Journal of Nonlinear Science (2024) 34:56
https://doi.org/10.1007/s00332-024-10035-5

Journal of Nonlinear Science

Elastic Metrics on Spaces of Euclidean Curves: Theory and Algorithms

Martin Bauer[1,4] · Nicolas Charon[2] · Eric Klassen[1] · Sebastian Kurtek[3] · Tom Needham[1] · Thomas Pierron[1]

Received: 26 May 2023 / Accepted: 21 March 2024 / Published online: 25 April 2024

Abstract

A main goal in the field of statistical shape analysis is to define computable and informative metrics on spaces of immersed manifolds, such as the space of curves in a Euclidean space. The approach taken in the elastic shape analysis framework is to define such a metric by starting with a reparametrization-invariant Riemannian metric on the space of parametrized shapes and inducing a metric on the quotient by the group of diffeomorphisms. This quotient metric is computed, in practice, by finding a registration of two shapes over the diffeomorphism group. For spaces of Euclidean curves, the initial Riemannian metric is frequently chosen from a two-parameter family of Sobolev metrics, called elastic metrics. Elastic metrics are especially convenient because, for several parameter choices, they are known to be locally isometric to Riemannian metrics for which one is able to solve the geodesic boundary problem explicitly—well-known examples of these local isometries include the complex square root transform of Younes, Michor, Mumford and Shah and square root velocity (SRV) transform of Srivastava, Klassen, Joshi and Jermyn. In this paper, we show that the SRV transform extends to elastic metrics for all choices of parameters, for curves in any dimension, thereby fully generalizing the work of many authors over the past two decades. We give a unified treatment of the elastic metrics: we extend results of Trouvé and Younes, Bruveris as well as Lahiri, Robinson and Klassen on the existence of solutions to the registration problem, we develop algorithms for computing distances

Communicated by David Martín de Diego.

✉ Nicolas Charon
ncharon@central.uh.edu

1 Department of Mathematics, Florida State University, Tallahassee, FL, USA

2 Department of Mathematics, University of Houston, Houston, TX, USA

3 Department of Statistics, Ohio State University, Columbus, OH, USA

4 Department of Mathematics, University of Vienna, Vienna, Austria

and geodesics, and we apply these algorithms to metric learning problems, where we learn optimal elastic metric parameters for statistical shape analysis tasks.

Keywords Shape analysis · Elastic metrics · Infinite-dimensional Riemannian geometry · Metric learning

Mathematics Subject Classification 49Q10 · 53Z50 · 68T10

1 Introduction

Shape is a fundamental physical property of objects and a key characteristic of their appearance in images. As a result, shape analysis plays a central role in various applications including computer vision, medical imaging, graphics, biology, bioinformatics and anthropology, among others. In these applications, one generally first extracts objects of interest from the imaging data and then studies their shapes via appropriate mathematical representations and metrics. In statistical shape analysis, each observed shape is treated as a random object with the primary goal of developing tools for shape registration, comparison, statistical summarization, exploration of variability, clustering, classification and other statistical procedures. Each of the aforementioned statistical tasks heavily depends on the underlying representation and associated metric chosen for shape analysis.

There is a rich literature on shape analysis that considers various representations of shape including deformable templates (Grenander and Miller 1998), ordered and unordered point sets (Dryden and Mardia 1998), level sets (Osher and Fedkiw 2003), medial axes (Gorczowski et al. 2010) and others. However, perhaps the most natural representation of a boundary of an object captured in an image is a parametrized curve. While accounting for the shape preserving transformations of rigid motion and scaling is fairly standard in this setting (Srivastava and Klassen 2016), one must additionally deal with parametrization variability inherent in the given data. Some past methods standardize parametrizations of observed curves to arc length (Zahn and Roskies 1972), but this has been shown to be suboptimal in many applications (Srivastava and Klassen 2016). A better solution is to determine optimal reparametrizations in a pairwise manner via a process referred to as *registration*. This, in turn, requires a metric on the space of parametrized curves that is invariant to reparametrizations. The metric plays a key role in shape analysis as it is used for joint registration and comparison of shapes. Further, it serves as a backbone of other statistical procedures for shape data including averaging and principal component analysis.

In this article, we focus on shapes which are represented as curves in Euclidean space $\mathbb{R}^d$, $d \geq 2$. Our shape metrics arise as geodesic distances with respect to Riemannian metrics on the (infinite-dimensional) manifold $\mathcal{I}(D, \mathbb{R}^d)$, whose points are immersions, with domain D either an interval or a circle. That is, each element of $\mathcal{I}(D, \mathbb{R}^d)$ is a smooth parametrized curve $c : D \to \mathbb{R}^d$ with nowhere-vanishing derivative. In order to induce a metric on the shape space of unparametrized curves, one requires the Riemannian metric on $\mathcal{I}(D, \mathbb{R}^d)$ to be invariant under reparametrizations. To be precise, the group $\mathcal{D}(D)$ of diffeomorphisms of D acts on $\mathcal{I}(D, \mathbb{R}^d)$ by precomposition,

and this action should be by isometries for the chosen Riemannian metric. Due to this property, the metric then descends to the quotient space $\mathcal{I}(D, \mathbb{R}^d)/\mathcal{D}(D)$ of curves considered up to precomposition with a diffeomorphism—that is, the quotient space can be considered as the space of unparametrized curves. The process of computing geodesic distances in that quotient space naturally involves solving a registration problem, so that the estimated geodesic distance eventually provides a meaningful metric for shape comparison. Moreover, the Riemannian formalism gives powerful tools for demonstrating the existence of optimal registrations, along with well-defined notions of tangent spaces, means, principal components and other statistical concepts.

In this setup, a variety of different Riemannian metrics have been proposed in the literature. The arguably simplest one, the invariant L^2-metric, has a surprising degeneracy: it induces vanishing geodesic distance on both parametrized (Bauer et al. 2012) and unparametrized (Mumford and Michor 2006) curves, i.e., any two curves (shapes) are regarded the same under the corresponding path-length distance. This behavior renders the L^2-metric unsuitable for any applications in shape analysis. Subsequently, several stronger Riemannian metrics have been proposed that consequently induce a meaningful measure of similarity on shape space. This includes the class of almost local metrics (Mumford and Michor 2006), but also the family of (higher order) Sobolev type metrics, see e.g. Michor and Mumford (2007), Bauer et al. (2014), Srivastava et al. (2010), Sundaramoorthi et al. (2007) and the references therein. In particular, Sobolev metrics of order one have attracted a large body of work and a two-parameter family of Riemannian metrics $G^{a,b}$, $a, b > 0$, has been proposed (Mio et al. 2007); elements of this family are usually called *elastic metrics*, for reasons that are highlighted in "Appendix A." The goal of this paper is to develop a comprehensive theoretical and computational framework for the $G^{a,b}$ metrics on $\mathcal{I}(D, \mathbb{R}^d)$ and the quotient $\mathcal{I}(D, \mathbb{R}^d)/\mathcal{D}(D)$, for all parameters $a, b > 0$ and all dimensions $d \geq 2$. Before precisely stating our main contributions, we first give an overview of related work to provide appropriate context.

A crucial component of an efficient algorithm for computing the desired geodesic distances in the quotient space $\mathcal{I}(D, \mathbb{R}^d)/\mathcal{D}(D)$ is a method for computing distances in the space $\mathcal{I}(D, \mathbb{R}^d)$. The family of elastic metrics is special: it has been shown that, for several values of the parameters a, b and d, the geodesic boundary value problem, and consequently the induced geodesic distance, can be solved *explicitly*. The typical method in the literature for deriving such explicit solutions is to construct an isometry $(\mathcal{I}(D, \mathbb{R}^d), G^{a,b}) \to (\mathcal{M}, G)$, for particular values of a, b and d, to some Riemannian manifold with explicit formulas for geodesics, and to then describe geodesics in the source space via pulling back. For the choice of parameters $a = 1$ and $b = \frac{1}{2}$, such an isometry is given by the well known *square root velocity transform*, as developed in Srivastava et al. (2010) for arbitrary $d \geq 2$. For $a = b = \frac{1}{2}$ and $d = 2$, an isometry is given by the *complex square root mapping* constructed in Younes et al. (2008), which is based on identifying $\mathbb{R}^2$ with the complex plane. The complex square root mapping was generalized to curves in $\mathbb{R}^3$ by replacing complex constructions with their quaternionic counterparts in Needham (2018, 2019b). For $a \leq 2b$ and $d = 2$, a related construction has been developed in Bauer et al. (2014), where the target manifold is a space of curves in a Euclidean cone. Finally, for curves with values in $\mathbb{R}^2$, the transformations of Younes et al. (2008), Srivastava et al. (2010), Bauer

et al. (2014) have been extended to general values of the parameters a and b using a local isometry defined, once again, via the identification of $\mathbb{R}^2$ with the complex plane (Needham and Kurtek 2020).

We can now state our main contributions and outline the structure of the paper.

- *Simplifying isometry for general elastic metric parameters (Sect. 2).* We show that the square root velocity transform mentioned above can, in fact, be used as a simplifying isometry of $G^{a,b}$ for general parameters a and b and for curves in arbitrary dimension (Theorem 2.1), thereby fully generalizing the results of Srivastava et al. (2010), Bauer et al. (2014), Needham and Kurtek (2020) and completing the story started over a decade ago in Younes et al. (2008). We use this result to characterize the metric completion of the space of curves and give an explicit formula for the distance in this space (Theorem 2.1). This completion includes curves of lower regularity—in particular, it includes the class of piecewise linear curves, which is important for representing smooth curves in a discrete computational setting. We also present a (to our knowledge, novel) relation between elastic metrics and classical elasticity theory ("Appendix A").
- *Existence of optimal reparametrizations (Sect. 3).* To obtain the distance on the quotient shape space, one needs to consider the following problem: given two curves $c_1, c_2 \in \mathcal{I}(D, \mathbb{R}^d)$, find a reparametrization $\gamma \in \mathcal{D}(D)$ such that the geodesic distance between c_1 and $c_2 \circ \gamma$, with respect to a given elastic metric $G^{a,b}$, is minimized. It has been shown that a minimizer exists (within certain extensions of the set $\mathcal{D}(D)$) for parameters $a = 1$ and $b = \frac{1}{2}$ (the original setting of the square root velocity transform), under certain regularity assumptions on the curves c_i (Lahiri et al. 2015; Bruveris 2016). We extend these results to general parameters $a, b > 0$, under technical regularity assumptions (Theorems 3.1 and 3.2) and lay out some precise open questions regarding dependence on regularity. Our proof uses classical results of Trouvé and Younes (2000)
- *Computational framework (Sect. 4).* We develop a comprehensive framework for computing geodesics for $G^{a,b}$-metrics in the quotient space $\mathcal{I}(D, \mathbb{R}^d)/\mathcal{D}(D)$ and arbitrary values of a and b. This computation involves optimizing geodesic distance over reparametrizations, as was described in the previous paragraph. Therefore, we generalize two algorithms, that have been originally developed for the case $a = 1$, $b = \frac{1}{2}$ to the present situation (Lahiri et al. 2015; Trouvé and Younes 2000; Mio et al. 2007; Srivastava and Klassen 2016): the first one is an explicit polynomial-time algorithm to find exact solutions in the setting of piecewise linear curves (Theorem 4.1) and the second one is a (faster) dynamic programming algorithm for approximating the optimizer (Sect. 4.2). We illustrate this approach with several computational examples (Sect. 4.3). Our code is available under an open source license.[1]
- *Metric Learning (Sect. 5).* Finally, we consider the following question: given a dataset of shapes, which metric from the family of elastic metrics gives the best performance on various statistical analysis tasks? We frame this as a *metric learning* problem. A general approach to learning the appropriate parameters for a given dataset is suggested (Sect. 5.1), based on foundational work in metric learn-

[1] https://github.com/charoncode/Gab_metrics

ing (Xing et al. 2002). We also give an alternative approach to parameter estimation with the aim of training for geometric protein classification (Sect. 5.2).

2 Simplifying Transform for General Elastic Metrics

In this section, we will introduce the basic concepts and spaces under consideration and introduce the class of Riemannian metrics that will be of central interest. We then define a new family of transforms for simplifying these Riemannian metrics.

2.1 Spaces of Curves and Elastic Metrics

In the following, assume $d \geq 2$ and let

$$\mathcal{I}(D, \mathbb{R}^d) = \{c \in C^\infty(D, \mathbb{R}^d) \,:\, c'(u) \neq 0 \; \forall u \in D\}\,,$$

where D is either the interval $I = [0, 1]$ for open curves or the unit circle S^1 for closed curves. This space is an open subset of the Fréchet space $C^\infty(D, \mathbb{R}^d)$ and is thus an infinite-dimensional manifold with tangent space $T_c\mathcal{I}(D, \mathbb{R}^d)$ at a curve $c \in \mathcal{I}(D, \mathbb{R}^d)$ satisfying $T_c\mathcal{I}(D, \mathbb{R}^d) \approx C^\infty(D, \mathbb{R}^d)$; specifically, the identification is made by taking $C^\infty(D, \mathbb{R}^d)$ to be the space of smooth vector fields (or *deformation fields*) along c.

This article is concerned with the family of elastic $G^{a,b}$-metrics on $\mathcal{I}(D, \mathbb{R}^d)$, indexed over pairs of positive constants $a, b > 0$:

$$G^{a,b}_c(h, k) = \int_D a^2(D_s h^\perp \cdot D_s k^\perp) + b^2(D_s h^\top \cdot D_s k^\top) ds\,, \tag{1}$$

where $h, k \in T_c\mathcal{I}(D, \mathbb{R}^d)$ are deformation fields (tangent vectors) to the curve $c \in \mathcal{I}(D, \mathbb{R}^2)$, $\cdot$ denotes the Euclidean inner product with norm $|\cdot|$ (evaluated pointwise), $D_s = \frac{1}{|c'|}\frac{rmd}{rmdu}$ and $ds = |c'|du$ are differentiation and integration with respect to arc length, respectively (u denoting the parameter of c), and $\bullet^\top$ and $\bullet^\perp$ denote projection onto the normal and tangential part of a tangent vector, i.e.,

$$D_s h^\top = \left(D_s h \cdot \tfrac{c'}{|c'|}\right)\tfrac{c'}{|c'|} \qquad \text{and} \qquad D_s h^\perp = D_s h - D_s h^\top.$$

The terminology of "elastic metrics" for (1) often used in the literature (Younes 1998; Mio et al. 2007; Jermyn et al. 2012; Needham 2019a) can be in fact justified from the theory of linear material elasticity, specifically as the limit of the linear elastic energy of a deforming shell as it becomes infinitely thin. Such a connection was recently emphasized in Charon and Younes (2022) for the class of first-order metrics on surfaces. We provide in "Appendix A" a similar and more direct derivation in the case of parametrized planar curves.

The group $\mathbb{R}^d$ acts on $\mathcal{I}(D, \mathbb{R}^d)$ by rigid translations. Each bilinear form (1) is degenerate on the space of all curves and therefore only defines a Riemannian metric

on the quotient $\mathcal{I}(D, \mathbb{R}^d)/\mathbb{R}^d$, which can be identified with the space $\mathcal{I}_0(D, \mathbb{R}^d)$ of curves starting at the origin. Once we have defined a Riemannian metric, we can consider the corresponding geodesic distance function

$$d_{a,b}(c_0, c_1) = \inf \int_0^1 \sqrt{G_c(c_t, c_t)} \mathrm{d}t,$$

where the infimum is taken over all paths $c : [0, 1] \to \mathcal{I}_0(D, \mathbb{R}^d) : t \mapsto c_t$ interpolating between the curves c_0 and c_1. For finite-dimensional Riemannian manifolds, geodesic distance is indeed a true metric, but this is not necessarily true in infinite dimensions: there are Riemannian metrics such that the corresponding geodesic distance function is degenerate or might even vanish identically (Mumford and Michor 2006; Bauer et al. 2020). For the $G^{a,b}$-metrics, this misbehavior has been ruled out (Mumford and Michor 2006; Srivastava et al. 2010), which consequently renders them as viable candidates for shape analysis.

