

A New Geometric Metric in the Space of Curves, and Applications to Tracking Deforming Objects by Prediction and Filtering*

Ganesh Sundaramoorthi[†], Andrea Mennucci[‡], Stefano Soatto[†], and Anthony Yezzi[§]

Abstract. We define a novel metric on the space of closed planar curves which decomposes into three intuitive components. According to this metric, centroid translations, scale changes, and deformations are orthogonal, and the metric is also invariant with respect to reparameterizations of the curve. While earlier related Sobolev metrics for curves exhibit some general similarities to the novel metric proposed in this work, they lacked this important three-way orthogonal decomposition, which has particular relevance for tracking in computer vision. Another positive property of this new metric is that the Riemannian structure that is induced on the space of curves is a smooth Riemannian manifold, which is isometric to a classical well-known manifold. As a consequence, geodesics and gradients of energies defined on the space can be computed using fast closed-form formulas, and this has obvious benefits in numerical applications. The obtained Riemannian manifold of curves is ideal for addressing complex problems in computer vision; one such example is the tracking of highly deforming objects. Previous works have assumed that the object deformation is *smooth*, which is realistic for the tracking problem, but most have restricted the deformation to belong to a finite-dimensional group—such as affine motions—or to finitely parameterized models. This is too restrictive for highly deforming objects such as the contour of a beating heart. We adopt the smoothness assumption implicit in previous work, but we lift the restriction to finite-dimensional motions/deformations. We define a dynamical model in this Riemannian manifold of curves and use it to perform filtering and prediction to infer and extrapolate not just the pose (a finitely parameterized quantity) of an object but its deformation (an infinite-dimensional quantity) as well. We illustrate these ideas using a simple first-order dynamical model and show that it can be effective even on image sequences where existing methods fail.

Key words. visual object tracking, shape analysis, filtering in the space of curves, Sobolev-type metrics

AMS subject classifications. 93, 53

DOI. 10.1137/090781139

1. Introduction. Shape theory is central in computer vision because shapes partially characterize objects in images. Shapes appear in two broad categories of applications:

- *shape optimization*, where we want to find the best shape according to a criterion; examples include image segmentation and object tracking; and
- *shape analysis*, where we study families of shapes for purposes of statistics, (automatic) cataloging, probabilistic modeling, etc.

*Received by the editors December 28, 2009; accepted for publication (in revised form) October 20, 2010; published electronically February 15, 2011. This work was funded by AFOSR FA9550-09-1-0427, NSF grant CCF-0728911, and SNS09MENNB of the Scuola Normale Superiore. A preliminary version of this paper appeared in [48].

<http://www.siam.org/journals/siims/4-1/78113.html>

[†]Computer Science Department, University of California Los Angeles, Los Angeles, CA 90095 (ganeshs@ucla.edu, soatto@ucla.edu).

[‡]Scuola Normale Superiore, 56126 Pisa, Italy (a.mennucci@sns.it).

[§]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 (ayezzi@ece.gatech.edu).

In this paper, the *shapes* we focus on are represented by smoothly immersed planar curves¹ which constitute the boundaries of compact domains (representing the boundaries of physical objects projected into the imaging plane). As is customary in the literature of *active contours* [23], we will somewhat abusively call these curves *embedded curves* in the rest of this paper. It should be noted that the *shape space* classically used in shape optimization (i.e., active contours) is more precisely identified as *the space of embedded curves, up to a choice of parameterization*, whereas in shape analysis the space is usually identified as *the space of embedded curves, up to rotation, translation, scaling, and reparameterization*.

The usage of the space of curves as a *shape space* in applications has predated the proper mathematical study of this shape space by almost two decades. Historically, in the active contour literature, many authors [23, 4, 29, 25, 5, 16] have defined energy functionals on curves, whose minima represent the desired object(s). In more recent works [45, 62, 59, 7], the curve was considered to be a contour partitioning the image into statistically distinct regions. In all cases, the authors utilized the calculus of variations to derive curve evolutions to search for the minima of the energy, often referring to these evolutions as gradient flows. Calling the minimizing flows *gradient flows*, however, implies a certain Riemannian metric on the space of curves.

Modeling the space of curves as a Riemannian manifold also has obvious benefits in shape analysis: indeed the distance between curves can be used for clustering, the geodesic can be used to define the average of two shapes, and so on. However, recently, in [33, 58] it was observed that nearly all previous works on geometric active contours that derive *gradient flows* to minimize energies (i.e., shape optimization) imply a natural notion of Riemannian metric, given by a geometric version of the standard \mathbb{L}^2 inner product, which we will call H^0 in (7). Subsequently, in [33, 57] a surprising property has been shown: the H^0 Riemannian metric on the space of curves is not meaningful, since the “distance” between *any* two curves is zero. (This phenomenon is an example of a more general property; indeed in [34] it is proved that the Fubini–Study metric induces geodesic distance 0 in the nonlinear Grassmannian of all submanifolds of type M in a Riemannian manifold (N, g) .)

This opened a new period of mathematical study, with the goal of finding a new metric in the space of curves that would provide a well-founded model. Many models have been presented, usually to be used either in shape analysis or in shape optimization but not both. The study of shapes as points on an infinite-dimensional space has thus been the subject of considerable interest [35, 32]; models and theory have been presented in [8, 36, 1, 13, 61, 46, 56].

Going back to the shape optimization tasks, many papers contain methods and studies that show that the active contour paradigm is successful in addressing object detection/image segmentation. Those methods can be extended to visual tracking, that is, to temporally varying data; the extension typically involves two steps. One is to collect local statistics from a single image (e.g., intensity histograms, spatial and temporal regularized derivatives, etc.) and use them to partition the image domain into regions that have homogeneous statistics

¹We note that all of the mathematical theory connected with the metric presented in this paper depends only upon the assumption of immersedness. The extra assumptions are needed only when pairing this theory together with various region-based energies that are typically minimized in various computer vision applications, including the tracking applications demonstrated at the end of this paper. Such energies are defined only for curves which have a well-defined interior/exterior, which is not the case for all immersed curves.

[5, 25, 38, 7, 12, 42]. The other step is to incorporate a model of the temporal variation of the deforming object into the tracking algorithm. The simplest way to extend the active contour methodology to time-varying imagery is to use the contour estimated at time “ t ” as initialization for the same gradient-based optimization at time “ $t + 1$ ” [7]. This approach implicitly assumes trivial dynamics (“constant position plus perturbation”), so its prediction would trail an object moving with constant speed with a constant error. Better dynamics (“constant velocity plus perturbation”) have been developed both for parametric [2, 53, 21, 43] and geometric [44, 39, 22] active contour models, the latter implemented using level set methods [40]. While these methods can more accurately predict the (*affine*) motion of the object, their deformation model remains overly simplistic, as—on average—they assume no deformation. So, the prediction of the motion of a jelly fish would extrapolate its affine trajectory (position, orientation, scale, and skew) but “freeze” its shape to the last observation. The dynamical model—and therefore the predictive ability of the tracking scheme—is restricted to the finite-dimensional portion of the actual object deformation.

Recent work has moved beyond the assumption of affine motion [11, 54]. In [11], the motion/deformation is described by a linear autoregressive model defined on combinations of distance functions given as a training set. The applicability of this method is therefore restricted by the availability of training data for every particular object class and its associated deformations. In [54], the authors use a small time-varying basis, which is finite-dimensional but goes beyond affine, to dynamically model local deformations of the contour. In [41], an optimal control approach is constructed to moderate between a model based on optical flow [20] and the results of image segmentation, which results in temporal consistency of the object when compared to frame-by-frame image segmentation. While the model (i.e., a transport equation) allows for deformations beyond affine, the model is defined on the entire domain of the image, and therefore the model is not intrinsic to the geometry of the deforming object. A model that is restricted to the object is natural for tracking because typically the dynamics of the object of interest are less complex than the object plus the background, which can have additional dynamics. Also, the model of [41] is tied to image measurements via optical flow and therefore may have problems in the case of noisy/corrupted image measurements or when the brightness constancy assumption does not hold. In [17], deformations that are not affine are considered by mapping views of a single three-dimensional (3D) object to the 2-sphere (in \mathbb{R}^3), and a constant velocity model is constructed on the mapped space. The mapping limits the shapes to projections from different viewpoints of the chosen 3D object, and, moreover, the approach assumes that the underlying 3D object being viewed is rigid.

1.1. Paper contributions. In the first part of the present work, we construct a Riemannian structure on the space of curves using a geometric-type Sobolev metric, which will be presented in Definition 3.1. While this metric will resemble some prior geometric Sobolev-type metrics for curves, it will exhibit a deliberate modification which was imposed to yield a special three-fold orthogonal decomposition (Theorem 3.4) chosen specifically for its usefulness in visual tracking applications. We present the relevant properties, as well as the formulas and methods to compute geodesics, parallel transport, and gradients of energies defined on curves according to this specialized metric. We show that, using our metric, all these operations can be numerically computed using fast algorithms. This metric builds on the experience of

previous examples of Sobolev metrics [60, 35, 51, 61]; in particular, it extends the ideas found in [61] to the space of all embedded curves, so that this Riemannian metric can be used to address problems in shape analysis and shape optimization and any desired combination of the two.

