it is related to the two-fold infimum definition of the metric. More generally, we have the following sequence of inequalities

\[
\inf_{\gamma \in \text{Diff}(I)} d(q_1, \sqrt{\gamma} q_1 \circ \gamma) \overset{a)}{\geq} \inf_{\tilde{q}_2 \in [q_2]} d(q_1, \sqrt{\gamma} q_1 \circ \tilde{\gamma}) \overset{b)}{\geq} \inf_{\gamma \in D^*} d(q_1, \sqrt{\gamma} q_1 \circ \gamma). \tag{32}
\]

Establishing equality in either $a)$ or $b)$ would mean that $\inf_{\tilde{q}_2 \in [q_2]} d(q_1, \tilde{q}_2)$ is a good metric on its own, resolving the apparent asymmetry of this definition. Indeed, both the infimum over $\text{Diff}(I)$ and the infimum over $D^*$ are symmetric expressions. This would induce a topology on $Q$ which is fully compatible with the quotient topology. Proving equality in $a)$ would be enough to show that closed reparametrization group orbits are actually identical to the equivalence classes in Fréchet equivalence. We already proposed in section 4.3 to take the closure of these orbits as an equivalence relation, but discarded this suggestion as too implicit to work with. For now, all we know is that $[q]_{\text{Diff}} \subseteq \overline{[q]}_{\text{Diff}} \subset [q]_{\text{Fréchet}},$

where $\overline{A}$ denotes the closure of a set $A$ in the $L^2$ topology. Both approaches have their benefit. On the one hand, taking the $L^2$ closure avoids the pathologies of the set $D^*$. The geodesic distance on the space of SRVFs is an $L^2$-continuous mapping and exchanges of limits work rather naturally between the orbits of $\text{Diff}(I)$ and their closure. On the other hand, $D^*$ enjoys the compactness of the weak* topology. We employed this to prove that the minimum distance between Fréchet equivalence is actually attained.

Seeing the benefit of establishing $a)$ and $b)$, how could we actually prove these equalities? One might attempt to show that $\text{Diff}(I)$ is dense in $D^*$. As this is only about pointwise convergence, this does not sound very demanding. For instance, Corollary 8.1. in [30] p. 55, asserts that the set of continuous $\varphi$ is dense in $D^*$. Using sequences $\varphi_n = \frac{1}{n} \text{Id} + (1 - \frac{1}{n}) \varphi$ one concludes that (continuous) homeomorphisms with $\dot{\varphi} > 0$ lie dense in $D^*$, as well. It seems only a matter of the right smoothing procedure to establish that $\text{Diff}(I)$ is dense in $D^*$. However, then again we have to deal with the difficulty that $\varphi \mapsto d(q_1, \sqrt{\varphi} q_2 \circ \varphi)$ is not a continuous mapping $D^* \to \mathbb{R}$ (c.f. example 4.12). We could overcome this difficulty with the following conjecture, which is slightly stronger then asserting denseness of $\text{Diff}(I)$.

**Conjecture 4.23.** For every $\gamma \in D^*$, there exists a sequence $\{\gamma_n\}_{n \in \mathbb{N}} \subset \text{Diff}(I)$, such that

\[
\gamma_n \to \gamma, \text{ pointwise a.e. as } n \to \infty
\]

\[
\dot{\gamma}_n \to \dot{\gamma}, \text{ pointwise a.e. as } n \to \infty
\]

It is known that pointwise convergence of $\gamma_n$ does not generally imply convergence of the derivatives. For instance, $\gamma_n(t) = t - \frac{1}{2\pi n} \sin(2\pi nt)$ exhibits such behaviour. However,

---

34This sequence of inclusions is to be understood within $C^\infty(I, \mathbb{R}^2)$, i.e. all sets are implicitly intersected with smooth curves. It is unclear how this extends to $L^2$ functions.
A similar procedure may be applied to the negative part \( Y = [34] \) for details on this. In section 4.1 we have introduced the completion of all in all, we could achieve very weak assumption to find some establishing for any \( \gamma \). An application of Fatou’s lemma yields

\[
\limsup_{n \to \infty} E_{q_1, q_2}^+ [\gamma_n] \leq E_{q_1, q_2}^+ [\gamma].
\]

Now, pointwise convergence of the \( \gamma_n \) would give us pointwise convergence of the integrand. An application of Fatou’s lemma yields

\[
\liminf_{n \to \infty} E_{q_1, q_2}^+ [\gamma_n] \geq E_{q_1, q_2}^+ [\gamma]
\]

establishing

\[
\lim_{n \to \infty} E_{q_1, q_2}^+ [\gamma_n] = E_{q_1, q_2}^+ [\gamma].
\]