On the space of immersions, there is a natural action by the orientation-preserving diffeomorphism group $\mathcal{D}(D)$ of the domain D: the reparametrization action. Given a curve $c \in \mathcal{I}_0(D, \mathbb{R}^d)$ and a diffeomorphism $\varphi \in \mathcal{D}(D)$, this action is given by composition from the right, i.e.,

$$\mathcal{I}_0(D, \mathbb{R}^d) \times \mathcal{D}(D) \to \mathcal{I}_0(D, \mathbb{R}^d), \qquad (c, \varphi) \mapsto c \circ \varphi.$$

A straightforward calculation shows that the $G^{a,b}$-metrics are invariant under this action, i.e.,

$$G^{a,b}_c(h, k) = G^{a,b}_{c\circ\varphi}(h \circ \varphi, k \circ \varphi), \qquad h, k \in T_c\mathcal{I}_0(D, \mathbb{R}^d), \quad \varphi \in \mathcal{D}(D).$$

Consequently, they descend to Riemannian metrics on the quotient shape space of immersions modulo parametrizations $\mathcal{S}(D, \mathbb{R}^d) := \mathcal{I}_0(D, \mathbb{R}^d)/\mathcal{D}(D)$. On the quotient space, the corresponding geodesic distance function can be calculated via

$$d^{\mathcal{S}}_{a,b}([c_0], [c_1]) = \inf_{\varphi\in\mathcal{D}(D)} d_{a,b}(c_0, c_1 \circ \varphi).$$

2.2 The Square Root Velocity Transform for General Elastic Metrics

As we overviewed in the introduction, there have been many approaches in the literature to understanding elastic metrics through *simplifying transformations* (Younes et al. 2008; Srivastava et al. 2010; Bauer et al. 2014; Needham and Kurtek 2020). That is, these works establish (local) isometries of the form $(\mathcal{I}(D, \mathbb{R}^d), G^{a,b}) \to (\mathcal{M}, G)$, for some choice of parameters a, b, and d, where the target space is some Riemannian manifold with an easy-to-describe geodesic structure. Such a transformation allows efficient computations involving the elastic metric by transferring them to the simple target space. Of particular interest for this paper is the *square root velocity transform*

of Srivastava et al., which we denote as

$$\begin{aligned} R : \mathcal{I}_0([0,1],\mathbb{R}^d) &\to C^\infty([0,1],\mathbb{R}^d \setminus \{0\}) \\ c &\mapsto \frac{c'}{\sqrt{|c'|}}. \end{aligned} \tag{2}$$

It was shown in Srivastava et al. (2010) that R is an isometry of the elastic metric $G^{1,\frac{1}{2}}$ and the standard L^2 metric on $C^\infty([0,1],\mathbb{R}^d \setminus \{0\})$, for any $d \geq 2$. Our first main result below will show that R is, for general parameters a, b, an isometry of $G^{a,b}$ and a Riemannian metric on $C^\infty([0,1],\mathbb{R}^d \setminus \{0\})$ which is non-Euclidean, but still simple enough to admit explicit geodesic distances.

In order to formulate this result, we first introduce a Riemannian metric on $\mathbb{R}^d \setminus \{0\}$. For $q \in \mathbb{R}^d \setminus \{0\}$, we identify $T_q\left(\mathbb{R}^d \setminus \{0\}\right)$ with $\mathbb{R}^d$ in the obvious way. We then decompose the tangent space via $T_q\left(R^d \setminus \{0\}\right) = V \oplus W$, where $V = \mathbb{R}\, q$ and $W = V^\perp$, where the orthogonal complement is with respect to the standard dot product on $\mathbb{R}^d$. We then define a Riemannian metric g^λ on $T_q\left(\mathbb{R}^d \setminus \{0\}\right)$ as follows:

$$g^\lambda_q(v,w) = \lambda^2(v^\perp \cdot w^\perp) + (v^\top \cdot w^\top),$$

where, in analogy with the notation used in Sect. 2.1, $w^\top = \frac{(w\cdot q)}{(q\cdot q)}q$ is the projection of w onto V and $w^\perp = w - w^\top$ is the projection of w onto W. For $\lambda = 1$, this is just a restriction of the standard Euclidean metric on $\mathbb{R}^d$; for $\lambda < 1$, it makes $\mathbb{R}^d \setminus \{0\}$ isometric to a dense subset of a cone in $\mathbb{R}^{d+1}$ with an acute angle at the cone point; for $\lambda > 1$, $(\mathbb{R}^d \setminus \{0\}, g^\lambda)$ does not isometrically embed in $\mathbb{R}^{d+1}$.

The metric g^λ induces an L^2-metric on the space of smooth curves in $\mathbb{R}^d \setminus \{0\}$:

$$G^{L^2_\lambda}_q(q_1,q_2) = \int_0^1 g^\lambda_q(q_1,q_2)du.$$

We now state our first main result, whose proof is postponed to "Appendix C."

Theorem 2.1 *For $\lambda = \frac{a}{2b}$, the square root velocity transform R, defined in (2), is an isometry of $\left(\mathcal{I}_0([0,1],\mathbb{R}^d), G^{a,b}\right)$ and $\left(C^\infty([0,1],\mathbb{R}^d\setminus\{0\}), 4b^2\, G^{L^2_\lambda}\right)$. Furthermore, for each $c_0 \in \mathcal{I}_0([0,1],\mathbb{R}^d)$ there exists a neighborhood $\mathcal{U}(c_0)$ such that the geodesic distance between c_0 and any $c_1 \in \mathcal{U}(c_0)$ is given by:*

$$dist_{a,b}(c_1,c_2) = 2b\sqrt{\ell_{c_1} + \ell_{c_2} - 2\int_D \sqrt{|c_1'||c_2'|}\cos\left(\tfrac{a}{2b}\theta\right)du}, \tag{3}$$

where

$$\theta(u) = \cos^{-1}(R(c_1)\cdot R(c_2)/|R(c_1)||R(c_2)|).$$

If $d \geq 3$, then the formula for the geodesic distance holds globally, i.e., for arbitrary $c_1 \in \mathcal{I}_0([0,1], \mathbb{R}^d)$, after replacing the formula for θ by

$$\theta(u) = min\left(\cos^{-1}(R(c_1) \cdot R(c_2)/|R(c_1)||R(c_2)|), \tfrac{2b\pi}{a}\right).$$

To deal with the difficulty in the $d = 2$ case (the formula of the geodesic distance being only valid locally), we can extend geodesics across the origin and obtain $C^\infty(I, \mathbb{R}^d_\lambda)$ as the geodesic completion in the sense of Wurzbacher (2004), Bruveris (2016). This allows us to interpret (3) as the geodesic distance on the geodesic completion. We will now extend the formula for the geodesic distance of Theorem 2.1 to the metric completion, which will be important in the next section where we will prove the existence of optimal reparametrizations.

Corollary 2.1 *The completion (in the sense of Lemma B.2) of $\mathcal{I}_0([0,1], \mathbb{R}^d)$ with the $G^{a,b}$ metric is the space of absolutely continuous open curves $AC_0([0,1], \mathbb{R}^d)$. For any two curves $c_1, c_2 \in AC_0([0,1], \mathbb{R}^d)$, their corresponding geodesic distance is given by*

$$dist_{a,b}(c_1, c_2) = 2b\sqrt{\ell_{c_1} + \ell_{c_2} - 2\int_0^1 \sqrt{|c_1'||c_2'|}\cos\left(\tfrac{a}{2b}\theta\right) du}.$$

where ℓ_{c_j} is the length of the curve c_j and

$$\theta(u) = \begin{cases} min\left(\cos^{-1}(c_1'(u) \cdot c_2'(u)/|c_1'(u)||c_2'(u)|), \tfrac{2b\pi}{a}\right) & \text{if } c_1'(u), c_2'(u) \neq 0 \\ \tfrac{2b\pi}{a} & \text{otherwise.} \end{cases}$$

2.3 Relation to Previous Work

In this subsection, we pin down the precise relationship between Theorem 2.1 and previous work on simplifying transforms (Younes et al. 2008; Srivastava et al. 2010; Bauer et al. 2014; Needham and Kurtek 2020).

The transform in the literature which is most relevant to our result is obviously the square root velocity transform (2), which was shown in Srivastava et al. (2010) to be an isometry of $G^{1,\frac{1}{2}}$ and the standard L^2 metric on $C^\infty([0,1], \mathbb{R}^d \setminus \{0\})$ for arbitrary $d \geq 2$. This result is recovered directly from Theorem 2.1.

The *complex square root map* of Younes et al. (2008) takes an immersion c in the plane to the curve $\sqrt{c'}$, where the square root is computed pointwise by considering c as a complex-valued function—there is some ambiguity here, so the square root curve is chosen in a way to make it continuous. This transform was shown to be a local isometry of $G^{\frac{1}{2},\frac{1}{2}}$ with the standard L^2 metric on $C^\infty([0,1], \mathbb{C}) \approx C^\infty([0,1], \mathbb{R}^2)$. In Needham and Kurtek (2020), it was shown that the complex square root map fits into a family of maps $F_{a,b}$, defined on a smooth plane curve c by $F_{a,b}(c) = 2b|c'|^{\frac{1}{2}}\left(c'/|c'|\right)^{\frac{a}{2b}}$, with exponentiation once again performed using the identification of $\mathbb{R}^2$ with the complex plane and choosing a continuous curve as the image; when $a = b = \frac{1}{2}$, $F_{a,b}$ reduces to

the complex square root map. It was shown in Needham and Kurtek (2020) that $F_{a,b}$ defines a local isometry between $G^{a,b}$ and the L^2 metric, for any choices of $a, b > 0$. In fact, the $F_{a,b}$ transform factors as

$$\left(\mathrm{Imm}_0([0,1],\mathbb{C}), G^{a,b}\right) \xrightarrow{R} \left(C^\infty([0,1],\mathbb{C}^*), 4b^2 G^{L^2_\lambda}\right) \xrightarrow{S_{a,b}} \left(C^\infty([0,1],\mathbb{C}^*), G^{L^2}\right)$$

$$\text{with } F_{a,b} = S_{a,b} \circ R$$

where $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$, $\lambda = \frac{a}{2b}$, $G^{L^2} = G^{L^2_1}$ is the standard L^2 metric and $S_{a,b}$ is a local isometry defined by $S_{a,b}(q) = 2b|q|^{1-\lambda}q^\lambda$—the local isometry claim can be seen via calculations in complex coordinates, similar to the proof of (Needham and Kurtek 2020, Theorem 2.3). The takeaway from Theorem 2.1 is that the geometry of the metric $G^{L^2_\lambda}$ is simple enough that we can work in the middle space of this diagram, allowing us to avoid technical issues with isometries only being local, while simultaneously allowing the result to be generalized to arbitrary dimension.

We should also mention (Bauer et al. 2014), which gave a similar family of generalizations of the complex square root map for plane curves, valid for $G^{a,b}$ with $a \leq 2b$. In this case, a curve c in $\mathbb{C}$ is mapped to the curve in $\mathbb{R}^3 \approx \mathbb{C} \times \mathbb{R}$ given by

$$R_{a,b}(c) = |c'|^{\frac{1}{2}} \left(a\frac{c'}{|c'|}, \sqrt{4b^2 - a^2} \right).$$

The image of $R_{a,b}$ lies on a certain cone in $\mathbb{R}^3$ and it is shown in Bauer et al. (2014) that the transform is an isometry of $G^{a,b}$ and the metric on the cone induced from the ambient Euclidean metric. As in the case of the $F_{a,b}$ transform described above, one can factor $R_{a,b}$ as the SRV transform R followed by an isometry between the space of curves in the plane and the space of curves in the cone.

3 Existence of Optimal Reparametrizations

3.1 Existence of Optimal Reparametrizations for Open Curves

In the previous section, we saw that the set of absolutely continuous functions provides the natural space for studying the geodesic distance function of the family of elastic metrics. In the following, we aim to prove the existence of optimal reparametrizations in this space. Let $\mathrm{dist}^{\mathcal{S}}_{a,b}$ be the induced distance of the elastic $G^{a,b}$-metric on the space of absolutely continuous, unparametrized and open curves $\mathcal{S}(I,\mathbb{R}^d) := \mathrm{AC}_0(I,\mathbb{R}^d)/\Gamma$, which is defined by

$$\mathrm{dist}^{\mathcal{S}}_{a,b}([c_1],[c_2]) = \inf_{\gamma\in\Gamma} \mathrm{dist}_{a,b}(c_1, c_2 \circ \gamma), \tag{4}$$

where $\mathrm{dist}_{a,b}$ denotes the geodesic distance function on the space of parametrized open curves $\mathrm{AC}_0(I,\mathbb{R}^d)$ and where Γ denotes the group of absolutely continuous

diffeomorphisms on I, i.e.,

$$\Gamma = \{\gamma \in AC(I, I) : \ \gamma(0) = 0, \ \gamma(1) = 1, \ \gamma' > 0 \quad a.e.\}.$$

We will also need the closure $\bar{\Gamma}$ (with respect to the $W^{1,1}$-norm topology) of Γ, which is the semigroup of weakly increasing, absolutely continuous functions, i.e.,

$$\bar{\Gamma} = \{\gamma \in AC(I, I) : \ \gamma(0) = 0, \ \gamma(1) = 1, \ \gamma' \geq 0 \quad a.e.\}.$$

With this notation, the induced distance $\text{dist}^{\mathcal{S}}_{a,b}$ on the quotient space $\mathcal{S}(I, \mathbb{R}^d)$ can be equivalently expressed via

$$\begin{aligned} \text{dist}^{\mathcal{S}}_{a,b}([c_1], [c_2]) &= \inf_{\gamma \in \Gamma} \text{dist}_{a,b}(c_1, c_2 \circ \gamma) = \inf_{\gamma_1, \gamma_2 \in \Gamma} \text{dist}_{a,b}(c_1 \circ \gamma_1, c_2 \circ \gamma_2) \\ &= \inf_{\gamma_1, \gamma_2 \in \bar{\Gamma}} \text{dist}_{a,b}(c_1 \circ \gamma_1, c_2 \circ \gamma_2). \end{aligned} \tag{5}$$

Here, the second equality in the first line follows from the invariance of the distance, whereas the third equality follows from the density of Γ in $\bar{\Gamma}$. Our main result of this section concerns the existence of optimal reparametrizations, i.e., the existence of reparametrization functions such that the infimum is attained. We will see that for $a \leq b$ we really need two reparametrization functions in $\bar{\Gamma}$, while for $a > b$ the infimum can attained by one reparametrization function. Before we formulate the theorem, we need to introduce the function space of piecewise differentiable functions:

$$\begin{aligned} PC^1(I, \mathbb{R}^d) := \Big\{ c \in C(I, \mathbb{R}^d) : \exists\, 0 = t_0 < t_1 < \ldots < t_n = 1 \\ \text{s.t. } c|_{(t_i, t_{i+1})} \in C^1\big((t_i, t_{i+1}), \mathbb{R}^d\big) \Big\}. \end{aligned}$$

Note that $PC^1(I, \mathbb{R}^d) \subset AC(I, \mathbb{R}^d)$. We also need to introduce the space $\mathcal{D}^*$ of all functions that can be written as $\phi(s) = \mu([0, s[)$, where μ is a probability measure on I. These are equivalently (c.f. Cohn 2013, Chapter 4) bounded variation (BV) functions, which are non-decreasing and left-continuous on I. We will denote the right limit of ϕ at x by $\phi(x + 0^+)$. Furthermore, we recall that since ϕ is nondecreasing, ϕ is differentiable almost everywhere. For $\phi \in \mathcal{D}^*$, we also introduce a generalized inverse $\phi^- \in \mathcal{D}^*$ defined by:

$$\phi^-(y) = \sup\{x \in [0, 1], \ \phi(x) < y\},$$

where we use the convention $\sup \emptyset = 0$; c.f. (Trouvé and Younes 2000, Section 5.2.3).