A Riemannian structure on the space of embedded curves, we show, also provides the means for constructing *dynamical systems on curves*, which is useful for modeling the dynamics of deforming objects. To illustrate these ideas, in the second part of this paper we consider the task of tracking highly deforming objects, such as a walking human, a maneuvering vehicle, a moving animal, etc., from time-varying images. We are interested in the changes of *shape* induced by motion on the boundaries of the projection of objects of interest onto an image. For instance, the silhouette of a walking person undergoes complex deformations, including changes of topology as gaps open between limbs and the trunk. We wish not only to predict the coarse motion or pose of the boundary curves but also to extrapolate their deformation.² To this end, we present a simple (infinite-dimensional) constant velocity dynamical model of the contour and then derive an associated filter that predicts and estimates the contour and its deformation based on local image statistics. For simplicity, in this work we consider intensity statistics, but local spatiotemporal filters could be used as well. Note that in [52] tracking is considered, but prediction and estimation, which is one of the main contributions of the present work, is not performed.

One benefit in using Sobolev-type metrics in tracking is that they favor *smooth* motions of the curve without restricting its deformation [51]. The metric studied in [51], however, did not allow for an efficient computation of geodesics, whereas the new metric presented in this paper allows for efficient calculation of geodesics.

It must be noted that one sure way to avoid infinite-dimensional Riemannian geometry (and the pathologies associated with some metrics) is to model the shape space as a *finite-dimensional manifold*: such is the case, for example, in the work of Kendall [24] (some more recent works include, e.g., [26, 3]). It is certainly possible to model curves using a finite-dimensional family of parameters, for example, using splines. This modeling, however, introduces later problems in tracking applications, since the motion of control points along the curve has to be factored out of the shape dynamics. Moreover, such finite-dimensional representations restrict the space of allowable deformations, which could have detrimental effects when tracking highly deforming objects. In this paper, we propose a tracking method where we define a dynamical model *directly* on the infinite-dimensional space of curves (allowing any smooth deformation), so as to model the deformation of the object of interest in a way that is natural with respect to the object's geometry.

A preliminary conference version of this paper appeared in [48]. The current version gives detailed mathematical proofs and computations of the statements in the conference version. In Appendix C, we have added a new gradient descent procedure for the image segmentation that now corresponds with the new Sobolev metric presented in section 3. We have also added a new experiment testing the approach on MRI data. A forthcoming paper will contain more

²We use the terms “motion” and “(shape) deformation” informally in this section but in a way compatible with the definitions of [47]. In particular, “deformation” means a change in shape, and shape is defined as the quotient of closed planar curve with respect to a finite-dimensional group [30] (in this paper, the group is the identity).

mathematical analysis of the Riemannian metric presented in Definition 3.1.

2. Geometry in the space of curves. We define the space of *smooth planar immersed curves* as

$$(1) \quad M = \{c \in C^\infty(\mathbb{S}^1, \mathbb{R}^2) : |c'(\theta)| \neq 0 \quad \forall \theta \in \mathbb{S}^1\},$$

where \mathbb{S}^1 is the circle, and $c'(\theta)$ is the usual parametric derivative of c . The tangent space $T_c M$ at c is the set of vector fields h on c , i.e., $h : \mathbb{S}^1 \rightarrow \mathbb{R}^2$, which represent *infinitesimal deformations* of c .

We will endow M with a Riemannian metric $\|h\|_c$ which we will then use to define distances between curves in M , geodesics (i.e., shortest paths), the exponential map, and all other standard tools that are found in common texts on Riemannian geometry, such as [14, 27]. All these tools will be essential, in particular, to define dynamical models in the infinite-dimensional space of curves.

2.1. Geometric curves. We are interested in *geometric* curves, i.e., curves considered up to reparameterizations. Let $\text{Diff}(\mathbb{S}^1)$ denote the group of diffeomorphisms of \mathbb{S}^1 ; given $\phi \in \text{Diff}(\mathbb{S}^1)$ and a curve $c \in M$, the composition $c \circ \phi$ is a reparameterization of the c . For technical reasons, we identify the slightly more restrictive subspace M_f of all *freely immersed curves* that are the curves $c \in M$ such that if $\phi \in \text{Diff}(\mathbb{S}^1)$ and $c(\phi(\theta)) = c(\theta)$ for all θ , then ϕ is the identity. This space is a dense open subset of the space M ; see [6] for a detailed proof. The advantage of this space is that we may define the space of *geometric curves* as the quotient space

$$(2) \quad B = M_f / \text{Diff}(\mathbb{S}^1),$$

and it can be shown that B remains a manifold (see [35] or [6, Thm. 1.5], whereas the quotient $M/\text{Diff}(\mathbb{S}^1)$ unfortunately is not a manifold. We will use the notation $[c]$ to indicate all possible parameterizations of the curve c ; that is, $[c]$ is an element of B .

Assuming that the chosen Riemannian metric $\|h\|_c$ in M is reparameterization-invariant, then we can project it to B , so as to define a metric on B . The distance in B can be defined as follows.

Definition 2.1. The distance $d : B \times B \rightarrow \mathbb{R}^+$ between two curves $[c_0] \in B$ and $[c_1] \in B$ is defined by

$$(3) \quad d([c_0], [c_1]) = \inf_{\phi \in \text{Diff}(\mathbb{S}^1)} \inf_{\gamma \in \Gamma(c_0, c_1 \circ \phi)} \text{Len}(\gamma),$$

where

$$\Gamma(c_0, c_1) = \{\gamma : [0, 1] \rightarrow M_f : \gamma(0) = c_0, \gamma(1) = c_1\}$$

is the set of all smooth paths (of intermediate curves) connecting c_0 to c_1 ,

$$(4) \quad \text{Len}(\gamma) = \int_0^1 \|\dot{\gamma}(t)\|_{\gamma(t)} dt$$

is the length of a path γ , $\dot{\gamma}(t) \in T_{\gamma(t)} M_f$ is the velocity (that is, the time derivative) of $\gamma(t)$, and $\|\cdot\|_{\gamma(t)}$ is the norm on $T_{\gamma(t)} M_f$.

Definition 2.2. A minimal geodesic in B between $[c_0]$ and $[c_1]$ is a path γ^* that attains the infimum in (3). Equivalently, up to a time reparameterization of the path γ^* , γ^* solves

$$(5) \quad \inf_{\phi \in \text{Diff}(\mathbb{S}^1)} \inf_{\gamma \in \Gamma(c_0, c_1 \circ \phi)} \mathbb{E}(\gamma),$$

where

$$(6) \quad \mathbb{E}(\gamma) = \int_0^1 \|\dot{\gamma}(t)\|_{\gamma(t)}^2 dt$$

is the action of the path γ .

A critical geodesic is a critical path for the action \mathbb{E} , that is, a solution to the associated Euler–Lagrange equations.

Definition 2.3. The exponential map $\exp : TB \rightarrow B$, where TB is the tangent bundle of B , is

$$\exp_{[c]}(h) = \gamma(1),$$

where $\gamma : [0, 1] \rightarrow B$ is the critical geodesic with $\dot{\gamma}(0) = h \in T_{[c]}B$.

3. A geometric Sobolev-type metric. In this section, we define a Riemannian metric on the space of curves M ; this metric is invariant with respect to reparameterizations of the curve. This metric will allow us to compute geodesics in B , distances between contours, gradients for active contours, etc.

For any fixed immersed curve c , let $L(c)$ be the length of c , and, given any $g : \mathbb{S}^1 \rightarrow \mathbb{R}^2$, let

$$D_s g \doteq \frac{g'}{|c'|}$$

be the derivative with respect to the arclength, and let

$$\begin{aligned} \int_c g(s) ds &\doteq \int_{\mathbb{S}^1} g(\theta) |c'(\theta)| d\theta, \\ \oint_c g(s) ds &\doteq \frac{1}{L(c)} \int_c g(s) ds \end{aligned}$$

denote the integral and the mean with respect to the arclength. We will often use the notation

$$\bar{g} = \oint_c g(s) ds$$

(not to be confused with the complex conjugate, which we will denote by g^*); we will call \bar{c} the *centroid* of c .