A similar procedure may be applied to the negative part

\[
E_{q_1, q_2}^- [\gamma] = \int_0^1 \max(-\langle q_1(t), \sqrt{\gamma(t)}q_2 \circ \gamma(t) \rangle_{\mathbb{R}^2}, 0) dt.
\]

All in all, we could achieve

\[
\lim_{n \to \infty} \langle q_1, \sqrt{\gamma_n}q_2 \circ \gamma_n \rangle_{L^2} = \langle q_1, \sqrt{\gamma}q_2 \circ \gamma \rangle_{L^2}
\]

for any \( \gamma \in D^* \) and the particular sequence of the conjecture 4.23. This is equivalent to

\[
\lim_{n \to \infty} d(q_1, \sqrt{\gamma_n}q_2 \circ \gamma_n) = d(q_1, \sqrt{\gamma}q_2 \circ \gamma),
\]

and would establish both equalities in eq. (32), when applied to a global minimizer \( \gamma^* \in D^* \) of the right hand side.

All arguments about the optimal matching of \( q_1 \) and \( q_2 \) have been, so far, restricted to continuous \( q \)’s. The assumption of continuity seems quite essential for the arguments leading to the existence of an optimal \( \gamma \in D^* \). Please refer to the paper of Trouvé and Younes [34] for details on this. In section 4.1 we have introduced the completion of \( S(V) \) in the weak \( L^2 \) topology. From a theoretical perspective it would be desirable to extend all our results to this space, as this would hold even non-smooth (i.e. not \( C^\infty \)) optimal matchings and guarantee the existence of generalized geodesics. Sadly, none of these issues have been resolved within this project. It is not even clear whether the full Fréchet equivalence class of some \( q \in L^2 \) is closed in \( L^2 \). \( L^2 \) convergence on its own seems like a very weak assumption to find some \( \gamma \in D^* \) that relates a limit point to an equivalence class. One might try to work with \( L^2 / \sim_F \) and define a similar metric as in theorem 4.21. However, as none of the results of theorem 4.10 may be applied, it is not clear how one might prove that this metric is well defined. In this extended set-up it seems even more natural to chose the ‘closure approach’ and see how it relates to Fréchet equivalence.

Beyond the SRV representation the phenomenon of unclosed reparametrization orbits also occurs for other metrics. In fact, it generally arises in the strong topology of \( C^\infty(I, \mathbb{R}^2) \) as a Fréchet space. This has been pointed out in example 3.2. The results established in this dissertation heavily rely on our particular choice of metric. It seems desirable to abstract these and make them applicable to other metrics as well. Fréchet equivalence is a solution to the closure problem which does not depend on the \( L^2 \) metric. Therefore, one could hope that it also works for more general metrics. In particular, it seems plausible that it would apply to the general class of elastic metrics \( G^{a,b} \) introduced in section 3.4.

All in all, this leaves many possibilities for subsequent research. Having outlined this bright perspective for further development, this dissertation shall come to an end. The author looks forward to reading about new research on this in the coming years.
A Implementation of Elastic Curve matching

This appendix is devoted to a modification of the already existing algorithms for elastic shape matching. These were introduced in Mio et al. [26]. The core part of the implementation is based on a dynamic programming procedure that performs the optimization over Diff(I). Minor modifications have to be made to carry out the extended optimization over D∗. For completeness, we describe the whole algorithm. In this project the code was implemented in Matlab.

A.1 Working with discretized shapes

One challenge in the implementation of elastic shape matching lies in the discretized representation of continuous curves. We will assume that a preprocessing procedure passes a discrete sample of some curve $c \in C^\infty(I, \mathbb{R}^2)$ as a vector $c \in \mathbb{R}^{2 \times N}$ to our program. A priori $N \in \mathbb{N}$ may not be the same for each curve. Some up- or down sampling algorithm may be used, however, to take care of that and we assume that $N$ is the same for all discretized curves. The original parametrization is not relevant. A vector $c$ will be interpreted in any case as $c_i = c((i - 1)/(N - 1))$, which implicitly assumes the sample points to be equally spaced. This, in turn, fixes the parametrization.

To compute the SRV representation of a curve $c$ we use finite differences

$$c'((i - 1)/(N - 1)) \approx \frac{c_{i+1} - c_{i-1}}{2h} =: \text{D}_c i,$$

with $h = 1/(N - 1), i = 2 \ldots, N - 1$.