We are now able to formulate the main result of this section.

Theorem 3.1 *Let $c_1, c_2 \in PC^1(I, \mathbb{R}^d)$. Then the distance $dist^{\mathcal{S}}_{a,b}$ on the quotient space $\mathcal{S}(I, \mathbb{R}^d)$ is equivalently given by a supremum over the space $\mathcal{D}^*$, i.e.,*

$$dist^{\mathcal{S}}_{a,b}([c_1], [c_2]) = 2b\sqrt{\ell_{c_1} + \ell_{c_2} - 2 \sup_{\phi \in \mathcal{D}^*} \int_I \sqrt{\dot{\phi}(x)}\, f_{a,b}\,(x, \phi(x))\, dx},$$

where $f_{a,b} : I \times I \to \mathbb{R}$ is defined by

$$f_{a,b}(x, y) = \begin{cases} \sqrt{|\dot{c}_1(x)||\dot{c}_2(y)|}\cos\left(\frac{a}{2b}\cos^{-1}\left(\frac{\dot{c}_1(x)\cdot\dot{c}_2(y)}{|\dot{c}_1(x)||\dot{c}_2(y)|}\right)\right) & \text{if } \frac{a}{2b}\cos^{-1}\left(\frac{\dot{c}_1(x)\cdot\dot{c}_2(y)}{|\dot{c}_1(x)||\dot{c}_2(y)|}\right) \leq \frac{\pi}{2} \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

We have the following statements concerning the existence of optimal reparametrizations:

1. $\mathbf{a} < \mathbf{b}$: *for any $c_1, c_2 \in PC^1(I, \mathbb{R}^d)$ there exists a strictly increasing homeomorphism such that the infimum in (4) is attained. If the derivatives $\dot{c}_1, \dot{c}_2$ are Lipschitz continuous then $\gamma \in Diff_{C^1}(I)$, i.e., the optimal reparametrization is a C^1-diffeomorphism.*
2. $\mathbf{a} \geq \mathbf{b}$: *for $c_1, c_2 \in PC^1(I, \mathbb{R}^d)$ there exists a pair of generalized reparametrizations $\gamma_1, \gamma_2 \in \bar{\Gamma}$ such that the infimum in (5) is attained. On the other hand, there exists a pair of curves $c_1, c_2 \in AC(I, \mathbb{R}^d)$ such that the infimum in (5) is not attained in $\bar{\Gamma}$ and consequently neither in Γ or $\mathcal{D}^*$.*

Our proof of this result, which makes repeated use of results by Trouve and Younes (2000), is presented in "Appendix C."

Remark 3.1 (Open questions). This result suggests the following questions, which remain open for future research.

- *Counterexample for $a < b$*: The proof of the non-existence result for $a \geq b$ will be based on constructing curves c_1, c_2 such that $f_{a,b}(x, x) \leq 0$ for $x \in B$, where B is a closed but nowhere dense subset. For $a < b$, $f_{a,b}$ is positive, and thus, the same strategy fails.
- *Higher regularity:* One would hope that a higher regularity of the curves c_1 and c_2 would lead to a higher regularity of the obtained optimal reparametrization functions. To deduce this result from the theorem of Trouve and Younes (2000), one would need that a higher regularity of the curves c_i also leads to a higher regularity of the function $f_{a,b}$. This function is, however, at best Lipschitz continuous and only locally of a higher regularity. One can use the local regularity of $f_{a,b}$ and localize the arguments of Trouvé and Younes (2000) to show that the optimal reparametrization functions are locally of class C^{k-1} provided that the curves are of class C^k, but as of now we do not know how one could go a step further and obtain a global regularity result.

Remark 3.2 In the limit (and degenerate) case $a = 0$, one can further show that for any regular curves $c_1, c_2 \in PC^1(I, \mathbb{R}^d)$ the infimum in (4) is attained by the constant speed reparametrizations of c_1 and c_2, i.e.,

$$\text{dist}^{\mathcal{S}}_{a,b}([c_1], [c_2]) = \text{dist}_{a,b}(c_1 \circ \psi_{c_1}^{-1} \circ \psi_{c_2}, c_2)$$

where $\psi_c(u) = \int_0^u |\dot{c}(s)|\mathrm{d}s$. Indeed, let us first assume that c_1 and c_2 are both constant speed parametrized, i.e., $|\dot{c}_1|$ and $|\dot{c}_2|$ are constant on I and equal to the curve lengths ℓ_{c_1} and ℓ_{c_2}, respectively. Therefore, $f_{a,b}$ is just constant, and the optimization problem becomes:

$$\mathrm{dist}^{\mathcal{S}}_{a,b}([c_1],[c_2]) = 2b\sqrt{\ell_{c_1} + \ell_{c_2} - 2 f_{a,b}\,(0,0) \sup_{\phi\in\mathcal{D}^*} \int_I \sqrt{\dot{\phi}(x)}dx}.$$

We have by Cauchy–Schwarz

$$\left(\int_I \sqrt{\dot{\phi}(x)}dx\right)^2 \leq 1 \cdot \int_I \dot{\phi}(x)dx \leq 1 \cdot 1 = 1$$

and for $\phi = \mathrm{id}$, the previous inequality is an equality. The result follows and we also obtain that the (pseudo-)distance is given by:

$$\mathrm{dist}^{\mathcal{S}}_{a,b}([c_1],[c_2]) = \sqrt{\ell_{c_1} + \ell_{c_2} - 2\sqrt{\ell_{c_1}\ell_{c_2}}} = \left|\sqrt{\ell_{c_1}} - \sqrt{\ell_{c_2}}\right|$$

In the general case, where $c_1, c_2 \in PC^1(I,\mathbb{R}^d)$, one has that $c_1 \circ \psi_{c_1}^{-1}$ and $c_2 \circ \psi_{c_2}^{-1}$ have constant speed, and

$$\mathrm{dist}^{\mathcal{S}}_{a,b}([c_1],[c_2]) = \mathrm{dist}_{a,b}(c_1 \circ \psi_{c_1}^{-1} \circ \psi_{c_2}, c_2).$$

3.2 Existence of Optimal Reparametrizations on the Space of Closed Curves

We will now extend the previous result to the case of closed curves, i.e., when $D = S^1$. For closed parametrized curves, there does not exist anymore an explicit formula for the geodesic distance associated with general elastic $G^{a,b}$-metrics—to our knowledge, the only $G^{a,b}$ metric with explicit geodesics for closed curves is the $a = b$ case for plane curves (Younes et al. 2008) and space curves (Needham 2019b), both of which rely on specific constructions involving Hopf maps. Nevertheless, the formula we obtained for open curves in Theorem 2.1 still defines a reparametrization invariant distance function on the space of closed, parametrized curves: the next result follows by the same analysis applied in the open curve setting.

Corollary 3.1 *For $c_1, c_2 \in AC_0(S^1,\mathbb{R}^d)$, let $\mathrm{dist}_{a,b}(c_1,c_2)$ be given by the same formula as in (3) with integration over I replaced by integration over S^1. Then $\mathrm{dist}_{a,b}$ defines a metric on the space of closed, absolutely continuous curves $AC_0(S^1,\mathbb{R}^d)$.*

As previously, this allows us to construct a distance on the quotient space of unparametrized closed curves by defining

$$\mathrm{dist}^{\mathcal{S},\mathrm{cl}}_{a,b}([c_1],[c_2]) := \inf_{\gamma\in\Gamma_{\mathrm{cl}}} \mathrm{dist}_{a,b}(c_1, c_2 \circ \gamma),$$

where $\text{dist}_{a,b}$ is given by (3) and where Γ_{cl} denotes the group of absolutely continuous reparametrizations on the circle, i.e.,

$$\Gamma_{\text{cl}} = \{\gamma \in AC(S^1, S^1) : \gamma \text{ is bijective and } \gamma' > 0 \quad a.e.\}$$

We introduce the shift operator on S^1:

$$S_\tau : \begin{cases} S^1 \to S^1 \\ \theta \mapsto \theta + \tau \end{cases}$$

where $\tau \in S^1$. Then we can rewrite the group of absolutely continuous reparametrizations on the circle as

$$\Gamma_{\text{cl}} = \{S_\tau \circ \gamma : \gamma \in \Gamma \text{ and } \tau \in S^1\},$$

and we also need, as before, the closure of this group $\bar{\Gamma}_{\text{cl}} = \{S_\tau \circ \gamma : \gamma \in \bar{\Gamma} \text{ and } \tau \in S^1\}$. Then, the induced distance $\text{dist}_{a,b}^{\mathcal{S},\text{cl}}$ can be expressed as

$$\begin{aligned} \text{dist}_{a,b}^{\mathcal{S},\text{cl}}([c_1], [c_2]) &= \inf_{\gamma \in \Gamma_{\text{cl}}} \text{dist}_{a,b}(c_1, c_2 \circ \gamma) = \inf_{\gamma_1, \gamma_2 \in \Gamma_{\text{cl}}} \text{dist}_{a,b}(c_1 \circ \gamma_1, c_2 \circ \gamma_2) \\ &= \inf_{\gamma_1, \gamma_2 \in \bar{\Gamma}_{\text{cl}}} \text{dist}_{a,b}(c_1 \circ \gamma_1, c_2 \circ \gamma_2). \end{aligned} \tag{7}$$

We can now formulate the existence result of optimal reparametrizations for closed curves. Our proof, which will use the same method as Hartman et al. (2021) where the result was shown for the SRV metric, is postponed to "Appendix D."

Theorem 3.2 *Let $c_1, c_2 \in PC^1(S^1, \mathbb{R}^d)$. We have the following statements concerning the existence of optimal reparametrizations:*

1. $\mathbf{a} < \mathbf{b}$: *for any $c_1, c_2 \in PC^1(S^1, \mathbb{R}^d)$ there exists a strictly increasing homeomorphism such that the infimum in (7) is attained. If the derivatives $\dot{c}_1, \dot{c}_2$ are Lipschitz continuous, then $\gamma \in Diff_{C^1}(S^1)$, i.e., the optimal reparametrization is a C^1-diffeomorphism..*
2. $\mathbf{a} \geq \mathbf{b}$: *for $c_1, c_2 \in PC^1(S^1, \mathbb{R}^d)$ there exists a pair of generalized reparametrization $\gamma_1, \gamma_2 \in \bar{\Gamma}_{cl}$ such that the infimum in (7) is attained.*

4 Algorithms for the Computation of Quotient Distances and Geodesics

In the following, we will describe two different algorithms for the numerical computation of the optimal reparametrization on the space of open curves: an exact algorithm based on the work of Lahiri et al. (2015) and a faster dynamic programming-based approximation. Solving the registration problem on the space of closed curves simply requires an additional optimization over the starting point, i.e., one has to solve the

registration problem on the space of open curves for any choice of starting point. Consequently, the numerical solution on the space of closed curves is significantly more expensive.

4.1 Exact Algorithm for Piecewise Linear Curves

By the results of the previous section, we obtain the existence of optimal reparametrizations in the case of open, piecewise linear curves. Furthermore, using a result of Lahiri et al. (2015) we obtain an explicit algorithm for the optimal reparametrizations γ_1 and γ_2. We have the following result that follows directly from the corresponding analysis for the SRV-metric.

Theorem 4.1 *Let c_1, c_2 be two piecewise linear curves with values in $\mathbb{R}^d$. Then the pair of generalized reparametrizations $\gamma_1, \gamma_2 \in \bar{\Gamma}$ that attains the infimum in (5) consists of two piecewise linear maps.*

Proof First we note that piecewise linear curves are piecewise smooth and thus in particular piecewise C^1. This guarantees the existence of optimal reparametrization by the results of Theorem 3.1.

Next, we introduce a notion from Lahiri et al. (2015) and let $f : I \times I \to \mathbb{R}$. We call f *rectangular* if there exist partitions $0 = i_0 < i_1 < \cdots < i_m = 1$ and $0 = j_0 < i_1 < \cdots < j_n = 1$ such that f is constant on each rectangle of the form $[i_{r-1}, i_r] \times [j_{s-1}, j_s]$.

Using again Theorem 3.1, we have shown that finding optimal reparametrizations for the geodesic distance is equivalent to the optimization problem

$$\sup_{\gamma_1, \gamma_2 \in \bar{\Gamma}} \int_I \sqrt{\dot{\gamma}_1(u)\dot{\gamma}_2(u)}\, f_{a,b}\left(\gamma_1(u), \gamma_2(u)\right) du,$$

where $f_{a,b}$ is defined in (6). It is clear that the function $f_{a,b}$ is rectangular if the curves c_1 and c_2 are piecewise linear. From here, the proof given in Lahiri et al. (2015) goes through verbatim. □

Consequently, the exact algorithm from Lahiri et al. (2015) can be adapted to find optimal reparametrizations in our setting. This algorithm can find the optimal piecewise linear trajectory γ_1, γ_2 that maximizes $\int_{I\times I} W(\gamma_1(u), \gamma_2(u))\sqrt{\dot{\gamma}_1(u)\dot{\gamma}_2(u)}$ for general $W \in L^2(I \times I, \mathbb{R})$. For the $G^{a,b}$-metrics, we simply have $W = f_{a,b}$. We use the implementation from Martins Bruveris [30] that computes reparametrization for the usual $G^{1,1/2}$-metric, i.e., with $W(x, y) = q_1(x) \cdot q_2(y)$. Recall that closed curves require an additional optimization step to determine the starting point. For each vertex of the piecewise linear curve c_1, we simply compute the resulting distance with c_2 and choose the minimum.

4.2 Dynamic Programming Approach

The previous algorithm gives the exact optimal piecewise linear reparametrization, but is in practice very slow to compute. An alternative method is to use a dynamic

programming scheme to estimate an approximation of the optimal reparametrization, similar to the approach proposed in Trouvé and Younes (2000), Mio et al. (2007), Srivastava and Klassen (2016). We now describe this method.

We define a discretization $\mathcal{I} = \{x_0, ..., x_n\}$ of the interval $[0, 1]$ and restrict the search to piecewise linear functions on the grid $\mathcal{I} \times \mathcal{I}$. We denote by $PL_{\mathcal{I}}$ the set of piecewise linear increasing functions with vertices on $\mathcal{I} \times \mathcal{I}$. For $k < i, l < j$ and $\phi \in PL_{\mathcal{I}}$, we define the partial cost function

$$E(k, l, \phi) = \int_{x_k}^{x_i} \sqrt{\dot{\phi}(x)} f_{a,b}(x, \phi(x))dx.$$

Let $L_{k,l,i,j}(x) = x_l + \frac{x_j - x_l}{x_i - x_k}(x - x_k)$ be the line between (x_k, x_l) and (x_i, x_j), and with a slight abuse of notation, we shall write $E(k, l, i, j) = E(k, l, L_{k,l,i,j})$ as the energy of the segment. By additivity, if $\phi \in PL_{\mathcal{I}}$ is defined by the segments $L_{k_1,l_1,k_2,l_2}, L_{k_2,l_2,k_3,l_3}, ..., L_{k_{p-1},l_{p-1},k_p,l_p}$, the total energy of ϕ is

$$E(\phi) = \sum_{n=1}^{p-1} E(k_n, l_n, k_{n+1}, l_{n+1}).$$

We thus have to find the sequence of nodes (k_n, l_n) that minimizes the energy. We define the partial value function to reach node (i, j) by

$$H(i, j) = \min_{k<i, l<j} E(k, l, i, j) + H(k, l) \tag{8}$$

with $H(0, 0) = 0$; in other words, H is defined recursively. Due to the specific additive form of E, if for all $k < i, l < j$, $H(k, l)$ is the minimal energy between $(0, 0)$ and (k, l), then by definition $H(i, j)$ is the minimal energy between $(0, 0)$ and (i, j). Consequently, the global minimal energy that we aim to find is given by $H(1, 1)$.