In the active contour literature, one defines an energy functional $E_{ac} : B \rightarrow \mathbb{R}$ that is constructed so that the minimal energy contour $[c] \in B$ represents the boundary of an object of interest in an image. Typically, a gradient descent procedure is used to optimize E_{ac} . To define a gradient, one needs a Riemannian metric that is often tacitly assumed to be

$$(7) \quad \langle h, k \rangle_{H^0} \doteq \oint_c h(s) \cdot k(s) ds,$$

where s is the arclength parameter of c .³ (We omit the dependency of the metric on c for ease of notation.) However, it is shown in [33] that in this metric the distance between any two curves in B is zero. Therefore, in [49], we considered the following geometric Sobolev-type metric:

$$(8) \quad \langle h, k \rangle_{\tilde{H}^1} \doteq \bar{h} \cdot \bar{k} + \lambda L(c) \int_c D_s h \cdot D_s k \, ds,$$

where $\lambda > 0$ is a (constant) weight.⁴ This metric \tilde{H}^1 was shown to yield favorable properties for active contours (i.e., gradient descent of E_{ac}) [51, 50, 52]. Moreover, this metric defines a well-behaved Riemannian geometry, in that the distance between different curves is positive. However, while having some of the appropriate properties for our applications, it is not particularly easy to compute geodesics in this metric. Moreover, the shape space obtained by coupling the immersed curves with this metric is not decomposable into components that are natural for visual tracking applications in computer vision. Therefore, we propose in the remainder of this paper a new variant which was constructed deliberately for its favorable properties in such applications.

3.1. A new Sobolev-type metric. We recall that the Gâteaux derivative of a function $f : M \rightarrow \mathbb{R}^k$ is defined by

$$Df(c; h) \doteq \left. \frac{d}{dt} f(c + th) \right|_{t=0} = \lim_{t \rightarrow 0} \frac{f(c + th) - f(c)}{t}.$$

The Gâteaux derivatives of the centroid and of the length are

$$\begin{aligned} D(\bar{c})(c; h) &= \oint_c p(h) \, ds, \\ DL(c)(c; h) &= - \int_c h \cdot D_s^2 c \, ds, \end{aligned}$$

where

$$p(h) \doteq h - (h \cdot D_s c) D_s c - (h \cdot D_s^2 c)(c - \bar{c}).$$

To define the new metric \mathbb{H} , we first define the following decomposition for $c \in M$ and $h \in T_c M$:

$$(9) \quad h = h^t + h^l(c - \bar{c}) + L(c)h^d,$$

where h^t is the component of h that changes the centroid of c , $h^l(c - \bar{c})$ is the component of h that changes the scale (length) of c , and h^d is the component of h that deforms c . The

³Note that in this section we are considering “geometric metrics” of curves, where the derivations and integrals are performed with respect to the arc parameter; these metrics are distinguished by using the letter “H” in different forms. Later we will instead use the notation \mathbb{L}^2 for the standard Hilbert metric where integration is performed in the parametric variable, as usual.

⁴More classical Sobolev-type metrics for active contours were also presented in [10, 9, 51, 52]. An overview of Sobolev-type metrics, methods, and mathematical results was presented in [35]; in section 4.3 of [35] it is shown that, for a large class of Sobolev metrics, critical geodesics exist for short time and smooth initial data.

components h^t and h^l of h are defined by

$$(10) \quad h^t = D(\bar{c})(c; h) = \oint_c p(h) \, ds \in \mathbb{R}^2,$$

$$(11) \quad h^l = D(\log L(c))(c; h) = - \oint_c h \cdot D_s^2 c \, ds \in \mathbb{R},$$

$$(12) \quad h^d = \frac{1}{L(c)} [h - h^t - h^l(c - \bar{c})].$$

The component h^d deforms the curve without scaling or translating, since

$$D(L(c))(c; h^d) = 0, \quad D\bar{c}(c; h^d) = 0.$$

Definition 3.1. If $h, k \in T_c M$ are decomposed as above, then we define the Riemannian metric \mathbb{H} as

$$(13) \quad \langle h, k \rangle_{\mathbb{H}} \doteq h^t \cdot k^t + \lambda_l h^l k^l + \lambda_d L(c)^2 \oint_c D_s h^d \cdot D_s k^d \, ds,$$

where the first two products are the Euclidean dot products, the last term is a normalized geometric Sobolev metric, and $\lambda_l, \lambda_d > 0$ are (constant) weights.

Note that in the notation $\langle h, k \rangle_{\mathbb{H}}$ we have omitted the dependency of the scalar product on c for ease of notation. Note also that the third term of the metric may be rewritten directly as a function of $h \in T_c M$ by using the identity

$$(14) \quad L(c)^2 \oint_c D_s h^d \cdot D_s k^d \, ds = \oint_c D_s h \cdot D_s k \, ds - \oint_c D_s h \cdot D_s c \, ds \oint_c D_s c \cdot D_s k \, ds.$$

This new metric enjoys the following properties:

1. Centroid translations, scale changes, and deformations of the curve are orthogonal. Moreover, the space of curves can be decomposed into a product space consisting of three components as shown in Theorem 3.4.
2. Sobolev-type metrics favor *smooth* but otherwise unrestricted infinite-dimensional deformations [51], and they have a coarse-to-fine evolution behavior [52]. For this reason, the old metric \tilde{H}^1 has proven useful in framewise object detection/image segmentation for visual tracking.

The new metric \mathbb{H} is indeed metrically equivalent to the old metric \tilde{H}^1 .

Theorem 3.2.

$$a_1 \frac{L(c)}{1 + L(c)} \|h\|_{\mathbb{H}} \leq \|h\|_{\tilde{H}^1} \leq a_2 (1 + L(c)) \|h\|_{\mathbb{H}},$$

where $0 < a_1 < a_2$ are constants depending on $\lambda, \lambda_l, \lambda_d$. (The proof is in section A.1.)

Therefore the new metric \mathbb{H} inherits the favorable properties of the old metric \tilde{H}^1 for shape optimization tasks.

3. As was the case for the old metric \tilde{H}^1 , there is a fast and easy way to compute gradients of commonly used energies with respect to the new metric \mathbb{H} . The method will be presented in Proposition 4.6.

4. Geodesics in this new metric (i.e., the optimization problem in (3)) can be numerically computed efficiently. The method is explained in the next section.

For all these reasons, this new metric seems a state of the art choice for addressing problems in shape theory when the shape of interest can be represented as a curve: indeed, it entails a sound mathematical model that can address problems in shape analysis and shape optimization and any desired combination of the two.

3.2. Space decomposition. We will find it useful to define a submanifold M_d of M .

Definition 3.3. *Let*

$$(15) \quad M_d = \{c \in M : L(c) = 1, \bar{c} = 0\},$$

which is the space of all smooth immersed curves with the unit length and with the centroid at the origin. This is a smooth submanifold of M (the proof follows from a corollary of the Nash–Moser theorem; see [19]). Its tangent space at $\tilde{c} \in M$ is

$$(16) \quad T_{\tilde{c}}M_d = \left\{ h \in T_{\tilde{c}}M \mid \int_{\tilde{c}} (D_s^2 \tilde{c}) \cdot h \, ds = 0, \int_{\tilde{c}} p(h) \, ds = 0 \right\}.$$

We associate the Euclidean metric to $\mathbb{R}^n \times \mathbb{R}$ and the metric

$$(17) \quad \langle h, k \rangle_{M_d} = \int_{\tilde{c}} D_s h \cdot D_s k \, ds$$

with M_d . This metric is the restriction of the metric \mathbb{H} to M_d . The metric \mathbb{H} is associated with an isometry between the space of curves M and the space $\mathbb{R}^2 \times \mathbb{R} \times M_d$.

Theorem 3.4. *Let $\lambda_l = \lambda_d = 1$ (in (3.1)) for simplicity. We define a map and its inverse by*

$$(18) \quad c \in M \mapsto \left(v = \bar{c}, \, l = \log L(c), \, \tilde{c} = \frac{c - \bar{c}}{L(c)} \right) \in \mathbb{R}^2 \times \mathbb{R} \times M_d,$$

$$(19) \quad (v, l, \tilde{c}) \in \mathbb{R}^2 \times \mathbb{R} \times M_d \mapsto v + e^l \tilde{c} \in M.$$

This map is an isometry. (The proof is in section A.2.)

To the best of our knowledge, the metric \mathbb{H} is the first example of a Sobolev-type metric of immersed curves to exhibit this useful decomposition of the entire space M . Other known metrics would provide a decomposition only of the infinitesimal motions h , i.e., a decomposition of the tangent space $T_c M$ rather than the space M itself. This decomposition, moreover, greatly simplifies the proof of some of the mathematical results in the following sections.

3.3. Computing geodesics and the exponential map. Let

$$C : \mathbb{S}^1 \times [0, 1] \rightarrow \mathbb{R}^2, \, (\theta, t) \mapsto C(\theta, t)$$

denote a time-varying family of closed curves (i.e., a homotopy) corresponding to a path $\gamma : [0, 1] \rightarrow M$, i.e., $C(\theta, t) = \gamma(t)(\theta)$. We will write either $\partial_t C$ or \dot{C} to denote the time derivative of C . We have that

$$(20) \quad \|\partial_t C\|_{\mathbb{H}}^2 = |\partial_t \bar{C}|^2 + \lambda_l (\partial_t (\log L(C)))^2 + \lambda_d L(c)^2 \oint_C |D_s (\partial_t C)|^2 \, ds.$$

Following (6), let

$$\mathbb{E}(C) \doteq \int_0^1 \|\partial_t C\|_{\mathbb{H}}^2 dt$$

be the *action* of the homotopy C . Using the above fact and some manipulations, one can show that geodesics in this metric are invariant to scale and translations.