This approximates the tangent vector of $c$. The discrete SRVF is now obtained as

$$q_i = \frac{\text{D}_c i}{\sqrt{\|\text{D}_c i\|}_{\mathbb{R}^2}}.$$

Integration is generally performed using a trapezoidal rule

$$\int_0^1 f(x)dx \approx h \frac{1}{2} \sum_{i=1}^{N-1} f_i + f_{i+1}. \quad (33)$$

This is used to approximate the $L^2$ inner product for discrete SRVFs $q_1, q_2 \in \mathbb{R}^{2 \times N}$, as well as to change back to the original curve (c.f. eq (21) in section 4.1).

One major issue in working with discrete curves is how to implement the composition $c \circ \gamma$ with a $\gamma \in D^*$. Within this project, this was generally dealt with using a cubic spline interpolation at the new sample nodes $\tilde{t}_i = \gamma((i - 1)/(N - 1))$. This way one can implement the reparametrization, despite working only with a discretized shape. However, in light of the discussion in section 4.5 one should be careful to apply a discontinuous $\gamma$ to a generic shape, as the SRV representation cannot store any information on relative placement of connected components. We introduced the possibility of decomposing such a $\gamma$ as

$$\gamma = \gamma_2 \circ \gamma_1^-$$

with continuous $\gamma_2$ and $\gamma_1$. The existence of such a decomposition was proven in lemma 4.19. The prove was a direct construction and can be implemented straightforwardly. Therefore, in a shape matching procedure of curves $c_1, c_2$, one should not apply a discontinuous $\gamma$ (which may arise as optimal matching prescription) to one shape, but rather apply the continuous changes of parameter $\gamma_2$ and $\gamma_1$ to both shapes.

\textsuperscript{35}One-sided differences at the boundaries
A.2 Dynamic Programming

Dynamic programming is a general procedure that may be applied to optimization problems that can be broken down into simpler parts. In our case we consider problems like

$$\text{minimize } E[J; \gamma] \text{ over } \gamma,$$

where $$\gamma : J \to [0, 1]$$ may be some generic non-decreasing function on the interval $$J$$, for instance, a diffeomorphism. The class of functionals $$E$$ this procedure may be applied to has the following property. If $$J = J_1 \cup J_2$$ is a disjoint union of intervals $$J_1, J_2$$, then

$$E[J_1; \gamma|_{J_1}] + E[J_2; \gamma|_{J_2}] = E[J; \gamma]. \tag{34}$$

Optimal matching of SRVFs $$q_1, q_2$$ corresponds to minimizing

$$E[\gamma] = -\int_0^1 \sqrt{\dot{\gamma}(t)} \langle q_1(t), q_2 \circ \gamma(t) \rangle \mathbb{R}^2 dt. \tag{35}$$

This is obviously divisible as in eq. (34). We usually refer to $$E$$ as an energy or cost of the path $$\gamma$$. The idea of the dynamic programming approach is to replace $$\gamma : [0, 1] \to [0, 1]$$ by a piecewise linear function on some $$n \times n$$ grid of $$[0, 1] \times [0, 1]$$

$$G = \{(i/(n-1), j/(n-1)) \mid i, j = 0, \ldots, n-1\},$$

and perform the optimization over this discrete set of $$\gamma$$’s. If $$\gamma$$ passes $$m$$ different nodes $$(k_s, l_s) \in G, s = 1, \ldots, m$$ on the grid, we can break up $$[0, 1]$$ into the intervals $$J_s = [k_s, k_{s+1}]$$ and, correspondingly, its cost as

$$E[[0, 1], \gamma] = \sum_{s=1}^{m-1} E[J_s, \gamma|_{J_s}].$$

This leads us to define the cost of a linear segment $$\gamma_{(k,l) \to (i,j)}$$ joining $$(k,l), (i,j) \in G$$ as

$$E[k,l; i,j] = E[[k,l], \gamma_{(k,l) \to (i,j)}],$$

with $$(k,l) < (i,j).$$\footnote{For two nodes $$(k,l)$$ and $$(i,j)$$ we write $$(k,l) < (i,j)$$ to mean $$k \leq i$$ and $$l \leq j$$, but not $$k = i$$ and $$l = j$$ at the same time.} For $$i = k$$ we take $$E[k,l; k,j] = 0$$. As this corresponds to a vanishing interval of integration, this is justified. Moreover, for the specific functional (35) this is symmetric to the case $$l = j$$, only switching the $$q$$’s. We now iteratively define a function $$H : G \to \mathbb{R}$$ that captures the minimal cost needed to reach a certain node in G.

```
for i, j = 0, \ldots, n - 1 do
    if i = 0 and j = 0 then
        V(0, 0, :) = (0, 0);
        H(0, 0) = 0;
    else
        \( (\hat{k}, \hat{l}) := \arg\min_{(k,l) < (i,j)} H(k,l) + E[k,l; i,j]; \)
        V(i, j, :) = (\hat{k}, \hat{l});
        H(i, j) = E[\hat{k}, \hat{l}; i,j] + H(\hat{k}, \hat{l});
    end if
end for
```
The array $V$ is of dimension $n \times n \times 2$ and saves the previous node that lead to a certain node $(i,j)$. The optimal path $\gamma$ can now be found by tracing back these nodes. This may be done as follows.