The algorithm proceeds in two steps. First, we compute different values of H on the grid $\mathcal{I} \times \mathcal{I}$ using equation (8). Then, we determine the optimal path by backtracking from vertex $(1, 1)$: if (x_i, x_j) is a vertex of the optimizer, we compute the node that joins (x_i, x_j) by solving the problem:

$$(\hat{k}, \hat{l}) = \underset{k<i, l<j}{\operatorname{argmin}} E(k, l, i, j) + H(k, l).$$

To speed-up the computation of the value function and optimal reparametrization, a standard approach (Mio et al. 2007) is to restrict the search of the node that connects to (i, j) to a smaller set than $\{k, l : k < i, l < j\}$. In our case, we define

$$N_{i,j} = \{k, l : i - 6 \leq k < i, j - 6 \leq l < j\}$$

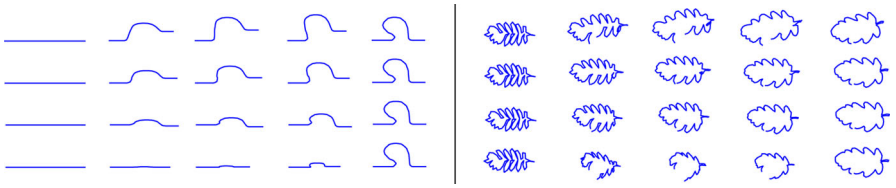

Fig. 1 Geodesics between open plane curves. Each panel shows several geodesics between a pair of curves, with each row corresponding to a different choice of parameters in the elastic metric. In the left panel, the source and target curves are simple synthetic curves, designed to clearly illustrate the dependence of the geodesic on the metric parameters. In the right panel, the source and target curves are real leaf shapes—although the source and target curves are actually closed, we compute each geodesic in the space of open curves. In both the left and right panels, the metric parameters vary by row and are given by $G^{a,\frac{1}{2}}$, with $a = 0.1$, 0.5, 1 and 5, as indicated

and the corresponding partial energy

$$H(i,j) = \min_{(k,l)\in N_{i,j}} E(k,l,i,j) + H(k,l).$$

This will lead to a restriction of the admissible slopes and in general to a less precise approximation of the reparametrization. Nevertheless, in all of our experiments, this restriction still yields good approximations of the true reparametrization functions, c.f. Fig. 2.

4.3 Examples

Figure 1 shows several geodesics between pairs of plane curves—a pair of simple synthetic curves and a pair of real leaf shapes, all sampled with 100 points in total. Geodesics are computed for a variety of elastic metrics $G^{a,b}$; we fix $b = \frac{1}{2}$ and compute geodesics for $a \in \{0.1, 0.5, 1, 5\}$. All geodesics in this figure were computed using our dynamic programming algorithm. These first examples clearly illustrate that the intermediate shapes along the geodesics strongly depend on the choice of metric parameters. Figure 2 compares the estimated reparametrization functions found via the dynamic programming algorithm to those found by the exact algorithm. We see here that the dynamic programming algorithm typically returns registrations which are close to the true optimal ones, while incurring a lower numerical burden. Indeed, for the synthetic example, the average computational times for exact registrations were 306 s, 269 s, 40 s, and 0.2s, respectively, for $a = 0.1$, 0.5, 1, 5; their counterparts computed using the dynamic programming algorithm were orders of magnitude smaller: 0.23s, 0.09s, 0.07s, and 0.06s. A similar trend held for the leaf shapes, where we got 299 s, 253 s, 34 s, and 0.19s for the exact algorithm and 0.22s, 0.03s, 0.025s, and 0.021s for dynamic programming. In the figure, we report the performance of the dynamic programming algorithm by giving its relative error $(d_{\text{dyn}} - d_{\text{ex}})/d_{\text{ex}}$, where d_{dyn} is geodesic distance (for given metric parameters) computed via dynamic programming and d_{ex} is the exact distance.

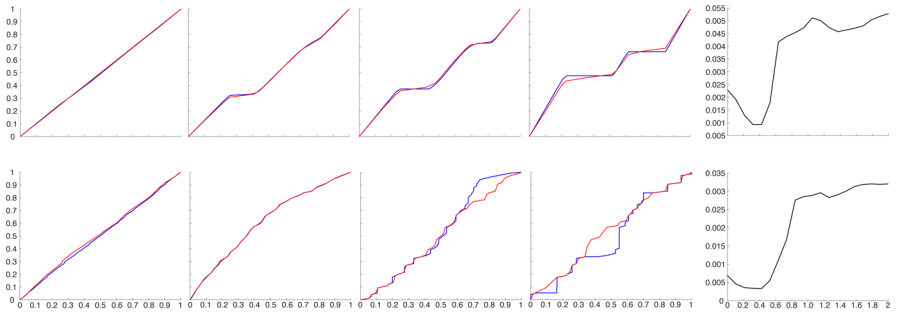

Fig. 2 Comparison of registrations via the exact and dynamic programming algorithms. The top row shows the registrations (i.e., optimal diffeomorphisms of [0, 1]) between the synthetic curves from Fig. 1 with respect to metric parameters $b = \frac{1}{2}$ and $a = 0.1, 0.5, 1, 5$, as indicated. In each figure, the exact registration is plotted in blue and the dynamic programming registration is plotted in red. The last graph in the row plots the relative error of dynamic programming versus the exact algorithm (see the text) for a finer range of parameters $a \in \{0, 0.1, 0.2, \ldots, 2\}, b = \frac{1}{2}$. The bottom row shows the same experiment for the registrations between the leaf shapes of Fig. 1. Observe that the relative error is generally on the order of a single digit percentage in each case (Color figure online)

Fig. 3 Geodesics between space curves. Each row shows a geodesic between the same pair of protein backbones, modeled as space curves, for a particular choice of elastic metric. The elastic metric parameters for the rows are $b = \frac{1}{2}, a = 0.1,\ 0.5,\ 1,\ 5$, respectively

Figure 3 applies our framework to compute geodesics between 3D curves (protein backbones), for a variety of elastic metric parameters. These geodesics were computed, once again, using the dynamic programming algorithm.

5 Metric Learning

The numerical experiments of the previous section illustrate both the influence and importance of the choice of metric when comparing and matching shapes. While certain heuristics may sometimes guide the selection of the parameters a and b of the elastic metric for a given dataset and application, it is often done via empirical trial and error approaches. Thus, in recent years, there has been a growing interest in developing methods for automatically estimating metrics on shape spaces.

Obviously, there is a priori no natural criterion to prefer one metric over another for the basic task of matching two shapes. However, when considering shape datasets and problems such as clustering or classification, it should be expected that different metrics will lead to different ways of quantifying differences across samples and consequently different properties from a statistical perspective. This suggests the idea of attempting to optimize (or learn) the choice of metric in order to improve the statistical power of shape analysis methods. The elastic framework of this paper appears quite amenable to such a task as learning the metric here reduces to learning a single parameter (the ratio of a and b). In this section, we consider the issue of metric learning for shape classification using two different models. Our primary focus is on demonstrating the feasibility and advantages of optimizing the choice of the metric in such a context; we leave an in depth study of the issue of developing efficient numerical algorithms for metric learning for future work.

The literature on metric learning is vast, and we only summarize some main ideas here—for more details, there are several surveys on the topic, e.g., Yang and Jin (2006), Bellet et al. (2013), Kulis et al. (2013), Kaya and Bilge (2019). Generally, metric learning is a supervised machine learning technique for choosing a metric from a parametric family which optimally separates data coming from different classes. Classically, the data consist of Euclidean feature vectors, and the metrics under consideration are *Mahalanobis distances* (Xing et al. 2002; Davis et al. 2007; Ying and Li 2012), which allows training via standard techniques from convex optimization theory. Closer to the topic of this paper, there has also been recent interest in metric learning on parametric families of Riemannian metrics on manifolds, such as spaces of SPD matrices (Vemulapalli and Jacobs 2015), spaces of histograms (Le and Cuturi 2015) and graphs (Heitz et al. 2021).

A common paradigm in metric learning is to represent data via *pairwise constraints*, where the training data consist of two sets $S = \{(x_i, x_j)\}$ and $D = \{(x_k, x_\ell)\}$ so that each pair $(x_i, x_j) \in S$ consists of *similar points* (coming from the same class) and each pair $(x_k, x_\ell) \in D$ consists of *dissimilar points* (coming from different classes). One then designs a loss function on the metric parameter space which encourages distances between similar points to be small and distances between dissimilar points to be large—the particulars of the loss function are application-dependent, and several choices are described in the survey papers cited above. An optimal metric with respect to a given loss can then be used for downstream distance-based analysis tasks, such as clustering, dimension reduction and k-nearest neighbors classification. This is the approach that we take in Sect. 5.1, where we us the pairwise constraints method to train metrics for various 2-dimensional shape datasets. On the other hand, if one has a particular classification task in mind, then it is sensible to learn the metric which optimizes performance on this task directly. We take this approach in Sect. 5.2 to learn a metric which optimally classifies 3-dimensional protein backbone curves.

5.1 Pairwise Constraints

Let us first consider the goal of estimating, in a supervised fashion, the metric that will best separate 2-dimensional shapes according to the pairwise constraints paradigm

described above. In other words, suppose that we have training data consisting of a collection of unparametrized curves $X = \{c_j\}_{j=1}^N$ together with known labels $y = \{y_j\}_{j=1}^N$ with each $y_j \in \{1, \ldots, K\}$ (that is, there are K distinct classes). We seek to determine the optimal parameters (a, b) for the elastic metric $G^{a,b}$ so that the geodesic distance $\text{dist}_{a,b}^{\mathcal{S}}$ optimally separates those classes. By a simple normalization argument, we may fix one of the two metric parameters, which we shall do in the following by setting $b = 1/2$ and optimize over $a \geq 0$.

The above task can be stated formally by introducing an adequate pairwise constraint loss function depending on the distances $\text{dist}_{a,b}^{\mathcal{S}}(c_i, c_j)$ between the curves of the training set. There have been various families of such loss functions appearing most notably in the machine learning literature. As a proof of concept, we choose here a simple loss which we can loosely think of as the ratio of the intra-class variance by the inter-class variance of the shape distances. Specifically, let $S \subset X \times X$ be the collection of ordered pairs (c_i, c_j) of curves with the same label $(y_i = y_j)$ and $D \subset X \times X$ the collection of ordered pairs with different labels $(y_i \neq y_j)$. We define our loss function by:

$$L(a) = \left(\frac{1}{|S|} \sum_{(c_i, c_j) \in S} \text{dist}_{a,1/2}^{\mathcal{S}}(c_i, c_j)^2 \right)^{\frac{1}{2}} \cdot \left(\frac{1}{|D|} \sum_{(c_k, c_\ell) \in D} \text{dist}_{a,1/2}^{\mathcal{S}}(c_k, c_\ell)^2 \right)^{-\frac{1}{2}}. \tag{9}$$

Minimizing $L(a)$ should be achieved when one strikes a balance between concentrating curves from the same class close to one another, while giving a large separation between different classes. As was recently observed in Ghojogh et al. (2019), Ghojogh et al. (2022), minimizing this loss function is essentially equivalent to solving the metric learning optimization problem described in the pioneering work of Xing et al. (2002).

For a given value of a, this loss function is calculated by first computing each of the $N(N-1)/2$ pairwise distances $\{\text{dist}_{a,1/2}^{\mathcal{S}}(c_i, c_j)\}$, which requires solving each of the corresponding registration problems. In our experiments, this is done using the dynamic programming scheme for the sake of computational efficiency. For the purpose of this work, we evaluate the loss function over a range of different values for a in order to determine the approximate value of the minimizer. Note that, while the expression of the loss when reparametrizations are fixed is relatively simple and could be optimized easily with respect to a, the difficulty is that the optimal reparametrizations leading to each pairwise distance value also depend on a. This makes the derivation of more sophisticated and efficient schemes for the minimization of (9) a non-trivial problem that we leave to future investigation, c.f. the discussion below.

Results. We tested our metric learning pipeline on two shape datasets. Results from the first dataset are reported in Fig. 4. Here, the data consisted of leaf shapes coming from four classes with 20 samples in each class. We chose 7 random examples from each class as training data and created pairwise data S and D, with S consisting of all pairs from the same class and D consisting of all pairs from different classes. We then minimized the loss function (9) via a grid search over the one-dimensional parameter space (a gradient descent algorithm was also implemented, which gave

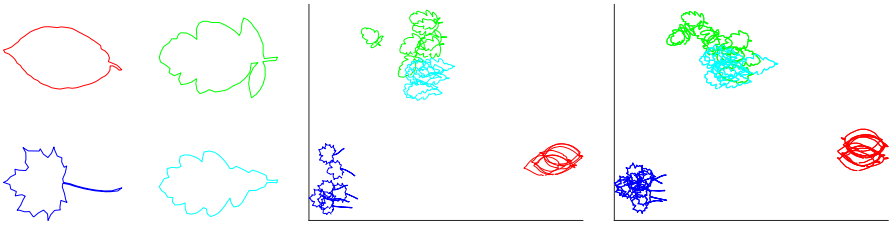

Fig. 4 Metric learning on the leaf dataset. The left panel shows representatives of each of the four classes in the leaf dataset. The middle and right panels show MDS (multidimensional scaling) plots of the pairwise distance matrices with respect to the learned metric for the training and testing data, respectively. The optimal metric parameter learned for the loss function (9) is $a = 0.71$. The Rand indices (see text for a full description) for the training and testing sets are 1 and 1, respectively

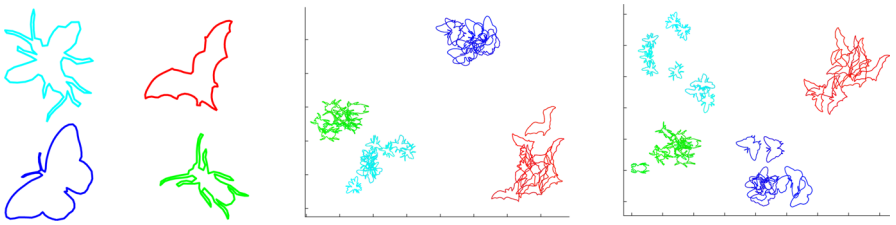

Fig. 5 Metric learning on the animal dataset. Similar to Fig. 5, from left to right the panels show representative shapes from each of the four classes, an MDS plot for the training data and an MDS plot for the testing data, both with respect to the pairwise distance matrices computed with the optimal metric parameter. For this experiment, the optimal parameter was $a = 0.52$ and the Rand indices for the training and testing sets were both equal to 1 (i.e., perfect clustering)

the same results). The efficacy of the learned metric was validated by testing the ability of the resulting geodesic distance to separate shapes from different classes. We measured this by computing the pairwise distance matrices for both the training and testing data, partitioning each dataset into four classes by applying complete linkage hierarchical clustering to the distance matrices and computing the Rand index of the inferred clusters against the ground truth classes. Similar results for the second dataset, consisting of shapes of four different species of animals, are reported in Fig. 5. Notably, the optimal metrics computed for the two datasets are quite different, indicating that the choice of optimal metric is data-dependent.

5.2 Metric Learning for Shape Classification

As an alternative to optimizing the metric parameters with respect to the loss (9), one may instead consider trying to maximize classification scores on the training set in a cross-validation fashion. A simple approach could be, for example, to evaluate leave-one-out nearest neighbor classification scores for varying values of a. A usually more robust way however is to rather rely on the estimation of conditional probabilities for different classes. We present one possible approach in the following description.