Proposition 3.5 (invariance of the action \mathbb{E}). *Let $c_0, c_1 \in M$, and let*

$$\tilde{c}_1 = v + e^\rho(c_1 - \overline{c}_1) + \overline{c}_1$$

be a scaling and translation of c_1 . Suppose that C is a homotopy connecting c_0 to c_1 , and let

$$\tilde{C} = tv + e^{t\rho}(C - \overline{C}) + \overline{C}$$

be a homotopy connecting c_0 to \tilde{c}_1 ; then

$$\mathbb{E}(\tilde{C}) = \mathbb{E}(C) + \text{const},$$

where the “constant term” depends only on the end curves c_0, c_1 and v, ρ . As a corollary, C is a geodesic connecting c_0 to c_1 if and only if \tilde{C} is a geodesic connecting c_0 to \tilde{c}_1 .

This result can be seen as a corollary of Theorem 3.4.

The above is also related to the following conservation laws.

Proposition 3.6 (momenta). *Suppose that C is a geodesic; then the following quantities are conserved:*

- [translation] $\partial_t \overline{C}$ is constant;
- [scaling] $\partial_t(\log L(C))$ is constant;
- [rotation] the angular momentum (that may be expressed in two equal ways)

$$(21) \quad \oint_C (AD_s C) \cdot (\partial_t D_s C) ds = \oint_C (AD_s C) \cdot (D_s \partial_t C) ds$$

is constant for any antisymmetric matrix A ;

- [reparameterization] $(D_s C) \cdot (D_s^2 \partial_t C)$ is constant in t for any fixed $\theta \in \mathbb{S}^1$.

(The proof is in section A.3.)

This, in particular, means that, along a geodesic C connecting c_0 to c_1 ,

$$(22) \quad \overline{C} = (1-t)\overline{c}_0 + t\overline{c}_1,$$

$$(23) \quad \log L(C) = (1-t)\log L(c_0) + t\log L(c_1).$$

The previous results imply that to compute a minimal geodesic in M between c_0 and c_1 we apply the following procedure:

1. define

$$\tilde{c}_0 \doteq \frac{c_0 - \overline{c}_0}{L(c_0)}, \quad \tilde{c}_1 \doteq \frac{c_1 - \overline{c}_1}{L(c_1)};$$

2. compute a geodesic \tilde{C} between \tilde{c}_0 and \tilde{c}_1 in the space M_d ;
3. rebuild the geodesic in M :

$$C(t, \cdot) = L^{1-t}(c_0)L^t(c_1)\tilde{C}(t, \cdot) + (1-t)\overline{c}_0 + t\overline{c}_1.$$

3.3.1. Representing smooth curves using the square root lifting. We have therefore reduced the problem of computing geodesics in M to computing geodesics in the space M_d of unit length curves with the centroid at the origin. To this end, we will identify the plane \mathbb{R}^2 with the complex numbers \mathbb{C} and consider curves as smooth maps $c : \mathbb{R} \rightarrow \mathbb{C}$ that are periodic of period 1.

Given two smooth functions $e, f : \mathbb{R} \rightarrow \mathbb{R}$, we define the map Φ introduced by Younes et al. in [60, 61] by

$$(24) \quad c(\theta) = \Phi(e, f)(\theta) \doteq c(0) + \frac{1}{2} \int_0^\theta (e + if)^2(\xi) d\xi,$$

where i denotes the imaginary unit; this map uniquely identifies a curve up to the choice of the base point $c(0)$ or, equivalently, up to the choice of the centroid \bar{c} .

Note that for $c = \Phi(e, f)$ to be a closed curve we must have that

$$(25) \quad 0 = c(1) - c(0) = \frac{1}{2} \int_0^1 (e + if)^2(\theta) d\theta = \frac{1}{2} \int_0^1 [e^2(\theta) - f^2(\theta) + 2ie(\theta)f(\theta)] d\theta,$$

and for the curve to be of unit length we must have that

$$(26) \quad 1 = \int_0^1 |\dot{c}(\theta)| d\theta = \frac{1}{2} \int_0^1 (e^2(\theta) + f^2(\theta)) d\theta.$$

The conditions (25) and (26) imply that the pair (e, f) belongs to

$$(27) \quad \mathbf{St}(2, C^\infty) = \{(e, f) \in C^\infty \times C^\infty \mid \|e\|_{\mathbb{L}^2} = \|f\|_{\mathbb{L}^2} = 1, \langle e, f \rangle_{\mathbb{L}^2} = 0\},$$

where the above \mathbb{L}^2 norms and inner product are the standard ones on $\mathbb{L}^2([0, 1])$. $\mathbf{St}(2, C^\infty)$ is known as a *Stiefel* manifold. It is a Riemannian manifold when we use the metric induced from the scalar product $\mathbb{L}^2 \times \mathbb{L}^2$ on the frames (e, f) .

Vice versa, let c be a closed unit length immersed smooth curve. We express c' in polar coordinates as

$$c'(\theta) = r(\theta)(\cos \varphi(\theta) + i \sin \varphi(\theta)).$$

We can then define an inverse of Φ (called the *square-root lifting*) by setting

$$(28) \quad e(\theta) + if(\theta) = \Phi^{-1}(c)(\theta) = \sqrt{2r(\theta)} \left(\cos \frac{\varphi(\theta)}{2} + i \sin \frac{\varphi(\theta)}{2} \right).$$

Note that (e, f) and $(-e, -f)$ are the only two inverses of c .

We identify inside $\mathbf{St}(2, C^\infty)$ an open subset \mathbf{St}^0 of all smooth frames (e, f) that represent the curves $c \in M_d$: then $\Phi : \mathbf{St}^0 \rightarrow M_d$ is a smooth two-fold covering. It is shown in [61] that Φ is a Riemannian isometry from \mathbf{St}^0 to M_d endowed with a Sobolev-type metric.

Theorem 3.7 (see Theorem 2.2 in [61]). Let $\tilde{c} \in M_d$, and let $h \in T_{\tilde{c}}M_d$ (see (16)) and $(e, f) \in \mathbf{St}^0$, $(\delta e, \delta f) \in T_{(e,f)}\mathbf{St}^0$ be the corresponding Stiefel representations, i.e.,

$$D\Phi(e, f; \delta e, \delta f) = h.$$

Then

$$\int_{\tilde{c}} |D_s h|^2 ds = 2 \int_0^1 (\delta e)^2 + (\delta f)^2 d\theta.$$

To exploit this theorem, we associate with M_d the (restriction of) the Sobolev-type metric $\int_{\tilde{c}} |D_s h|^2 ds$; then Φ maps isometrically the (restriction of) the metric $\mathbb{L}^2 \times \mathbb{L}^2$ in \mathbf{St}^0 to the chosen metric on M_d . So this result couples perfectly with the isometry shown in Theorem 3.4.

3.3.2. Completing $\mathbf{St}(2, C^\infty)$ to $\mathbf{St}(2, \mathbb{L}^2)$. The space $\mathbf{St}(2, C^\infty)$ is not a complete smooth Riemannian manifold; its metric completion is the space $\mathbf{St}(2, \mathbb{L}^2)$ of all orthonormal frames (e, f) of two generic vectors $e, f \in \mathbb{L}^2 = \mathbb{L}^2([0, 1])$. $\mathbf{St}(2, \mathbb{L}^2)$ has many interesting properties:

- $\mathbf{St}(2, \mathbb{L}^2)$ is a smooth embedded submanifold of $\mathbb{L}^2 \times \mathbb{L}^2$.
- $\mathbf{St}(2, \mathbb{L}^2)$ is a complete smooth Riemannian manifold modeled on a Hilbert space. This implies that the exponential map is well defined. We will show in the next section that the exponential map can be written in closed form and computed efficiently. (The formula proves that any two given pairs (e_0, f_0) and (e_1, f_1) can be connected by a critical geodesic; that is, the exponential map is surjective.)
- Completeness is also an important hypothesis in any mathematical proof that would aim to prove that an optimization method is well posed.
- A frame $(e, f) \in \mathbf{St}(2, \mathbb{L}^2)$ can be mapped to a closed (possibly nonsmooth) curve using the map Φ ; but the map is not a two-fold covering; that is, a curve has many representations in $\mathbf{St}(2, \mathbb{L}^2)$.
- Vice versa, any closed curve c that is absolutely continuous (that is, c' exists as an integrable function) can be represented by a pair $(e, f) \in \mathbf{St}(2, \mathbb{L}^2)$.

For these reasons, we will consider $\mathbf{St}(2, \mathbb{L}^2)$ instead of $\mathbf{St}(2, C^\infty)$ in the rest of this paper.

3.3.3. Computing critical geodesics. Due to the above theorems and remarks, we now present the calculus of geodesics in $\mathbf{St}(2, \mathbb{L}^2)$.