\[
W(0,:) = V(n-1, n-1,:);
\]
\[
a = 0;
\]

\[\text{while } W(a,:) \neq (0,0) \text{ do}
\]
\[a++;\]
\[W(a,:) = V(W(a-1,1), W(a-1,2,:));\]
\[\text{end while}\]

The array $W$ now contains the nodes for the optimal $\gamma$, in reverse order.

![Intensity plot of $H$ with optimal $\gamma$](image1.png)

![Restricted Neighbourhood $N_{(i,j)}$](image2.png)

Figure 14: a) Intensity plot of $H$ with optimal $\gamma$; b) restricted Neighbourhood $N_{(i,j)}$

Reparametrization with $\gamma$ may be performed as was described in section A.1, in combination with the decomposition of $\gamma$ into two continuous parts. Turning to the computational cost of this algorithm, we find that it needs $\frac{1}{4}n^2(n+1)^2 - n^2 = \mathcal{O}(n^4)$ evaluations of the energy functional. This can be very costly, as the evaluation of the energy functional itself needs a spline interpolation and an integration. To reduce this cost one performs the search for the optimal $(\hat{k}, \hat{l})$ only in a limited neighbourhood $N_{(i,j)}$ of $(i,j)$. This way one achieves $\mathcal{O}(n^2)$ evaluations of the energy functional, significantly increasing the performance. An example of such a neighbourhood is sketched in figure 14(b).

Remark A.1. This is almost the same algorithm as used by Mio et al.. However, they never allow horizontal or vertical segments. Using a similar restricted neighbourhood $N_{i,j}$, this effectively places an upper bound on the admitted slope of $\gamma$. Presumably, this is the reason that their results never feature singular changes of parameter. Having removed this restriction, the optimization is now performed over functions $\gamma \in D^*$, as well.
A.3 Further examples of Shape Matching

We conclude the discussion of the implementation by showing some more examples, produced by this algorithm. As rotational alignment and closure constraints are not considered here, these results do not yet constitute ideal shape matchings. To take these features into account, one would need to combine the dynamic programming with an exhaustive search over $SO(2)$. Also one would have to allow for different initial points of the parametrization, in case of closed curves, effectively resulting in another exhaustive search over $SO(2)$. The grid size used here was $n = 100$. The neighbourhood $N_{i,j}$ was taken to be of size 5 in $x$ and $y$ direction. Each group of pictures features the optimal $\gamma$ (a), the initial parametrization of $q_1$ and $q_2$ (b)-(c), the final parametrization of $q_2$ (d) and the connecting geodesic (e). (b)-(d) are in low resolution to emphasize the changes in parameter. Gaps are to be interpreted as particularly stretched segments and do not occur in a higher resolution. Occurrence of discontinuities in the optimal $\gamma$ are particularly notable in figure 15(a) at $t = 0.14, t = 0.71$ and $t = 0.98$. Figure 16(a) features less pronounced discontinuities at $t = 0.42$ and $t = 1$. No singularities at all are seen in figure 17(a). Apparently, in this case the shapes are close enough not to require any ‘over-stretching’. The last shape matching, again, features a small discontinuity of the optimal $\gamma$ (figure 18(a)) at $t = 0.93$. Reparametrization was always performed on both shapes, using the decomposition $\gamma = \gamma_2 \circ \gamma_1^{-1}$ (as in lemma 4.19). However, only the changes of parameter in the second shape are notable, as $\gamma_1$ simply introduces intervals of constancy to the first shape. These examples further emphasize that discontinuous optimal matching prescriptions do arise for generic shape constellations and not only in examples that are specifically designed to exhibit such behaviour.

![Figure 15: Geodesic distance $d = 0.678$](image)
Figure 16: Geodesic distance \( d = 0.747 \)

Figure 17: Geodesic distance \( d = 0.345 \)

Figure 18: Geodesic distance \( d = 0.751 \)
References