With the same notation as in the previous section, we introduce a leave-one-out scheme in which for each $j = 1, \ldots, N$, we denote by $p_j(\ell|\hat{X}_j)$ the probability of

c_j to be in the class $\ell \in \{1, \ldots, K\}$ knowing only the curves in $\hat{X}_j = \{c_i\}_{i \neq j}$ and their labels. In order to estimate the conditional probabilities p_j, we adapt a standard approach in many machine learning works, see e.g. Goodfellow et al. (2016), Chapter 6.2. Namely, we introduce the estimator $p^{(j)}(\ell|\hat{X}_j) = \sigma(z^{(j)})_\ell$ where the vector $z^{(j)} \in \mathbb{R}^k$ is defined by

$$(z^{(j)})_\ell = -\frac{1}{|\{i \neq j : y_i = l\}|} \sum_{i \neq j, y_i = l} \text{dist}^a(c_i, c_j)^2$$

and σ is the softmax function given by

$$\sigma(z)_\ell = \frac{e^{\beta z_\ell}}{\sum_{m=1}^K e^{\beta z_\ell}}$$

with $\beta > 0$ a fixed (or tunable) parameter. Then, for each $j = 1, \ldots, N$, $p^{(j)}(y_j|\hat{X}_j)$ gives an estimate of the probability for the correct class y_j of curve j. Thus, we seek to maximize the sum of the log-likelihoods over all instances of the leave-one-out scheme. In other words, we define the loss function to minimize to be:

$$L(a) = -\sum_{j=1}^N \log(p^{(j)}(y_j|\hat{X}_j)).$$

Note that $-\log(p^{(j)}(y_j|\hat{X}_j))$ can also be interpreted as the Kullback–Leibler divergence between the true probability distribution δ_{y_j} and the estimated one $p^{(j)}(\cdot|\hat{X}_j)$. As in the previous section, we can calculate the above loss for each value of the metric parameter a by first solving all pairwise matching problems to obtain the set of distances $\{\text{dist}^a(c_i, c_j)\}$.

Results. We used data from the 3D Shape Retrieval Contest 2010 (SHREC'10) (Mavridis et al. 2010) which contains a training dataset of 1000 protein structures from 100 classes, each class containing 10 proteins. Only the 3D curves representing the protein backbones were extracted and used in our experiments; two examples of such protein shapes are shown in Fig. 3. In addition, 50 more proteins from random classes formed the testing dataset that we used to evaluate the metric learning process. Our goal is to compare results obtained with the above approach to the methods from Mavridis et al. (2010). We calculated the classification loss for a sample of values of the metric parameter a between 0 and 2 (Fig. 6) and found the optimal value to be $a = 0.61$.

For evaluation, we use this optimal parameter to compute a matrix of the distances between the 50 proteins in the testing set and each of the proteins in the training set. The performance of the method is measured in the same two ways as in the original contest.

- Nearest neighbor: for each of the 50 proteins, we find the closest protein from the training dataset to predict the class of the testing protein. We calculate the overall

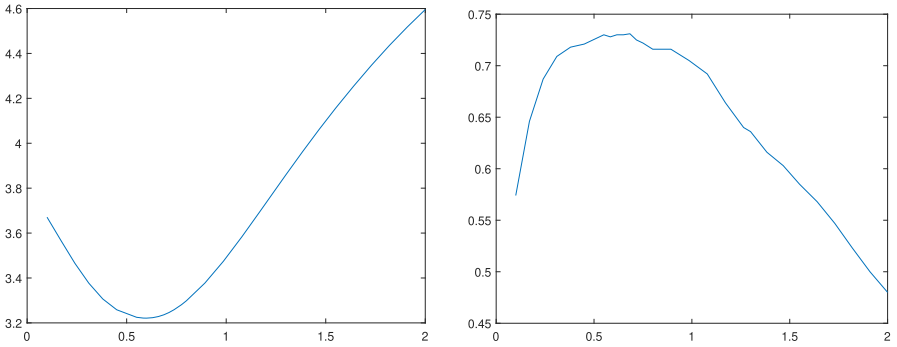

Fig. 6 Scores for different values of a, where the parameter a is plotted on the horizontal axis. In the left panel the vertical axis shows the loss function described in Sect. 5.2. For comparison, on the right, we also evaluate the cross-validation leave-one-out accuracy based on nearest neighbor classification

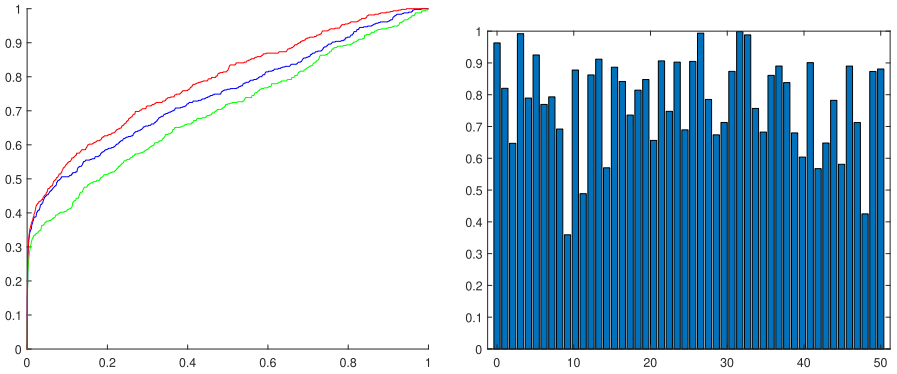

Fig. 7 Left: Aggregate ROC curves for different metric parameters: $a = 2$ (green), $a = 1$ (blue), and the optimal value $a = 0.61$ (red). Right: Bar chart showing the AUC for optimal metric parameter value $a = 0.61$ for each testing protein (Color figure online)

percentage of correct predictions. For the optimal value of the metric parameter determined by our method—$a = 0.61$—we obtain 82% of correct predictions, which beats all methods from Mavridis et al. (2010) (the best method in the paper reaches 80%).

- Receiver operating characteristic (ROC) curve: for each of the 50 proteins in the testing dataset, we create a ranked list of the proteins from the training dataset, from the closest protein to the most distant. This ranked list contains 10 proteins in the actual class of the testing protein (the true positives) and 990 proteins in a different class (the true negatives). The ranked list is traversed sequentially and we plot the cumulative rate of true positives against the cumulative rate of false positives. Figure 7 shows aggregate ROC curves for all of the 50 test protein curves, for different parameters of the metric. For the optimal value of a, we also compute the area under the curve (AUC) for the ROC curves of each testing protein and plot it in the right panel of the figure.

Acknowledgements M. Bauer has been supported by NSF-grants DMS-1912037 and DMS-1953244 and by the Austrian Science Fund grant P 35813-N, N. Charon by NSF grants DMS-1953267 and DMS-1945224, S. Kurtek by NSF grants CCF-1740761, CCF-1839252 and DMS-2015226, and NIH grant R37-CA214955, T. Needham by NSF grant DMS–2107808 and T. Pierron by NSF-grant DMS-1912037. We also thank Cy Maor for helpful discussions during the preparation of the manuscript.

Author Contributions Authors are listed in alphabetical order. All authors participated in the development of the mathematical and methodological part of the manuscript. Thomas Pierron was the primary contributor in the numerical experiments of this work.

Funding and Conflict of interest M. Bauer has been supported by NSF-grants DMS-1912037 and DMS-1953244 and by the Austrian Science Fund grant P 35813-N, N. Charon by NSF grants DMS-1953267 and DMS-1945224, S. Kurtek by NSF grants CCF-1740761, CCF-1839252 and DMS-2015226, and NIH grant R37-CA214955, T. Needham by NSF grant DMS-2107808 and T. Pierron by NSF-grant DMS-1912037. The authors have no further Conflict of interest to declare that are relevant to the content of this article.

Appendix A. An Interpretation from Linear Elasticity Theory

In the following section, we will formally derive the $G^{a,b}$-metric as a membrane limit (order-one energy density) of a linear elastic energy model. With the definitions and notations of Sect. 2.1, let us consider a portion of an open smooth curve which is parametrized on an interval we will denote I, i.e., $c : I \to \mathbb{R}^2$ is an immersion of class C^∞, and define the thin shell domain $\Omega \subset \mathbb{R}^2$ of "thickness" $\delta > 0$ around that curve. Specifically, we take Ω as being given by the parametrization $\omega : [-\delta/2, \delta/2] \times I \to \Omega$ defined by $\omega(t, u) = c(u) + tn(u)$ where $n(u)$ is the unit normal vector to the curve c at u. It is easy to see that for δ small enough, ω is a diffeomorphism that we can view as a foliation of Ω. Indeed, we note that $\omega(0, \cdot) = c(\cdot)$ and for any $t \in [-\delta/2, \delta/2]$, $\omega_t : u \mapsto \omega(t, u)$ defines a parametrized curve which corresponds to layer t of the foliation. Moreover, $\omega_t'(u) = c'(u) + tn'(u)$ and, as $n(u)$ is a unit vector orthogonal to $c'(u)$, we get that $\omega_t'(u)$ is parallel to $c'(u)$ which implies that $n(u)$ is also the unit normal vector to ω_t at u. Thus, for any $x = \omega(t, u) \in \Omega$, we can define the orthonormal vector frame $F(x) = (\tau(x), n(x))$ by:

$$\tau(x) = \frac{\omega_t'(u)}{|\omega_t'(u)|} = \frac{\partial_u \omega(t, u)}{|\partial_u \omega(t, u)|}, \quad n(x) = n(u)$$

See the illustration given in Fig. 8.

Now, we model Ω as a linear elastic material which undergoes an infinitesimal deformation given by a smooth vector field $v : \Omega \to \mathbb{R}^2$. We shall further assume that this deformation field is uniform along the transversal direction, in other words that it takes the following form: for any $x = \omega(t, u)$, $v(x) = h(u)$ where h is a vector field defined along the curve c as in the previous section, c.f. again Fig. 8 for visualization. Note that this is a natural assumption in the small thickness laminar model that we are interested in here. Then, following the approach of classical linear elasticity (Gurtin 1981; Ciarlet 1988), one introduces the (2×2) symmetric tensor field defined for all $x \in \Omega$ as $\varepsilon(x) = \frac{dv(x) + dv(x)^T}{2}$. This is known as the strain tensor associated with the deformation field v and expressed in the canonical basis. Given the specific laminar structure of the domain here, it will be more convenient to instead consider

the strain tensor relative to the above orthonormal frame $F(x)$, which is specifically $S(x) = F(x)^T \varepsilon(x) F(x)$. The linear elastic energy associated with the deformation field is obtained from Hooke's law and takes the general form:

$$E(v) = \int_\Omega U(x, S(x))\mathrm{d}x$$

where $U(x, \cdot)$ is a quadratic form on the space of symmetric (2×2) matrices which is usually referred to as the elastic or stiffness tensor. In the present context, we will restrict the class of such elastic tensors by making a few additional assumptions. First, we will consider the elastic properties of the material to be uniform in the sense that $U(x, \cdot)$ does not depend on $x \in \Omega$. Then, viewing any symmetric matrix $S = (s_{ij})_{i,j=1,2}$ as the (3×1) vector $(s_{11}, s_{22}, s_{12})^T$, the quadratic form U can be identified with a single symmetric positive definite (3×3) matrix which we write as:

$$U = \begin{pmatrix} u_{1,1} & u_{1,2} & u_{1,12} \\ u_{1,2} & u_{2,2} & u_{2,12} \\ u_{1,12} & u_{2,12} & u_{12,12} \end{pmatrix}.$$

The coefficients in U assign weights to different terms in the elastic energy in the following way. Both coefficients $u_{1,1}$ and $u_{2,2}$ correspond to spring-like stiffness coefficients in the tangential and normal directions, respectively. Coefficient $u_{1,2}$, on the other hand, weighs the relative compression/stretching between tangential and normal directions. The coefficient $u_{12,12}$ can be associated with bending energy that results from a change of angle between the two directions. We will make some further symmetry assumptions on the material, namely that it is orthotropic with respect to the two directions $\tau(x)$ and $n(x)$ at each point x. This leads to the conditions $u_{1,12} = u_{2,12} = 0$. Note that the orthotropy assumption is relatively common in many materials (with the exception of certain crystals) and include in particular fully isotropic materials, c.f. Remark A.1.

Going back more specifically to the deformation of the foliated domain Ω, due to the particular form of the vector field v, we can see that for any $x = \omega(t, u) \in \Omega$,

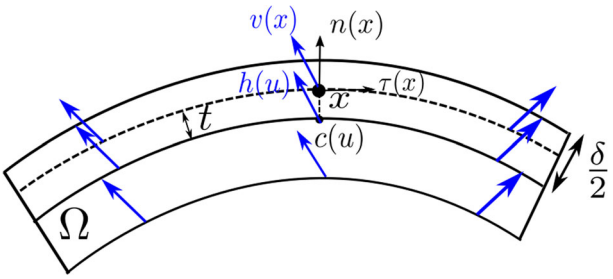

Fig. 8 Illustration of the thin shell elastic domain model

$dv(x) \cdot n(x) = 0$. This implies that the strain tensor is of the form:

$$S(x) = \begin{pmatrix} \tau(x)^T dv(x) \cdot \tau(x) & \frac{1}{2}(dv(x) \cdot \tau(x))^T n(x) \\ \frac{1}{2}(dv(x) \cdot \tau(x))^T n(x) & 0 \end{pmatrix}$$

and so, identified as a (3×1) vector, we have $S(x) = (s_{11}(x), 0, s_{12}(x))^T$. Under the previous assumptions on the elastic tensor, we find the following elastic energy:

$$\begin{aligned} E(v) &= \int_\Omega (u_{1,1} s_{11}(x)^2 + u_{12,12} s_{12}(x)^2) \mathrm{d}x \\ &= \int_{-\frac{\delta}{2}}^{+\frac{\delta}{2}} \int_I (u_{1,1} s_{11}(\omega(t,u))^2 + u_{12,12} s_{12}(\omega(t,u))^2) \, |J_\omega(t,u)| \, \mathrm{d}u \mathrm{d}t. \end{aligned}$$

where J_ω denotes the Jacobian determinant of ω. We have seen that $\partial_t \omega(t,u) = n(u)$ and $\partial_u \omega(t,u) = c'(u) + t n'(u)$ with $c'(u)$ and $n'(u)$ being parallel vectors both orthogonal to $n(u)$. Thus, for all (t,u), $|J_\omega(t,u)| = |c'(u) + t n'(u)|$.

Now, using the continuity with respect to t of the inside integral and the mean value theorem, we obtain that:

$$\lim_{\delta \to 0} \frac{1}{\delta} E(v) = \int_I (u_{1,1} s_{11}(\omega(0,u))^2 + u_{12,12} s_{12}(\omega(0,u))^2) |c'(u)| du.$$

In addition, from $v(\omega(0,u)) = h(u)$, we get by differentiating that $dv(\omega(0,u)) \cdot \tau(\omega(0,u)) = h'(u)/|c'(u)|$. We can then rewrite the above expressions of s_{11} and s_{12} as:

$$\begin{aligned} s_{11}(\omega(0,u)) &= \tau(\omega(0,u))^T \left(\frac{h'(u)}{|c'(u)|} \right) = \left(\frac{c'(u)}{|c'(u)|} \right)^T \frac{h'(u)}{|c'(u)|} \\ s_{12}(\omega(0,u)) &= \frac{1}{2} \frac{n(\omega(0,u))^T h'(u)}{|c'(u)|} \end{aligned}$$

which finally leads to:

$$\lim_{\delta \to 0} \frac{1}{\delta} E(v) = \int_I \left[u_{1,1} (D_s h^\top)^2 + \frac{u_{12,12}}{4} (D_s h^\perp)^2 \right] ds$$

In summary, we have shown that the expression of the Riemannian metric $G^{a,b}(h,h)$ of (1) is obtained as the limit of the elastic energy of an orthotropic laminar thin shell domain as the thickness $\delta \to 0$, in which $a^2 = u_{11}$ and $b^2 = u_{12,12}/4$ can be interpreted as stretching and bending energy coefficients, respectively.