Classically, the Stiefel manifold $\mathbf{St}(p, \mathbb{R}^n)$ is defined as the set of all frames composed of p orthonormal vectors in \mathbb{R}^n (with $1 \leq p \leq n$); those frames are represented as $n \times p$ matrices. Geodesics in Stiefel manifolds $\mathbf{St}(p, \mathbb{R}^n)$ are known to have closed form solutions as demonstrated by Edelman, Arias, and Smith [15].⁵

Proposition 3.8 (exponential map in $\mathbf{St}(p, \mathbb{R}^n)$). *Let $Y : [0, 1] \rightarrow \mathbf{St}(p, \mathbb{R}^n)$ be a path, and suppose that $\mathbf{St}(p, \mathbb{R}^n)$ is endowed with the Euclidean metric, i.e.,*

$$\langle A, B \rangle = \text{tr}(A^T B);$$

then the geodesic equation is

$$(29) \quad \ddot{Y} + Y(\dot{Y}^T \dot{Y}) = 0.$$

The solution is

$$(30) \quad (Y(t)e^{At}, \dot{Y}(t)e^{At}) = (Y(0), \dot{Y}(0)) \exp t \begin{pmatrix} A & -S \\ \text{Id} & A \end{pmatrix},$$

⁵Edelman, Arias, and Smith [15] credit a personal communication by R. A. Lippert for the final closed form formula (30).

where $A = Y^T(0)\dot{Y}(0)$, $S = \dot{Y}^T(0)\dot{Y}(0)$, and Id is the $p \times p$ identity matrix.

The proof and discussion of these results is in section 2.2.2 of [15].

The solution in (30), while written for $\mathbf{St}(p, \mathbb{R}^n)$, extends to $\mathbf{St}(2, \mathbb{L}^2)$. Indeed, (30) shows that the columns of $Y(t), \dot{Y}(t)$ remain in the space spanned by the columns of $Y(0), \dot{Y}(0)$ for all t .

Proposition 3.9 (exponential map in $\mathbf{St}(2, \mathbb{L}^2)$). *Let $(e, f) : [0, 1] \rightarrow \mathbf{St}(2, \mathbb{L}^2)$ be a geodesic path such that*

$$(e(0), f(0)) = (e^*, f^*) \in \mathbf{St}(2, \mathbb{L}^2)$$

and

$$(\dot{e}(0), \dot{f}(0)) = (\delta e, \delta f) \in T_{(e^*, f^*)}\mathbf{St}(2, \mathbb{L}^2).$$

Define an orthonormal set⁶

$$(31) \quad \mathcal{B} = \{e^*, f^*, \tilde{e}, \tilde{f}\} \subset \text{span}(\{e^*, f^*, \delta e, \delta f\})$$

according to the usual \mathbb{L}^2 metric. Then the geodesic in $\mathbf{St}(2, \mathbb{L}^2)$ is given by

$$\begin{aligned} e(t) &= Y_1^1(t)e^* + Y_1^2(t)f^* + Y_1^3(t)\tilde{e} + Y_1^4(t)\tilde{f}, \\ f(t) &= Y_2^1(t)e^* + Y_2^2(t)f^* + Y_2^3(t)\tilde{e} + Y_2^4(t)\tilde{f}, \end{aligned}$$

where Y_j^i denotes the i th component of $Y_j \in \mathbb{R}^4$, $Y : [0, 1] \rightarrow \mathbf{St}(2, \mathbb{R}^4)$ is the geodesic in $\mathbf{St}(2, \mathbb{R}^4)$ that satisfies

$$(32) \quad Y(0) = ((1, 0, 0, 0)^T, (0, 1, 0, 0)^T),$$

$$(33) \quad \dot{Y}(0) = (a, b) \in T_{Y(0)}\mathbf{St}(2, \mathbb{R}^4),$$

and (a, b) are the representations of $\delta e, \delta f$ relative to \mathcal{B} .

The geodesic $\gamma : [0, 1] \rightarrow M_d$ can then be recovered from the geodesic in $\mathbf{St}(2, \mathbb{L}^2)$ via the isometry Φ ; but note that even if the initial curve is smooth and immersed and the vector field is smooth, it is not guaranteed that the curve will be immersed for all t (cf. section 3.3.2 and the examples in [61]).

3.3.4. Computing minimal geodesics. The formula (30) gives the geodesic as a function of the initial position and direction; this is the exponential map. However, to compute geodesics between two curves (the so-called *logarithmic map*), it is necessary to have a formula for Y in terms of the boundary conditions $Y(0)$ and $Y(1)$. We are not aware of such an explicit formula, and therefore we use an iterative algorithm that computes the initial direction $\dot{Y}(0)$ of the geodesic Y such that $Y(0) = Y_0$ and $Y(1) = Y_1$.

As in the previous proposition, we can reduce the computation to $\mathbf{St}(2, \mathbb{R}^4)$: indeed, we represent the end curves as two frames (e_0, f_0) and (e_1, f_1) , respectively, and then define an orthonormal set

$$\mathcal{B} = \{e_0, f_0, \tilde{e}, \tilde{f}\} \subset \text{span}(\{e_0, f_0, e_1, f_1\}).$$

⁶If $\{e^*, f^*, \delta e, \delta f\}$ do not span a four-dimensional (4D) space, then \tilde{f} may be chosen arbitrarily.

In this coordinate system, the end curves are represented as $Y_0, Y_1 \in \mathbb{R}^{4 \times 2}$, and Y_0 is given by (32). All possible tangent vectors at Y_0 are

$$\dot{Y}(0) = ((0, \alpha, v_1, v_3)^T, (-\alpha, 0, v_2, v_4)^T)$$

with $\alpha, v_1, v_2, v_3, v_4 \in \mathbb{R}$, in conformity with the representation in (2.6) of [15].

To compute the geodesic, we minimize the energy $E : \mathbb{R}^5 \rightarrow \mathbb{R}^+$,

$$(34) \quad E(\alpha, v_1, v_2, v_3, v_4) = |Y(1) - Y_1|^2,$$

where $Y(1)$ is given according to (30),

$$Y(1) = (Y(0), \dot{Y}(0)) \exp \begin{pmatrix} A & -S \\ \text{Id}_{2 \times 2} & A \end{pmatrix} \text{Id}_{4 \times 2} e^{-A},$$

$$A = \begin{pmatrix} 0 & -\alpha \\ \alpha & 0 \end{pmatrix}, \quad S = \begin{pmatrix} \alpha^2 + v_1^2 + v_3^2 & v_1 v_2 + v_3 v_4 \\ v_1 v_2 + v_3 v_4 & \alpha^2 + v_2^2 + v_4^2 \end{pmatrix},$$

where $\text{Id}_{4 \times 2} = ((1, 0, 0, 0)^T, (0, 1, 0, 0)^T)$. Note that the energy E is not convex.

We minimize the energy in (34) by standard gradient descent in \mathbb{R}^5 , initializing the descent with $(\alpha, v_1, v_2, v_3, v_4) = (0, 0, 0, 0, 0)$. The gradient is computed as follows.

Proposition 3.10. *The partial derivatives of the energy E in (34) are given by*

$$(35) \quad \begin{aligned} \partial_* E(\alpha, v_1, v_2, v_3, v_4) &= (Y(1) - Y_1) \cdot \left[(0_{4 \times 2}, \partial_* \dot{Y}(0)) \exp(N) \text{Id}_{4 \times 2} e^{-A} \right. \\ &\quad + (Y(0), \dot{Y}(0)) \int_0^1 \exp(tN) \partial_* N \exp((1-t)N) dt \text{Id}_{4 \times 2} e^{-A} \\ &\quad \left. + (Y(0), \dot{Y}(0)) \exp(N) \text{Id}_{4 \times 2} e^{-A} \partial_* A \right], \end{aligned}$$

where $*$ = $\alpha, v_1, v_2, v_3, v_4$,

$$N = \begin{pmatrix} A & -S \\ \text{Id}_{2 \times 2} & A \end{pmatrix},$$

the partials of A are

$$\partial_\alpha A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \text{and } \partial_{v_i} A = 0_{2 \times 2} \quad \text{for } i = 1, 2, 3, 4,$$

and the partials of S are

$$\begin{aligned} \partial_\alpha S &= 2\alpha \text{Id}_{2 \times 2}, \quad \partial_{v_1} S = \begin{pmatrix} 2v_1 & v_2 \\ v_2 & 0 \end{pmatrix}, \quad \partial_{v_2} S = \begin{pmatrix} 0 & v_1 \\ v_1 & 2v_2 \end{pmatrix}, \\ \partial_{v_3} S &= \begin{pmatrix} 2v_3 & v_4 \\ v_4 & 0 \end{pmatrix}, \quad \partial_{v_4} S = \begin{pmatrix} 0 & v_3 \\ v_3 & 2v_4 \end{pmatrix}. \end{aligned}$$

Proof. This is a standard calculation based on the matrix differentiation formula

$$D(\exp X)(X; Z) = \int_0^1 \exp(tX) Z \exp((1-t)X) dt,$$

which can be found in [31].⁷ Note that when X and Z commute, then $D(\exp X)(X; Z) =$

⁷A more general result for infinite-dimensional Lie manifolds has been proven in [18].