Remark A.1 In the special case of an isotropic elastic domain Ω (still with respect to the frame vectors $\tau(x)$ and $n(x)$), the elastic tensor takes the particular form:

$$U = \begin{pmatrix} 2\mu + \lambda & \lambda & 0 \\ \lambda & 2\mu + \lambda & 0 \\ 0 & 0 & 2\mu \end{pmatrix}.$$

in which $\lambda, \mu \geq 0$ are the so called Lamé coefficients of the material. This leads to $G^{a,b}$ metric for which $a^2 = 2\mu + \lambda$ and $b^2 = \mu/2$ i.e.

$$\frac{b}{a} = \frac{1}{2}\sqrt{\frac{2\mu}{2\mu+\lambda}} \leq \frac{1}{2}.$$

It is thus interesting to note that this stronger isotropy assumption on the elastic domain imposes the constraint that $b \leq a/2$, the limiting case $b = a/2$ corresponding precisely to the square root velocity (SRV) metric of Srivastava et al. (2010).

Remark A.2 The previous derivations can be extended to curves in higher dimensions relatively easily. In the case of a 3D curve for example, one can introduce a tubular neighborhood of small radius δ around that curve that is again deformed by a vector field uniform in the transverse direction. Then, assuming a material with transversely isotropic elastic properties, one can show that, as δ goes to 0, the resulting elastic energy is again given by (1).

Appendix B. Proof of Theorem 2.1 and Corollary 2.1

To derive the formula for geodesic distance (3), we will first prove a lemma on distances in the finite-dimensional space $(\mathbb{R}^d \setminus \{0\}, g^\lambda)$. Clearly $\mathbb{R}^d \setminus \{0\}$ is incomplete with respect to g^λ for any $\lambda > 0$. We can complete it as a metric space by reinserting the origin. Let $\mathbb{R}^d_\lambda$ denote the metric space that is the completion with respect to the g^λ metric. Note that, as a point set, $\mathbb{R}^d_\lambda$ is just $\mathbb{R}^d$, but $\mathbb{R}^d_\lambda$ is not a Riemannian manifold when $\lambda \neq 1$, as the Riemannian metric g^λ cannot be smoothly extended to the origin in this case. We obtain the following explicit formula for the distance d_λ on this completion:

Lemma B.1 *For $q_1, q_2 \in \mathbb{R}^d_\lambda$. we let*

$$d_\lambda(q_1, q_2) = \sqrt{|q_1|^2 + |q_2|^2 - 2|q_1||q_2|\cos(\lambda\theta)},$$

where

$$\theta = \begin{cases} min\left(\cos^{-1}(q_1 \cdot q_2/|q_1||q_2|), \frac{\pi}{\lambda}\right) & \text{if } q_1 \text{ and } q_2 \text{ are both nonzero} \\ \frac{\pi}{\lambda} & \text{if } q_1 \text{ or } q_2 \text{ is zero} \end{cases}$$

Then $(\mathbb{R}^d_\lambda, d_\lambda)$ is the metric completion of the Riemannian manifold $(\mathbb{R}^d \setminus \{0\}, g^\lambda)$.

Proof First we consider the case $d = 2$. Without loss of generality, assume that $q_1 = (h, 0)$ and $q_2 = (k\cos\theta, k\sin\theta)$, where $h, k > 0$ and $0 \leq \theta \leq \pi$. (This can easily be arranged, since reflections in the x-axis and rotations about the origin are isometries of $\mathbb{R}^2_\lambda$.) Define the sector $V_\theta \subset \mathbb{R}^n_\lambda$ by $V_\theta = \{(r\cos\alpha, r\sin\alpha) : 0 \leq \alpha \leq \theta \text{ and } r \geq 0\}$. Define a function $F : V_\theta \to \mathbb{R}^2$ by

$$F(r\cos\alpha, r\sin\alpha) = (r\cos\lambda\alpha, r\sin\lambda\alpha).$$

We now consider the case $\lambda\theta \leq \pi$, and we show that F is an isometry from V_θ with metric given by the completion of the Riemannian manifold $(\mathring{V}_\theta, g^\lambda)$ into $\mathbb{R}^2$ with the Euclidean metric. In the following computations, we fix a basepoint $q = (r\cos\alpha, r\sin\alpha)$ in the interior $\mathring{V}_\theta$, and $v, w \in T_q\mathring{V}_\theta$. We can easily see that $F \circ O_{-\alpha} = O_{-\lambda\alpha} \circ F$ where $O_\phi : \mathbb{R}^2 \to \mathbb{R}^2$ is rotation by angle ϕ. Therefore, as rotations are isometries for the Euclidian metric g^1, we have

$$\begin{aligned} g^1_{F(q)}(d_qF(v), d_qF(w)) &= g^1_{O_{-\lambda\alpha}\circ F(q)}(d_{F(q)}O_{-\lambda\alpha} \circ d_qF(v), d_{F(q)}O_{-\lambda\alpha} \circ d_qF(w)) \\ &= g^1_{F\circ O_{-\alpha}(q)}(d_{O_{-\alpha}(q)}F \circ d_qO_{-\alpha}(v), d_{O_{-\alpha}(q)}F \circ d_qO_{-\alpha}(w)) \\ &= g^1_{F(r,0)}(d_{(r,0)}F(\tilde{v}), d_{(r,0)}F(\tilde{w})) \end{aligned}$$

with $\tilde{v} := d_qO_{-\alpha}(v) = O_{-\alpha}(v)$ and $\tilde{w} := d_qO_{-\alpha}(w) = O_{-\alpha}(w)$. We can easily compute that $d_{(r,0)}F(\tilde{v}) = (\tilde{v}^\top, \lambda\tilde{v}^\perp)$, and putting this together with the calculation above yields

$$\begin{aligned} g^1_{F(q)}(d_qF(v), d_qF(w)) &= g^1_{F(r,0)}\left((\tilde{v}^\top, \lambda\tilde{v}^\perp), (\tilde{w}^\top, \lambda\tilde{w}^\perp)\right) \\ &= \tilde{v}^\top\tilde{w}^\top + \lambda^2\tilde{v}^\perp\tilde{w}^\perp \\ &= g^\lambda_{(r,0)}(\tilde{v}, \tilde{w}) \end{aligned}$$

As g^λ is invariant under rotations, we have

$$g^\lambda_{(r,0)}(\tilde{v}, \tilde{w}) = g^\lambda_{O_\alpha(r,0)}(O_\alpha(v), O_\alpha(w)) = g^\lambda_q(v, w)$$

and it follows that F is a Riemannian isometric embedding from $(\mathring{V}_\theta, g^\lambda)$ into $(\mathbb{R}^2, g^1)$. It extends to a metric isometric embedding from V_θ into $\mathbb{R}^2$ by completion.

F is also injective and $F(V_\theta)$ is a convex subset of $\mathbb{R}^2$, so the geodesic (i.e., straight line) in $F(V_\theta)$ joining $F(q_1)$ to $F(q_2)$ remains in $F(V_\theta)$. Thus, if we apply F^{-1} to this straight line, we obtain a geodesic in $\mathbb{R}^2_\lambda$ joining u to v. Since F is an isometry, the length of the geodesic in $\mathbb{R}^2_\lambda$ is the same as the length of the straight line, and is given by the desired formula, as a simple application of the law of cosines in $\mathbb{R}^2$ shows.

In case that $\lambda\theta \geq \pi$, let $\lambda' = \frac{\lambda\theta}{\pi}$. We have $\lambda' \geq 1$. We consider once again $F_{\frac{\pi}{\theta}} : V_\theta \to V_\pi$:

$$F_{\frac{\pi}{\theta}}(r\cos\alpha, r\sin\alpha) = \left(r\cos\frac{\pi}{\theta}\alpha, r\sin\frac{\pi}{\theta}\alpha\right)$$

Then $F_{\frac{\pi}{\theta}}$ is an isometry from (V_θ, g^λ) to $(V_\pi, g^{\frac{\lambda\theta}{\pi}})$, and therefore:

$$\mathrm{dist}_{g^\lambda}((h,0), (r\cos\theta, r\sin\theta)) = \mathrm{dist}_{g^{\lambda'}}((h,0), (-r,0))$$

Because $\lambda' \geq 1$, for all $\omega \in \mathbb{R}^2$, $g^{\lambda'}(\omega,\omega) \geq |\omega'|^2_{\mathrm{euc}}$, and therefore, $\mathrm{dist}_{g^{\lambda'}} \geq \mathrm{dist}_{\mathrm{euc}}$. But taking the straight line between $(h,0)$ and $(-r,0)$, we then have $\mathrm{dist}_{g^{\lambda'}}((h,0),(-r,0)) \leq h + r$, and therefore $\mathrm{dist}_{g^{\lambda'}}((h,0),(-r,0)) = h + r$. This yields again the desired formula.

Having proved the theorem for $d = 2$, it follows for general d, since any pair of elements u, v is contained in a totally geodesic copy of $\mathbb{R}^2_\lambda \subset \mathbb{R}^d_\lambda$. □

Proof of Theorem 2.1 and Corollary 2.1 The statement that R is a diffeomorphism is clear from the definition of the involved spaces; see also Bruveris (2016), Lahiri et al. (2015). It remains to show that R is a Riemannian isometry. To this end, we calculate the derivative of R at $c \in \mathcal{I}_0([0,1], \mathbb{R}^d)$ in the direction $h \in T_c I_0([0,1], \mathbb{R}^d)$ as

$$d_c R(h) = \frac{1}{\sqrt{|c'|}} \left(h' - \frac{1}{2} \left(h' \cdot \frac{c'}{|c'|} \right) \frac{c'}{|c'|} \right).$$

The component of $d_c R(h)$ tangential to $R(c)$ is

$$\begin{aligned} \big(d_c R(h)\big)^\top &= \left(d_c R(h) \cdot \frac{c'}{|c'|} \right) \frac{c'}{|c'|} = \frac{1}{2\sqrt{|c'|}} \left(h' \cdot \frac{c'}{|c'|} \right) \frac{c'}{|c'|} \\ &= \frac{\sqrt{|c'|}}{2} \left(\frac{1}{|c'|} h' \cdot \frac{c'}{|c'|} \right) \frac{c'}{|c'|} = \frac{\sqrt{|c'|}}{2} D_s h^\top. \end{aligned}$$

Similarly, the orthogonal component is given by $\big(d_c R(h)\big)^\perp = \sqrt{|c'|} D_s h^\perp$. Therefore, for $h, k \in T_c I_0([0,1], \mathbb{R}^d)$, we have

$$\begin{aligned} 4b^2 G^{L^2_\lambda}_{R(c)}(d_c R(h), d_c R(k)) &= 4b^2 \int_0^1 \lambda^2 |c'| \big(D_s h^\perp \cdot D_s k^\perp \big) + \frac{|c'|}{4} \big(D_s h^\top \cdot D_s k^\top \big)\, du \\ &= \int_0^1 a^2 \big(D_s h^\perp \cdot D_s k^\perp \big) + b^2 \big(D_s h^\top \cdot D_s k^\top \big)\, ds = G^{a,b}_c(h,k), \end{aligned}$$

where, in the last line, we use that $\lambda = \frac{a}{2b}$ and $ds = |c'(u)| du$. The proves that R is an isometry.

It remains to derive the geodesic distance formula. To do so, we recall a general fact about geodesics in path spaces. Let (M, g) be a (finite-dimensional) Riemannian manifold and consider the space

$$\mathcal{M} := C^\infty([0,1], M).$$

By Bruveris (2018), a path in $\mathcal{M}$ given by $t \mapsto c_t$ is a length minimizing geodesic with respect to the L^2-Riemannian metric (defined by

$$G^{L^2}_c(h,k) = \int_0^1 g_{c(u)}(h(u), k(u)) du$$

for $h, k \in T_c \mathcal{M}$ parametrized smooth vector fields along c) if and only if for (almost all) fixed u_0, the curve given by $t \mapsto c_t(u_0)$ is a length-minimizing geodesic in M. Consequently, the geodesic distance between $c_0, c_1 \in \mathcal{M}$ is given by

$$\sqrt{\int_0^1 \operatorname{dist}^M(c_0(u), c_1(u))^2 du},$$

where dist^M denotes geodesic distance in the finite-dimensional manifold M.

We can now apply the above result with $(M, g) = (\mathbb{R}^d \backslash \{0\}, g^\lambda)$. Let $c_0, c_1 \in \mathcal{I}_0([0,1], \mathbb{R}^d)$ such that the geodesic in $(\mathbb{R}^d, g^\lambda)$ between $R(c_0)(u)$ and $R(c_1)(u)$ does not pass through the origin for any $u \in [0,1]$. Then the formula for the geodesic distance follows directly by the formula from Lemma B.1 and the above considerations – note that in this case the minimum in the definition of θ is always given by the arccos term. This proves the local formula for the geodesic distance. To obtain the global formula one needs to smoothly perturb any path that passes through the origin in such a way that the perturbed path avoids the origin. It is easy to see that this is possible if $d \geq 3$, which shows the global formula for the geodesic distance. For $d = 2$ and $\lambda = 1$ a counterexample, i.e., two curves where the minimizing path cannot be perturbed to avoid the origin, has been constructed in Bruveris (2016). A similar argument works for general λ, and thus, the formula for the geodesic distance is only valid locally in this case. □

To prove the statements on the metric completion, we will first study these completions in the space of R-transforms, i.e., on $C^\infty(I, \mathbb{R}^d \setminus \{0\})$:

Lemma B.2 *For $q_1, q_2 \in L^2([0,1], \mathbb{R}^d)$, we let*

$$d_{G^{L^2_\lambda}}(q_1, q_2) = \sqrt{\int_0^1 |q_1(u)|^2 + |q_2(u)|^2 - 2|q_1(u)||q_2(u)| \cos(\lambda\theta(u)) du},$$

where

$$\theta(u) = \begin{cases} min\left(\cos^{-1}(q_1(u) \cdot q_2(u)/|q_1(u)||q_2(u)|), \frac{\pi}{\lambda}\right) & \text{if } q_1(u), q_2(u) \neq 0 \\ \frac{\pi}{\lambda} & \text{otherwise.} \end{cases}$$

We have the following two statements regarding the completion of the space of smooth functions: The space $\left(L^2([0,1], \mathbb{R}^d_\lambda), d_{G^{L^2_\lambda}}\right)$ is the metric completion of the geodesic completion $C^\infty([0,1], \mathbb{R}^d_\lambda)$. If $d \geq 3$, then $\left(L^2([0,1], \mathbb{R}^d_\lambda), d_{G^{L^2_\lambda}}\right)$ is also the metric completion of $\left(C^\infty([0,1], \mathbb{R}^d \backslash \{0\}), G^{L^2_\lambda}\right)$.

Proof This follows directly from the definition of the geodesic distance on $C^\infty([0,1], \mathbb{R}^d \setminus \{0\})$, the proof of Theorem 2.1, and Lemma B.1. □

Now corollary 2.1 follows from the results above and the formula for R.

Appendix C. Proof of Theorem 3.1.

The main ingredient for the existence proof is the following result by Trouve and Younes, concerning the existence of minimizers for a wide class of optimization problems:

Theorem C.1 (Theorem 3.1 and Prop. 5.1 in Trouvé and Younes (2000)). *Let* $f : [0,1] \times [0,1] \to \mathbb{R}_{\geq 0}$ *be a bounded measurable function that satisfies the following condition*

(H1) There exists a finite family of closed segments $([a_j, b_j])_{j\in J}$ *such that each of them is horizontal or vertical and* f *is continuous on* $[0,1]^2 \setminus \bigcup_{j\in J}[a_j, b_j]$

Then there exists an non-decreasing BV-function $\phi \in \mathcal{D}^*$ *that maximizes the functional* $\phi \mapsto \int_D \sqrt{\dot{\phi}(x)} f(x, \phi(x))dx$. *Let* f_s *be defined by*

$$f_s(x_0, y_0) = \lim_{\delta\to 0}\left(\inf\left\{ f\,(x,y) \,|\, (x,y) \in [0,1]^2 \setminus \bigcup_{j\in J}[a_j, b_j], |\,(x,y) - (x_0, y_0)\,| < \delta \right\}\right)$$

Assume that f_s *satisfies in addition the condition:*

(H2) There does not exist any nonempty open vertical or horizontal segment $]a, b[$ *such that* f_s *vanishes on* $]a, b[$.

Then the optimizer $\phi^* \in \mathcal{D}^*$ *is a strictly increasing homeomorphism.*

For the statements regarding the higher regularity case, we will in addition need the following technical result:

Theorem C.2 (Theorem 3.3 in Trouvé and Younes (2000)). *Let* f *be a nonnegative measurable function on* $[0,1]^2$, *and assume that* $U_f : \mathcal{D}^* \to \mathbb{R}, \phi \mapsto \int_D \sqrt{\dot{\phi}(x)} f(x, \phi(x))dx$ *reaches its maximal value at a strictly increasing continuous function* $\phi^* \in \mathcal{D}^*$. *Then for any* $x_0 \in [0,1]$, *if* $f(x_0, \phi(x_0)) > 0$ *and if* f *is locally Hölder continuous, then* ϕ^* *is differentiable at* x_0, *with strictly positive derivative, and* $\dot{\phi}^*$ *is continuous in a neighborhood of* x_0.

Proof of Theorem 3.1 We start by proving the formula for the geodesic distance. Using the explicit formula for parametrized curves that are obtained in Theorem 2.1, we can write the geodesic distance as

$$\mathrm{dist}^{\mathcal{S}}_{a,b}([c_1],[c_2]) = 2b\sqrt{\ell_{c_1} + \ell_{c_2} - 2\sup_{\gamma_1,\gamma_2\in\bar{\Gamma}}\int_I \sqrt{\dot{\gamma}_1(u)\dot{\gamma}_2(u)}\,\tilde{f}_{a,b}\left(\gamma_1(u), \gamma_2(u)\right)du},$$

where $\tilde{f}_{a,b} : I \times I \to \mathbb{R}$ is defined by

$$\tilde{f}_{a,b}(x,y) = \begin{cases} \sqrt{|\dot{c}_1(x)||\dot{c}_2(y)|}\cos\left(\frac{a}{2b}\cos^{-1}\left(\frac{\dot{c}_1(x)\cdot\dot{c}_2(y)}{|\dot{c}_1(x)||\dot{c}_2(y)|}\right)\right) & \text{if } \frac{a}{2b}\cos^{-1}\left(\frac{\dot{c}_1(x)\cdot\dot{c}_2(y)}{|\dot{c}_1(x)||\dot{c}_2(y)|}\right) \leq \pi \\ -1 & \text{otherwise.} \end{cases}$$

This formulation is, however, not convenient for us, as the function $\tilde{f}_{a,b}$ is not non-negative, and thus, one cannot directly apply the results of Trouvé and Younes (2000). Thus, we will first show that the above optimization problem does not change when we substitute $\tilde{f}_{a,b}$ by the non-negative function $f_{a,b}$, as defined in (6). Since $f_{a,b} =$

$\max(\tilde{f}_{a,b}, 0)$, we have

$$\sup_{\gamma_1,\gamma_2\in\bar{\Gamma}} \int_I \sqrt{\dot{\gamma}_1(u)\dot{\gamma}_2(u)}\, \tilde{f}_{a,b}\left(\gamma_1(u), \gamma_2(u)\right)) du \leq \sup_{\gamma_1,\gamma_2\in\bar{\Gamma}} \int_I \sqrt{\dot{\gamma}_1(u)\dot{\gamma}_2(u)}\, f_{a,b}\left(\gamma_1(u), \gamma_2(u)\right)) du$$

Let $\gamma_1, \gamma_2 \in \bar{\Gamma}$, and let $A = \{u \in I,\ \tilde{f}_{a,b}\left(\gamma_1(u), \gamma_2(u)\right)) < 0\}$ the open set of negative parts of $\tilde{f}_{a,b}$. We can write A as an at most countable disjoint union of open intervals $A = \bigcup_n I_n$ with $I_n =]u_n^-, u_n^+[$. Now let us construct reparametrizations $\tilde{\gamma}_1, \tilde{\gamma}_2$ with derivatives equal to zero on A. We set $\tilde{\gamma}_1(u) = \gamma_1(u)$, $\tilde{\gamma}_2(u) = \gamma_2(u)$ for $u \in I\backslash A$ and:

$$\tilde{\gamma}_1(u) = \begin{cases} \gamma_1(2u - u_n^-) & \text{for } u \in]u_n^-, 1/2(u_n^- + u_n^+)[\\ \gamma_1(u_n^+) & \text{for } u \in]1/2(u_n^- + u_n^+), u_n^+[\end{cases}$$

$$\tilde{\gamma}_2(u) = \begin{cases} \gamma_2(u_n^-) & \text{for } u \in]u_n^-, 1/2(u_n^- + u_n^+)[\\ \gamma_2(2u - u_n^+) & \text{for } u \in]1/2(u_n^- + u_n^+), u_n^+[\end{cases}$$

We clearly have $\tilde{\gamma}_1, \tilde{\gamma}_2 \in \bar{\Gamma}$ and for $u \in I_n$, $\sqrt{\dot{\gamma}_1(u)\dot{\gamma}_2(u)} = 0$, so

$$\begin{aligned} \int_I \sqrt{\dot{\tilde{\gamma}}_1(u)\dot{\tilde{\gamma}}_2(u)}\, \tilde{f}_{a,b}\left(\tilde{\gamma}_1(u), \tilde{\gamma}_2(u)\right)) du &= \int_{I\backslash A} \sqrt{\dot{\gamma}_1(u)\dot{\gamma}_2(u)}\, \tilde{f}_{a,b}\left(\gamma_1(u), \gamma_2(u)\right)) \\ &= \int_I \sqrt{\dot{\gamma}_1(u)\dot{\gamma}_2(u)}\, f_{a,b}\left(\gamma_1(u), \gamma_2(u)\right)), \end{aligned}$$

which proves the equivalence of the two optimization problems.

Next we will prove the equivalent definition of the distance, where the infimum is taken over one BV function in the space $\mathcal{D}^*$. Thus, we consider the two functionals

$$\begin{aligned} \mathcal{D}^* &\to \mathbb{R} \\ \phi &\mapsto \textstyle\int_I \sqrt{\dot{\phi}(x)}\, f_{a,b}(x, \phi(x)) dx \end{aligned} \tag{C.1}$$

and

$$\begin{aligned} \bar{\Gamma} \times \bar{\Gamma} &\to \mathbb{R} \\ (\gamma_1, \gamma_2) &\mapsto \textstyle\int_I \sqrt{\dot{\gamma}_1(u)\dot{\gamma}_2(u)}\, f_{a,b}\left(\gamma_1(u), \gamma_2(u)\right) du. \end{aligned}$$

We will show a slightly stronger statement, namely the following claim:
Claim A: For $\phi \in \mathcal{D}^*$, there exist $\gamma_1, \gamma_2 \in \bar{\Gamma}$, such that

$$\int_I \sqrt{\dot{\gamma}_1(u)\dot{\gamma}_2(u)}\, f_{a,b}\left(\gamma_1(u), \gamma_2(u)\right) du = \int_I \sqrt{\dot{\phi}(x)}\, f_{a,b}(x, \phi(x)) dx.$$

Conversely, for $\gamma_1, \gamma_2 \in \bar{\Gamma}$, there exists $\phi \in \mathcal{D}^*$ such that

$$\int_I \sqrt{\dot{\phi}(x)}\, f_{a,b}(x, \phi(x))dx = \int_I \sqrt{\dot{\gamma}_1(u)\dot{\gamma}_2(u)}\, f_{a,b}\left(\gamma_1(u), \gamma_2(u)\right) du.$$

To show Claim A, let $\phi \in \mathcal{D}^*$ and μ the corresponding probability measure given by 3.1. By Lebesgue's decomposition theorem, we may write $\mu = \omega_\phi dx + \nu_\phi + \sum_{n\in\mathbb{N}} a_n \delta_{x_n}$, where $\omega_\phi \mathrm{d}x$ is the absolutely continuous part of μ with respect to the Lebesgue measure, ν_ϕ the singular continuous part, and for all $n \in \mathbb{N}$, $x_n \in I$ and $a_n \geq 0$. The latter part can be seen as the (at most countable) jumps of ϕ located at the points x_n and of amplitude a_n. We then consider the set

$$C = \mathrm{Graph}(\phi) \cup \bigcup_{n\in\mathbb{N}} \{x_n\} \times [\phi(x_n), \phi(x_n) + a_n]$$

Then C is compact as it is clearly bounded and its closedness can be shown from the definition using the left continuity of ϕ. Furthermore, C is connected and $\mathcal{H}_1(C) < \infty$, since ϕ is of bounded variation and $\sum_n a_n \leq 1 < \infty$. Therefore, C is the image of a rectifiable curve, that can be reparametrized as an injective, Lipschitz continuous curve γ, cf. (Falconer 1985, Lemmas 3.1 and 3.12). We write

$$\gamma : \begin{cases} [0,1] \to [0,1]^2 \\ u \quad \mapsto (\gamma_1(u), \gamma_2(u)) \end{cases},$$

where γ_1 and γ_2 are Lipschitz continuous, non-decreasing and differentiable almost everywhere (and we will let $\dot{\gamma}_i(u) = 0$ if γ_i is not differentiable in u). We then calculate:

$$\begin{aligned}
&\int_I \sqrt{\dot{\gamma}_1(u)\dot{\gamma}_2(u)}\, f_{a,b}\left(\gamma_1(u), \gamma_2(u)\right) du \\
&= \int_I \sqrt{\dot{\gamma}_1(u)\dot{\gamma}_2(u)}\, f_{a,b}(\gamma_1(u), \gamma_2(u)) 1_{\dot{\gamma}_1(u)>0}(u) du \\
&= \int_I \dot{\gamma}_1(u) \sqrt{\frac{\dot{\gamma}_2(u)}{\dot{\gamma}_1(u)}}\, f_{a,b}(\gamma_1(u), \gamma_2(u)) 1_{\dot{\gamma}_1(u)>0}(u) du \\
&= \int_I \sum_{u\in\gamma_1^{-1}(x)} \sqrt{\frac{\dot{\gamma}_2(u)}{\dot{\gamma}_1(u)}}\, f_{a,b}(\gamma_1(u), \gamma_2(u)) 1_{\dot{\gamma}_1(u)>0}(u) dx,
\end{aligned}$$

where $\gamma_1^{-1}(x) = \{u \in I, \gamma_1(u) = x\}$ and 1_C denotes the indicator function for condition C. The last equality follows from the area formula (Federer 1968, Theorem 3.2.3). Indeed, given the assumptions on γ_1, it is differentiable almost everywhere and we have by (Mattila 1999, p.103) that $\{\gamma_1(u), \dot{\gamma}_1(u) = 0\}$ is of Lebesgue measure zero. Thus, for almost all $x \in [0,1]$, $\gamma_1^{-1}(x)$ is reduced to a single point. Then, by setting $x = \gamma_1(u)$, we have $(\gamma_1(u), \gamma_2(u)) = (x, \phi(x))$ with $\phi(x) = \gamma_2(\gamma_1^{-1}(x))$.

It follows that for almost all u such that $\dot{\gamma_1}(u) > 0$, one has:

$$\dot{\gamma_2}(u) = \frac{d}{du}(\phi \circ \gamma_1(u)) = \dot{\phi}(\gamma_1(u))\dot{\gamma_1}(u) = \dot{\phi}(x)\dot{\gamma_1}(u),$$

and thus $\dot{\phi}(x) = \frac{\dot{\gamma_2}(u)}{\dot{\gamma_1}(u)}$. Going back to the original equality, this leads to

$$\begin{aligned}\int_I \sqrt{\dot{\gamma_1}(u)\dot{\gamma_2}(u)}\, f_{a,b}\,(\gamma_1(u), \gamma_2(u))\, du &= \int_I \sum_{u \in \gamma_1^{-1}(x)} \sqrt{\frac{\dot{\gamma_2}(u)}{\dot{\gamma_1}(u)}}\, f_{a,b}(\gamma_1(u), \gamma_2(u)) 1_{\dot{\gamma_1}(u)>0}(u) dx \\ &= \int_I \sqrt{\dot{\phi}(x)}\, f_{a,b}(x, \phi(x)) dx,\end{aligned}$$

which proves the first direction of Claim A.

To prove the converse direction of Claim A, we let $\gamma_1, \gamma_2 \in \bar{\Gamma}$, where we can choose γ_1, γ_2, up to a reparametrization, to be Lipschitz continuous. We consider the generalized inverse $\gamma_1^- \in \mathcal{D}^*$, and let $\phi = \gamma_2 \circ \gamma_1^-$. By (Trouvé and Younes 2000, Lemma 5.8) the generalized inverse is again an element of $\mathcal{D}^*$. Since γ_2 is Lipschitz continuous, and since composition with Lipschitz functions keeps $\mathcal{D}^*$ invariant, we have that $\phi \in \mathcal{D}^*$, see e.g. (Josephy 1981, Theorem 4). Now one can obtain the desired equality

$$\int_I \sqrt{\dot{\phi}(x)}\, f_{a,b}(x, \phi(x)) dx = \int_I \sqrt{\dot{\gamma_1}(u)\dot{\gamma_2}(u)}\, f_{a,b}\,(\gamma_1(u), \gamma_2(u))\, du,$$

by a similar computation as above, which concludes the proof of Claim A.

Now the first statement of item 1—$a < b$ and $c_1, c_2 \in PC^1(I, \mathbb{R}^d)$– follows directly from Theorem C.1: since c_1 and c_2 are assumed to be piecewise C^1, the function $f_{a,b}$ is bounded and continuous when $\dot{c}_1$ and $\dot{c}_2$ are continuous. For x (resp. y) a point of discontinuity of $\dot{c}_1$ (resp. $\dot{c}_2$), $f_{a,b}$ is not continuous on the vertical segment $\{x\} \times [0, 1]$ (resp. the horizontal segment $[0, 1] \times \{y\}$. Thus, $f_{a,b}$ satisfies (H1). For $a < b$, we have in addition that $f_{a,b} > c$ where $c > 0$, and thus, we also have $f_s > c$ does not vanish. By Theorem C.1 this implies that the minimizer exists in $\mathcal{D}^*$ and is a strictly increasing homeomorphism.

It remains to prove the statements assuming additional smoothness of the curves c_1 and c_2, namely that $\dot{c}_i$ are Lipschitz continuous. Therefore, we show that in this case the function $f_{a,b}$ is also Lipschitz continuous: the application $\theta \mapsto \cos(\frac{a}{2b}\cos^{-1}(\theta))$ is differentiable on $]-1, 1[$, and its derivative is bounded; therefore, $\theta \mapsto \cos(\frac{a}{2b}\cos^{-1}(\theta))$ is Lipschitz continuous on $[-1, 1]$. As $\dot{c}_1$ and $\dot{c}_2$ are Lipschitz continuous, the function $x, y \mapsto \frac{\dot{c}_1(x)\cdot\dot{c}_2(y)}{|\dot{c}_1(x)||\dot{c}_2(y)|}$ is also Lipschitz continuous. Therefore, by composition, $f_{a,b}$ is Lipschitz continuous and thus also Hölder continuous since we are working on a compact domain. We have already shown that the minimizer ϕ is strictly increasing and continuous. As $f_{a,b}$ is strictly positive everywhere we obtain by Theorem C.2 that ϕ is of class $\mathcal{C}^1$ on all of I, which concludes the proof of the second statement of point 1.