$\exp(X)Z$. ■

The derivative of the matrix exponential can be computed efficiently using a technique resembling the fast Fourier transform (see Appendix B for details). The geodesic connecting Y_0 to Y_1 is then obtained by setting $Y(0) = Y_0$ and $\dot{Y}(0) = ((0, \alpha^*, v_1^*, v_3^*)^T, (-\alpha^*, 0, v_2^*, v_4^*)^T)$, where $(\alpha^*, v_1^*, v_2^*, v_3^*, v_4^*)$ is the minimum point of (34).

3.4. Geodesics in the space of geometric curves, B . Up to this point, we have specified how to compute geodesics and the exponential map in M according to the metric \mathbb{H} ; however, we are interested in these operations in the geometric space B . To be mathematically precise, in this section we will consider M to be the space of all freely immersed smooth curves.

We first define two objects of interest. We define the *vertical space* as

$$(36) \quad V_c M = \{h \in T_c M : h = \beta c', \beta : \mathbb{S}^1 \rightarrow \mathbb{R}\}.$$

This is the set of infinitesimal deformations of c that do not change the geometry of the curve c but only its parameterization. We then define the *horizontal space* as

$$(37) \quad W_c M \doteq (V_c M)^\perp = \{h \in T_c M : \langle h, k \rangle_{\mathbb{H}} = 0 \forall k \in V_c M\}.$$

We use the horizontal space $W_c M$ as a model of the tangent space $T_{[c]} B$.

Geodesics in B (with the metric induced from M) correspond to geodesics in M , provided they are horizontal, i.e., $\dot{\gamma}(t) \in W_{\gamma(t)} M$ for all t . Equivalently, it is enough that $\dot{\gamma}(1) \in W_{\gamma(1)} M$ and γ is a geodesic in M for γ to be a geodesic in B . We now give the condition to determine whether $\dot{\gamma}(1) \in W_{\gamma(1)} M$, which can be found in [61]. As a first step we consider $\phi \in \text{Diff}(\mathbb{S}^1)$ such that $\phi(0) = 0$: then, by direct computation,

$$\begin{aligned} \Phi(e, f)(\phi^{-1}(\theta)) &= c(0) + \frac{1}{2} \int_0^{\phi^{-1}(\theta)} (e + if)^2(\xi) d\xi \\ &= c(0) + \frac{1}{2} \int_0^\theta (e + if)^2(\phi(\xi)) \phi'(\xi) d\xi = \Phi\left(\sqrt{\phi'}(e \circ \phi), \sqrt{\phi'}(f \circ \phi)\right)(\theta), \end{aligned}$$

where $\phi' = d\phi/d\theta$ is the derivative of ϕ . Therefore, the action of reparameterization on a point $(e, f) \in \mathbf{St}(2, C^\infty)$ is

$$(e, f) \mapsto \sqrt{\phi'}(e \circ \phi, f \circ \phi),$$

and the differential of the action above evaluated at the identity in the direction $\beta : \mathbb{S}^1 \rightarrow \mathbb{R}$ is

$$\left(\frac{1}{2}\beta'e + \beta e', \frac{1}{2}\beta'f + \beta f'\right).$$

The collection of all such differentials above for all β is the vertical space at (e, f) . For $(\delta e, \delta f) \in T_{(e, f)} \mathbf{St}(2, C^\infty)$ to be in the horizontal space, it must be orthogonal to all vertical perturbations:

$$\left\langle \delta e, \frac{1}{2}\beta'e + \beta e' \right\rangle_{\mathbb{L}^2} + \left\langle \delta f, \frac{1}{2}\beta'f + \beta f' \right\rangle_{\mathbb{L}^2} = \frac{1}{2} \langle \beta, e'\delta e - e(\delta e)' \rangle_{\mathbb{L}^2} + \frac{1}{2} \langle \beta, f'\delta f - f(\delta f)' \rangle_{\mathbb{L}^2} = 0$$

for all β , that is,

$$(38) \quad \Omega(e, \delta e) + \Omega(f, \delta f) = 0,$$

where

$$(39) \quad \Omega(a, b) \doteq ab' - ba'.$$

To compute a geodesic between $[c_0], [c_1] \in B$, we use the following algorithm.

Algorithm 3.11 (computing geodesics in $\mathbf{St}(2, C^\infty)/\text{Diff}(\mathbb{S}^1)$). Let $\phi_0 \in \text{Diff}(\mathbb{S}^1)$ be an initial reparameterization of the end curve c_1 in its Stiefel representation (one possible initialization is given in Remark 3.12). Define a sequence $\phi_k \in \text{Diff}(\mathbb{S}^1)$ iterating the following steps:

1. Compute the geodesic $(e_{\phi_k}(t), f_{\phi_k}(t))$, $t \in [0, 1]$, in $\mathbf{St}(2, C^\infty)$ between (e_0, f_0) and $\sqrt{\dot{\phi}_k}(e_1 \circ \phi_k, f_1 \circ \phi_k)$.
2. Compute a $\beta : \mathbb{S}^1 \rightarrow \mathbb{R}$ so that defining

$$(40) \quad (v_e, v_f) = \left(\frac{1}{2}\beta' e_{\phi_k}(1) + \beta e'_{\phi_k}(1), \frac{1}{2}\beta' f_{\phi_k}(1) + \beta f'_{\phi_k}(1) \right)$$

we have that $(e'_{\phi_k}(1) - v_e, f'_{\phi_k}(1) - v_f)$ is horizontal; this β must solve

$$\Omega(e_{\phi_k}(1), e'_{\phi_k}(1) - v_e) + \Omega(f_{\phi_k}(1), f'_{\phi_k}(1) - v_f) = 0,$$

that is,

$$(41) \quad \Omega(e_{\phi_k}(1), v_e) + \Omega(f_{\phi_k}(1), v_f) = \Omega(e_{\phi_k}(1), e'_{\phi_k}(1)) + \Omega(f_{\phi_k}(1), f'_{\phi_k}(1)).$$

Note that (41) simplifies to

$$(42) \quad \frac{1}{2}\beta''(e^2 + f^2) + \beta'(ee' + ff') - \beta((e')^2 - ee'' + (f')^2 - ff'') = \Omega(e, \dot{e}) + \Omega(f, \dot{f}),$$

where we have used a simplified notation $e = e_{\phi_k}(1)$, $f = f_{\phi_k}(1)$, $\dot{e} = \dot{e}_{\phi_k}(1)$, and $\dot{f} = \dot{f}_{\phi_k}(1)$. Note that the discretization of (42) is given in Appendix D.

3. Set $\phi_{k+1} = \phi_k - \varepsilon\beta$, where $\varepsilon > 0$ is small.

Figure 1 illustrates this process.

Remark 3.12. The above algorithm is not guaranteed to converge to the global optimum reparameterization ϕ^* of the geodesic distance in $\mathbf{St}(2, C^\infty)$. In order to help avoid convergence to a local minimum, we perform a direct search for the optimal base point rotation $\phi_0(\theta) = \theta + a$, where $a \in \mathbb{S}^1$, before iterating the above steps.

The geodesic $(e_{\phi_k}(t), f_{\phi_k}(t))$, $t \in [0, 1]$ (in $\mathbf{St}(2, C^\infty)$), for large k will approximate the geodesic in $\mathbf{St}(2, C^\infty)/\text{Diff}(\mathbb{S}^1)$, and hence $\Phi(e_{\phi_k}(t), f_{\phi_k}(t))$, $t \in [0, 1]$, approximates the geodesic in B .

Figure 2 shows an example geodesic in B .

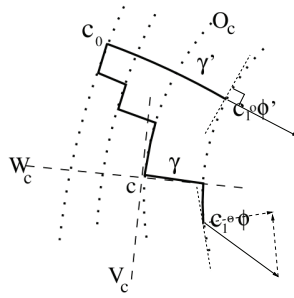


Figure 1. The dashed lines represent equivalence classes of curves (O_c is the orbit $[c]$), and W_c is the space horizontal to c . To compute geodesics in B , we compute the geodesic in M between c_0 and $c_1 \circ \phi$ (the staircase path), project $\dot{\gamma}$ to its vertical component (tangent to $O_{c_1 \circ \phi}$), move c_1 to another representative determined by the vertical component, and iterate the process until the vertical component becomes zero.

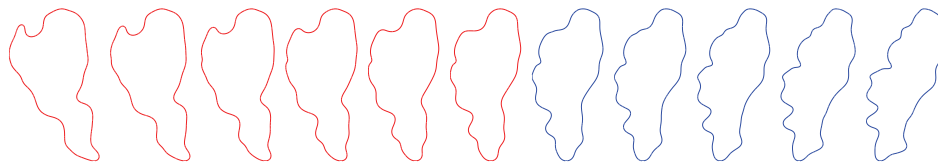


Figure 2. Example illustration of a geodesic in B . The geodesic is computed between the first (leftmost) and the last red curve (time $t = 0$ and $t = 1$), and intermediate curves (interpolation) are shown in between. The blue curves ($t = 1$ to $t = 2$) are the continuation of the geodesic (extrapolation) from the last red curve. It can be seen that the extrapolation does not simply change the pose of the contour but also alters its shape (i.e., it “deforms”).