For the second item, $a \geq b$, there may exist areas where $f_{a,b} = 0$, which leads to optimizers that have jumps and are thus not continuous. To deal with these difficulty, we

will follow the same approach as in Bruveris (2016) and consider a pair of generalized reparametrization functions, that might have vertical parts but no jumps. Using Claim A we can still focus on maximizing (C.1) on the space of BV functions, which allows us to use again the result of Trouvé and Younes (2000). In particular, by Theorem C.1, cf. (Trouvé and Younes 2000, Proposition 5.1), there exists $\phi \in \mathcal{D}^*$ that maximizes (C.1), and thus by Claim A, there exist $\gamma_1, \gamma_2 \in \bar{\Gamma}$ such that $\operatorname{dist}^{\mathcal{S}}_{a,b}([c_1], [c_2]) = \operatorname{dist}_{a,b}(c_1 \circ \gamma_1, c_2 \circ \gamma_2)$, which proves the existence result in item 2.

Finally, we shall construct a counter-example when $a > b$ if the curves c_1 and c_2 are only in the space $AC(I, \mathbb{R}^d)$. To that end, we adapt the counter-example from (Bruveris 2016, Section 6). Let $0 < \epsilon < \frac{1}{6}$ and define

$$v_1(t) = \begin{pmatrix} \cos \frac{2a\pi}{b}\epsilon t \\ \sin \frac{2a\pi}{b}\epsilon t \end{pmatrix}, \quad v_2 = \begin{pmatrix} \cos \frac{4a\pi}{3b} \\ \sin \frac{4a\pi}{3b} \end{pmatrix}, \quad \text{and} \quad v_3 = \begin{pmatrix} \cos \frac{4a\pi}{3b} \\ -\sin \frac{4a\pi}{3b} \end{pmatrix}.$$

Then we have that $\frac{b}{2a} \cos^{-1}(\frac{v_i(t)\cdot v_j(t)}{|v_i(t)||v_j(t)|}) \geq \frac{\pi}{2}$ for each $i \neq j$, and therefore $f_{a,b} \leq 0$ for those vectors. We define two curves $c_1, c_2 \in AC(I, \mathbb{R}^d)$ such that:

$$\begin{aligned} \dot{c}_1(u) &= v_1(u) 1_A(u) + v_2 1_B(u) \\ \dot{c}_2(u) &= v_1(u) 1_A(u) + v_3 1_B(u) \end{aligned}$$

where $B \subset I$ a modified Cantor set such that B is closed, nowhere dense with $\lambda(B) = \frac{1}{2}$, and $A = I \backslash B$. Following the same proof as in (Bruveris 2016, section 6), we have that

$$\sup_{\gamma_1, \gamma_2 \in \bar{\Gamma}} \int_I f_{a,b}\left(\gamma_1(u), \gamma_2(u)\right) dt = \lambda(A)$$

and is not attained. □

Appendix D. Proof of Theorem 3.2

Proof For $c_1, c_2 \in PC^1(S^1, \mathbb{R}^d)$ and $a, b > 0$ we define the functional

$$F : \begin{cases} S^1 \to \mathbb{R} \\ \tau \mapsto \inf_{\gamma_1, \gamma_2 \in \bar{\Gamma}} \operatorname{dist}_{a,b}(c_1 \circ \gamma_1, c_2 \circ S_\tau \circ \gamma_2) \end{cases}$$

We aim to show that F is continuous, which will directly lead to the desired conclusion. Therefore, let $\epsilon > 0$, $\tau, \tau' \in S^1$. Then

$$\begin{aligned}
|F(x\tau) - F(\tau')| &= |\inf_{\gamma_1,\gamma_2\in\bar{\Gamma}} \mathrm{dist}_{a,b}(c_1\circ\gamma_1, c_2\circ S_\tau\circ\gamma_2) \\
&\quad - \inf_{\gamma_1',\gamma_2'\in\bar{\Gamma}} \mathrm{dist}_{a,b}(c_1\circ\gamma_1', c_2\circ S_{\tau'}\circ\gamma_2')| \\
&= |\mathrm{dist}^{\mathcal{S}}_{a,b}([c_1],[c_2\circ S_\tau]) - \mathrm{dist}^{\mathcal{S}}_{a,b}([c_1],[c_2\circ S_{\tau'}])| \\
&\le \mathrm{dist}^{\mathcal{S}}_{a,b}([c_2\circ S_\tau],[c_2\circ S_{\tau'}] \\
&\le \inf_{\gamma,\gamma'\in\bar{\Gamma}} \mathrm{dist}_{a,b}(c_2\circ S_\tau\circ\gamma, c_2\circ S_{\tau'}\circ\gamma') \\
&= \inf_{\gamma,\gamma'\in\Gamma} \mathrm{dist}^{L^2_{a/(2b)}}\left(Q(c_2\circ S_\tau\circ\gamma), Q(c_2\circ S_{\tau'}\circ\gamma')\right) \\
&\le \mathrm{dist}^{L^2_{a/(2b)}}\left(Q(c_2\circ S_\tau), Q(c_2\circ S_{\tau'})\right).
\end{aligned}$$

By definition, we have that $G^{L^2_{a/(2b)}} \le \max(\frac{a}{2b}, 1)^2\, G^{L^2}$; therefore, we can deduce that

$$\mathrm{dist}^{L^2_{a/(2b)}}\left(Q(c_2\circ S_\tau), Q(c_2\circ S_{\tau'})\right) \le \max\left(\frac{a}{2b}, 1\right)\|Q(c_2\circ S_\tau) - Q(c_2\circ S_{\tau'})\|_{L^2}.$$

Since the space $C(I,\mathbb{R}^d)$ is dense in $L^2(I,\mathbb{R}^d)$, we can choose $g \in C(I,\mathbb{R}^d)$ such that $\|Q(c_2) - g\|_{L^2} \le \epsilon/3$. By change of variable, we also have $\|Q(c_2)\circ S_\tau - g\circ S_\tau\|_{L^2} \le \epsilon/3$ and $\|Q(c_2)\circ S_{\tau'} - g\circ S_{\tau'}\|_{L^2} \le \epsilon/3$. Since g is continuous and I is compact, g is also uniformly continuous by the Heine–Borel theorem. Thus, we have, for $|\tau - \tau'|$ small enough,

$$\|g\circ S_\tau - g\circ S_{\tau'}\|_{L^2} \le \epsilon/3$$

Finally, we have:

$$\begin{aligned}
\left|F(\tau) - F(\tau')\right| &\le \max\left(\frac{a}{2b}, 1\right)\|Q(c_2\circ S_\tau) - Q(c_2\circ S_{\tau'})\|_{L^2} \\
&\le \max\left(\frac{a}{2b}, 1\right)\left(\|Q(c_2\circ S_\tau) - g\circ S_\tau\|_{L^2}\right. \\
&\quad \left. + \|g\circ S_\tau - g\circ S_{\tau'}\|_{L^2} + \|g\circ S_{\tau'} - Q(c_2\circ S_{\tau'}\|_{L^2}\right) \\
&\le \max\left(\frac{a}{2b}, 1\right)(\epsilon/3 + \epsilon/3 + \epsilon/3) = \frac{a}{2b}\epsilon
\end{aligned}$$

Thus, we have shown that F is continuous function on the compact set S^1. Consequently, there exists an optimal $\tau \in S^1$ such that $F(\tau) = \inf_{S^1} F$. Now the remaining statement follows directly from Theorem 3.1. □

References

Bauer, M., Bruveris, M., Harms, P., Michor, P.W.: Vanishing geodesic distance for the Riemannian metric with geodesic equation the kdv-equation. Ann. Glob. Anal. Geom. **41**(4), 461–472 (2012)

Bauer, M., Bruveris, M., Marsland, S., Michor, P.W.: Constructing reparameterization invariant metrics on spaces of plane curves. Differ. Geom. Appl. **34**, 139–165 (2014)

Bauer, M., Bruveris, M., Michor, P.W.: Overview of the geometries of shape spaces and diffeomorphism groups. J. Math. Imaging Vis. **50**(1), 60–97 (2014)

Bauer, M., Harms, P., Preston, S.C.: Vanishing distance phenomena and the geometric approach to SQG. Arch. Ration. Mech. Anal. **235**(3), 1445–1466 (2020)

Bellet, A., Habrard, A., Sebban, M.: A survey on metric learning for feature vectors and structured data (2013). arXiv preprint arXiv:1306.6709

Bruveris, M.: The L^2-metric on $C^\infty(M, N)$ (2018). arXiv:1804.00577

Bruveris, M.: Optimal reparametrizations in the square root velocity framework. SIAM J. Math. Anal. **48**(6), 4335–4354 (2016)

Charon, N., Younes, L.: Shape spaces: from geometry to biological plausibility (2022). arXiv preprint arXiv:2205.01237

Ciarlet, P.G.: Three-Dimensional Elasticity, vol. 20. Elsevier, Amsterdam (1988)

Cohn, D.L.: Measure Theory, vol. 1. Springer (2013)

Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-theoretic metric learning. In: Proceedings of the 24th International Conference on Machine Learning, pp. 209–216 (2007)

Dryden, I.L., Mardia, K.: Statistical Shape Analysis. Wiley, New York (1998)

Falconer, K.J.: The Geometry of Fractal Sets. Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge (1985)

Federer, H.: Geometric Measure Theory. Springer, Berlin (1969)

Ghojogh, B., Ghodsi, A., Karray, F., Crowley, M.: Spectral, probabilistic, and deep metric learning: tutorial and survey (2022). arXiv preprint arXiv:2201.09267

Ghojogh, B., Karray, F., Crowley, M.: Fisher and kernel fisher discriminant analysis: tutorial (2019). arXiv preprint arXiv:1906.09436

Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016)

Gorczowski, K., Styner, M., Jeong, J., Marron, J., Piven, J., Hazlett, H., Pizer, S., Gerig, G.: Multi-object analysis of volume, pose, and shape using statistical discrimination. IEEE Trans. Pattern Anal. Mach. Intell. **32**(4), 652–666 (2010)

Grenander, U., Miller, M.I.: Computational anatomy: an emerging discipline. Quart. Appl. Math. LV **I**(4), 617–694 (1998)

Gurtin, M.E.: An Introduction to Continuum Mechanics (1981)

Hartman, E., Sukurdeep, Y., Charon, N., Klassen, E., Bauer, M.: Supervised deep learning of elastic SRV distances on the shape space of curves. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4425–4433 (2021)

Heitz, M., Bonneel, N., Coeurjolly, D., Cuturi, M., Peyré, G.: Ground metric learning on graphs. J. Math. Imaging Vis. **63**(1), 89–107 (2021)

Jermyn, I.H., Kurtek, S., Klassen, E., Srivastava, A.: Elastic shape matching of parameterized surfaces using square root normal fields. In: European Conference on Computer Vision, pp. 804–817. Springer (2012)

Josephy, M.: Composing functions of bounded variation. Proc. Am. Math. Soc. **83**(2), 354–356 (1981)

Kaya, M., Bilge, H.Ş: Deep metric learning: a survey. Symmetry **11**(9), 1066 (2019)

Kulis, B., et al.: Metric learning: a survey. Found. Tends Mach. Learn. **5**(4), 287–364 (2013)

Lahiri, S., Robinson, D., Klassen, E.: Precise matching of pl curves in $\mathbb{R}^n$ in the square root velocity framework. Geom. Imaging Comput. **2**(3), 133–186 (2015)

Le, T., Cuturi, M.: Unsupervised Riemannian metric learning for histograms using Aitchison transformations. In: International Conference on Machine Learning, pp. 2002–2011. PMLR (2015)

Martins Bruveris. libSRV. Available at https://github.com/martinsbruveris/libsrvf

Mattila, P.: Geometry of Sets and Measures in Euclidean Spaces: Fractals and Rectifiability, vol. 44. Cambridge University Press, Cambridge (1999)

Mavridis, L., Venkatraman, V., Ritchie, D., Morikawa, N., Andonov, R., Cornu, A., Malod-Dognin, N., Nicolas, J., Temerinac-Ott, M., Reisert, M., Burkhardt, H., Axenopoulos, A., Daras, P.: Shrec'10 track: protein model classification, pp. 117–124 (2010)

Michor, P.W., Mumford, D.: An overview of the Riemannian metrics on spaces of curves using the Hamiltonian approach. Appl. Comput. Harmon. Anal. **23**(1), 74–113 (2007)
Mio, W., Srivastava, A., Joshi, S.: On shape of plane elastic curves. Int. J. Comput. Vis. **73**(3), 307–324 (2007). (**Publisher: Springer**)
Mumford, D.B., Michor, P.W.: Riemannian geometries on spaces of plane curves. J. Eur. Math. Soc. **8**(1), 1–48 (2006)
Needham, T.: Kähler structures on spaces of framed curves. Ann. Glob. Anal. Geom. **54**(1), 123–153 (2018)
Needham, T.: Knot types of generalized Kirchhoff rods. J. Knot Theory Ramifications **28**(11), 1940010 (2019)
Needham, T.: Shape analysis of framed space curves. J. Math. Imaging Vis. **61**(8), 1154–1172 (2019)
Needham, T., Kurtek, S.: Simplifying transforms for general elastic metrics on the space of plane curves. SIAM J. Imag. Sci. **13**(1), 445–473 (2020)
Osher, S., Fedkiw, R.: Level Set Methods and Dynamic Implicit Surfaces. Springer (2003)
Srivastava, A., Klassen, E.: Functional and Shape Data Analysis. Springer, Berlin (2016)
Srivastava, A., Klassen, E., Joshi, S.H., Jermyn, I.H.: Shape analysis of elastic curves in Euclidean spaces. IEEE Trans. Pattern Anal. Mach. Intell. **33**(7), 1415–1428 (2010)
Sundaramoorthi, G., Yezzi, A., Mennucci, A.C.: Sobolev active contours. Int. J. Comput. Vis. **73**(3), 345–366 (2007)
Trouvé, A., Younes, L.: Diffeomorphic matching problems in one dimension: designing and minimizing matching functionals. In: European Conference on Computer Vision, pp. 573–587. Springer (2000)
Trouvé, A., Younes, L.: On a class of diffeomorphic matching problems in one dimension. SIAM J. Control Optim. **39**, 1112–1135 (2000)
Vemulapalli, R., Jacobs, D.W.: Riemannian metric learning for symmetric positive definite matrices (2015). arXiv preprint arXiv:1501.02393
Wurzbacher, T., Khesin, B., Michor, P.W.: The flow completion of burgers' equation. Infinite dimensional groups and manifolds. Editor. IRMA Lect. Math. Theor. Phys. **5**, 17–26 (2004)
Xing, E., Jordan, M., Russell, S.J., Ng, A.: Distance metric learning with application to clustering with side-information. Adv. Neural Inf. Process. Syst. **15**, 30 (2002)
Yang, L., Jin, R.: Distance metric learning: a comprehensive survey. Michigan State Univ. **2**(2), 4 (2006)
Ying, Y., Li, P.: Distance metric learning with eigenvalue optimization. J. Mach. Learn. Res. **13**(1), 1–26 (2012)
Younes, L.: Computable elastic distances between shapes. SIAM J. Appl. Math. **58**(2), 565–586 (1998). (**Publisher: Society for Industrial and Applied Mathematics**)
Younes, L., Michor, P.W., Shah, J.M., Mumford, D.B.: A metric on shape space with explicit geodesics. Rendiconti Lincei **19**(1), 25–57 (2008)
Zahn, C.T., Roskies, R.Z.: Fourier descriptors for plane closed curves. IEEE Trans. Comput. **21**(3), 25 (1972)