3.5. Parallel transport. In the next section we will discuss a dynamical model for tracking deforming shapes. To this end, we recall this definition, which is standard in Riemannian geometry.

Definition 3.13. Suppose that M is a Riemannian manifold. Given a path $\gamma : [a, b] \rightarrow M$, the parallel transport

$$P_\gamma : T_{\gamma(a)}M \rightarrow T_{\gamma(b)}M$$

along γ of the tangent vector $h \in T_{\gamma(a)}M$ is defined as

$$P_\gamma(h) = V(b),$$

where the vector field $V(t) \in T_{\gamma(t)}M$ is such that

$$\begin{cases} \nabla_{\dot{\gamma}(t)} V(t) = 0 & \forall t \in [a, b], \\ V(a) = h, \end{cases}$$

and $\nabla_{\dot{\gamma}}$ is the covariant derivative along γ . The parallel transport is a linear isometry between $T_{\gamma(a)}M$ and $T_{\gamma(b)}M$.

The parallel transport in the finite-dimensional Stiefel manifolds $\mathbf{St}(p, \mathbb{R}^n)$ is described in section 2.2.3 of [15]. Given an $n \times p$ orthogonal matrix Y , a tangent vector Δ at Y

is represented as an $n \times p$ real matrix such that $Y^T \Delta$ is skew symmetric. Given a curve $Y = Y(t)$ in $\mathbf{St}(p, \mathbb{R}^n)$, the equation for the parallel transport of Δ is

$$(43) \quad \dot{\Delta} = -Y(\dot{Y}^T \Delta + \Delta^T \dot{Y})/2$$

by (2.16) in [15]. According to [15], no closed form solution is known for this equation, even for the case when Y is a geodesic.

The equation for the parallel transport in $\mathbf{St}(2, \mathbb{L}^2)$ can be easily adapted from the equation above. As for the case of geodesics, we can reduce the problem of computing the parallel transport in $\mathbf{St}(2, \mathbb{L}^2)$ along a geodesic to a computation in a finite-dimensional Stiefel manifold. We first note the following proposition.

Proposition 3.14.

- Suppose that $Y = Y(t)$ is a curve and that v is a vector that is orthogonal to all columns in $Y(t)$; then it is also orthogonal to all columns in $\dot{\Delta}(t)$.
- If Δ is a constant matrix and all its columns are orthogonal to all the columns of $Y(t)$ (for all t), then Δ satisfies (43).

This proposition is stated for the case of $\mathbf{St}(p, \mathbb{R}^n)$ but easily extends to $\mathbf{St}(p, \mathbb{L}^2)$. So we obtain this simplified method to compute the parallel transport in $\mathbf{St}(2, \mathbb{L}^2)$.

Corollary 3.15. Suppose that $(e(t), f(t))$ is a geodesic in $\mathbf{St}(2, \mathbb{L}^2)$. Let \mathcal{B} be the base used in Proposition 3.9, and let $Y(t)$ be the geodesic expressed in this base (note that $Y(t)$ is a geodesic in $\mathbf{St}(2, \mathbb{R}^4)$). Suppose that $(b(t), d(t))$ is the parallel transport of $(b(0), d(0))$ along $(e(t), f(t))$. We decompose both $b(t)$ and $d(t)$ into two components

$$b(t) = \tilde{b}(t) + \hat{b}(t), \quad d(t) = \tilde{d}(t) + \hat{d}(t)$$

with \tilde{b} and \tilde{d} in the 4D space spanned by \mathcal{B} and \hat{b}, \hat{d} orthogonal to this space. From the previous proposition and the uniqueness of solution of the linear first-order system of ODE (43), we obtain that \hat{b}, \hat{d} are constant; if we express \hat{b}, \hat{d} as the two columns of a matrix Δ , using the base \mathcal{B} , then Δ satisfies (43) in $\mathbf{St}(2, \mathbb{R}^4)$.

4. Filtering and prediction for deforming shapes.

4.1. Dynamical model. The geometry in the space of curves described in the previous section gives the foundations for defining dynamical systems on curves. In this section, we show how to construct the simplest possible nontrivial dynamical model. Later, in experiments, we show the usefulness of the model. Defining more complex dynamical models is beyond the scope of this paper, which is to develop the tools for which any dynamical model can be defined.

We start by considering a simple “constant velocity plus perturbation” model for a point moving in \mathbb{R}^n :

$$(44) \quad \begin{aligned} \mu_k &= \mu_{k-1} + \nu_{k-1}, \\ \nu_k &= \nu_{k-1} + \eta_{k-1}, \end{aligned}$$

where the state is $x_k = (\mu_k, \nu_k)$, η_{k-1} is a noise process, and μ represents the position and ν the velocity. When $\{\eta_k\}$ is a white Gaussian process, this is a discrete-time Brownian motion, or first-order random walk.

We assume that we are given noisy measurements of the first component of the state, i.e.,

$$(45) \quad y_k = \mu_k + \xi_k,$$

where ξ_k is the measurement noise.

We now generalize the above dynamical model from \mathbb{R}^n to the case of curves. Denote with $\mu_k \in B$ the *deforming contour* and with $\nu_k \in T_{\mu_k} B$ its *velocity* at time k . The state at time k is $x_k = (\mu_k, \nu_k)$. In the Riemannian manifold B we may define the operation analogous to addition, i.e., $\mu_k + \nu_k$, by using the exponential map. Also, since ν_k and ν_{k-1} are not in the same space (i.e., $\nu_k \in T_{\mu_k} B$ and $\nu_{k-1} \in T_{\mu_{k-1}} B$), the expression $\nu_k = \nu_{k-1} + \eta_{k-1}$ is not defined, and we must transport ν_{k-1} to $T_{\mu_k} B$ via parallel transport.

The “constant velocity plus perturbation” model in the space of curves becomes the following definition.

Definition 4.1 (discrete Brownian motion of curves).

$$(46) \quad \mu_k = \exp_{\mu_{k-1}}(\nu_{k-1}),$$

$$(47) \quad \nu_k = P_{\mu_{k-1}, \mu_k}(\nu_{k-1} + \eta_{k-1}),$$

where $x_k = (\mu_k, \nu_k) \in TB$ is the state, $\eta_{k-1} \in T_{\mu_{k-1}} B$ is a noise process, and P_{μ_{k-1}, μ_k} denotes the parallel transport along the geodesic connecting μ_{k-1} to μ_k . Note that the noise process lives in a linear space, where it is easy to define a Gaussian distribution.

Since the parallel transport is a linear isometry, (47) can be replaced by

$$(48) \quad \nu_k = P_{\mu_{k-1}, \mu_k}(\nu_{k-1}) + \eta_{k-1}$$

simply by choosing $\eta_{k-1} \in T_{\mu_k} B$ directly.

The two models (46), (47) and (46), (48) have equivalent descriptive power. The latter model is, however, easier to implement numerically, since the parallel transport $P_{\mu_{k-1}, \mu_k}(\nu_{k-1})$ is trivial to compute: it is the parallel transport along a geodesic of its own tangent vector, so it is obtained as $\dot{\gamma}(1)$, where γ is the geodesic between $\gamma(0) = \mu_{k-1}$ and $\gamma(1) = \mu_k$; and γ is exactly the geodesic computed in (46).

We will assume that noisy samples of the contour μ_k are available at each time k , for instance, from a segmentation scheme from the active contour literature.

Definition 4.2 (measurement model).

$$(49) \quad y_k = \exp_{\mu_k}(\xi_k),$$

where $\xi_k \in T_{\mu_k} B$ is the measurement noise.

The measurement is a noisy version of the first component of the state, μ_k . Again, notice that ξ_k lives on a linear space, where a Gaussian distribution can be defined easily.

4.2. Filtering deforming shapes. In this section, the goal is to devise a recursive (causal) procedure to estimate the state of the dynamical system, (μ_k, ν_k) , i.e., the shape and velocity of a moving object, introduced in the previous section, from measurements y_k obtained from the time-varying image, I_k . We start by reviewing the classical linear finite-dimensional (Luenberger) observer in \mathbb{R}^n [28] and then generalize it to the space of curves.

An observer in \mathbb{R}^n for the dynamical system (44) and measurement model (45) itself is a dynamical model with state $(\hat{\mu}, \hat{\nu})$ that evolves according to two pairs of equations, the *state prediction*

$$(50) \quad \hat{\mu}_{k|k-1} = \hat{\mu}_{k-1|k-1} + \hat{\nu}_{k-1|k-1},$$

$$(51) \quad \hat{\nu}_{k|k-1} = \hat{\nu}_{k-1|k-1}$$

and the *update*

$$(52) \quad \hat{\mu}_{k|k} = \hat{\mu}_{k|k-1} + K_\mu(y_k - \hat{\mu}_{k|k-1}),$$

$$(53) \quad \hat{\nu}_{k|k} = \hat{\nu}_{k|k-1} + K_\nu(y_k - \hat{\mu}_{k|k-1}).$$

Note how the Luenberger observer structure involves a direct effect of the measurement y_k on the state through two components $K_\mu, K_\nu \geq 0$, which are called the *gains*.⁸ The gains can be chosen to satisfy some optimality criterion. The weakest requirement, for time-invariant models, is that the error

$$e_k = x_k - \hat{x}_{k|k}$$

between the state estimate and the true state approaches zero as $k \rightarrow +\infty$.

The analogous observer in the case of the dynamical system on the space of curves may take the following form.

Definition 4.3 (curve observer). *The prediction is*

$$(54) \quad \hat{\mu}_{k|k-1} = \exp_{\hat{\mu}_{k-1|k-1}}(\hat{\nu}_{k-1|k-1}),$$

$$(55) \quad \hat{\nu}_{k|k-1} = P_{\hat{\mu}_{k-1|k-1}, \hat{\mu}_{k|k-1}}(\hat{\nu}_{k-1|k-1}),$$

where $\hat{\nu}_{k|k-1} \in T_{\hat{\mu}_{k|k-1}}B$, and $P_{\hat{\mu}_{k-1|k-1}, \hat{\mu}_{k|k-1}}$ denotes the parallel transport along the geodesic from $\hat{\mu}_{k-1|k-1}$ to $\hat{\mu}_{k|k-1}$; again $P_{\hat{\mu}_{k-1|k-1}, \hat{\mu}_{k|k-1}}(\hat{\nu}_{k-1|k-1})$ is trivial to compute.

The general form of the update equations may be expressed as

$$(56) \quad \hat{\mu}_{k|k} = \exp_{\hat{\mu}_{k|k-1}}(K_\mu \log(\hat{\mu}_{k|k-1}, y_k)),$$

$$(57) \quad \hat{\nu}_{k|k} = P_{\hat{\mu}_{k|k-1}, \hat{\mu}_{k|k}}(\hat{\nu}_{k|k-1} + K_\nu \log(\hat{\mu}_{k|k-1}, y_k)).$$

To better understand the above equations (and give meaning to \log), we identify the update geodesic ζ_k such that

$$\begin{aligned} \zeta_k(0) &= \hat{\mu}_{k|k-1}, \\ \dot{\zeta}_k(0) &= \log(\hat{\mu}_{k|k-1}, y_k), \\ \zeta_k(1) &= y_k, \\ \zeta_k(K_\mu) &= \hat{\mu}_{k|k}, \end{aligned}$$

⁸In general, the gains can be matrices such as in the Kalman filter, where these matrices are chosen to minimize the expected square error between the state and estimated state when the noise process is chosen to be Gaussian. We choose the simpler case of an isotropic gain that reduces to a scalar, since such a case generalizes easily in the infinite-dimensional case, which we examine next.

where ζ_k in the interval $[0, 1]$ is a minimal geodesic. Since $P_{\hat{\mu}_{k|k-1}, \hat{\mu}_{k|k}}$ is the parallel transport along this geodesic, (57) can be rewritten as

$$(58) \quad \hat{\nu}_{k|k} = P_{\hat{\mu}_{k|k-1}, \hat{\mu}_{k|k}}(\hat{\nu}_{k|k-1}) + K_\nu \dot{\zeta}_k(K_\mu).$$

In any case, the update requires parallel transport of a tangent vector that is not tangent to the geodesic path to be transported along, which entails solving a differential equation numerically. While this is feasible (and not particularly burdensome, as explained in Corollary 3.15), we will consider in this paper only the simplified observer structure where the correction occurs at the velocity level, and therefore $K_\mu = 0$. Then the update takes the form

$$(59) \quad \hat{\mu}_{k|k} = \hat{\mu}_{k|k-1},$$

$$(60) \quad \hat{\nu}_{k|k} = \hat{\nu}_{k|k-1} + K_\nu \log(\hat{\mu}_{k|k-1}, y_k).$$

The constant $K_\nu > 0$ should be chosen to trade off asymptotic tracking error with convergence speed. Ideally, it should be chosen within bounds that guarantee at least stability of the filtering or asymptotic decay of prediction error. In practice, stability can be guaranteed only under additional assumptions on the uncertainty process $\{\eta\}$. In this paper we do not examine this issue.

Remark 4.4. The contour of the state μ represents the boundary of the object of interest in the image, which is typically a simple closed curve. As we remarked in section 3.3.2, for general $\nu \in T_\mu B$, it is not guaranteed that $\exp_\mu(\nu)$ is simple (that is, immersed and non-self-intersecting). However, we have generally observed in the experiments (section 4.4) that, since $\hat{\nu}_{k|k}$ is not chosen arbitrarily (i.e., it is determined by measurements y_k , which are constructed to be simple (see section 4.3)), $\hat{\mu}_{k|k-1}$ is generally simple. Moreover, since we are performing only a one-step prediction (54), which corresponds to following the geodesic for a short time step, the curve generally remains simple. For more than a one-step prediction, a more elaborate model is needed to guarantee that the curve remains simple.

Remark 4.5. We discuss the computational complexity of the above observer for numerical computations. Let the curve μ_k be discretized by N sample points. Computing the exponential map (54) requires first converting to its Stiefel representation (28), computing the basis \mathcal{B} (31), and then applying (30). The parallel transport (55) is automatically computed by the previous formula. Therefore, computing (54), (55) has complexity $O(N)$. Computing log in (60) is more expensive, since it requires an iterative procedure (Algorithm 3.11). Each iteration requires the solution of a tridiagonal system (Appendix D), and this can be done in $O(N)$. Thus, the complexity for (60) is $O(N * P)$, where P is the number of iterations in Algorithm 3.11. Hence, the overall complexity of the curve observer at each time k is $O(N * P)$.

4.3. Obtaining pseudomeasurements via image segmentation. In this section, we describe the procedure to obtain the pseudomeasurements y_k of the state contour μ_k using the image sequence $I_k : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}^+$. We call these pseudomeasurements because what is measured is the irradiant energy impinging on a small area element on the image plane, i.e., the intensity of a pixel. However, for the purpose of this section, we assume that an intermediate process is available that converts these measurements into an ideal, closed, planar simple curve. In practice, implementing such an intermediate process may be rather difficult

when not impossible, depending on the photometric, geometric, and dynamic characteristics of the scene. This issue goes to the heart of much of low-level vision and is well beyond the scope of this paper. Therefore, we will take this assumption with the due caveats, and for the rest of this section we will assume that such pseudomeasurements y_k are available and, for simplicity, call *them* “measurements.” They are the boundary of two disjoint regions obtained by partitioning the image domain Ω into two regions that have distinctive local statistics (e.g., intensity histograms, spatial and temporal regularized derivatives, etc.). In order to find the partitioning boundary, we minimize an energy $E_{ac}(\cdot; I) : B \rightarrow \mathbb{R}$ that depends on an image, I , and is defined on the space of curves, B . The goal of this section is not to show how one constructs such an energy but rather, given any energy, to show how to optimize it via a steepest descent algorithm that is intrinsic to the Riemannian geometry of the space B under the metric \mathbb{H} introduced in section 3. This is different than the Sobolev metric presented in [51, 52], so we present the computation of the gradient of an energy E_{ac} with respect to the metric \mathbb{H} . Indeed, the next proposition shows how one calculates the \mathbb{H} gradient from the usual \mathbb{L}^2 gradient of an energy.

Proposition 4.6. *Let $E_{ac} : M \rightarrow \mathbb{R}$, and suppose that $f \doteq \nabla_{H^0} E_{ac}(c)$ and $g \doteq \nabla_{\mathbb{H}} E_{ac}(c)$ exist. Then $g = g^t + g^l(c - \bar{c}) + L(c)g^d$ is related to f by the following:*

$$\begin{aligned} (61) \quad & g^t = \bar{f}, \\ (62) \quad & g^l = \frac{1}{\lambda_l} \overline{f \cdot (c - \bar{c})}, \\ (63) \quad & g^d = \frac{1}{\lambda_d L(c)} \left(\overline{f \cdot (c - \bar{c})} (c - \bar{c}) - \hat{w} + k \right), \end{aligned}$$

where

$$\begin{aligned} A &\stackrel{\text{def}}{=} (D_s c)(c - \bar{c})^T, & \hat{f}(\sigma) &\stackrel{\text{def}}{=} \int_0^\sigma f(s) \, ds - \sigma \bar{f}, \\ v &\stackrel{\text{def}}{=} \oint_c \hat{f} + A \bar{f} \, ds, & \hat{w}(\sigma) &\stackrel{\text{def}}{=} \int_0^\sigma \hat{f}(s) + A \bar{f} \, ds - \sigma v, \\ k &\stackrel{\text{def}}{=} \oint_c \hat{w} + A^T (\hat{f} + A \bar{f} - v) \, ds \end{aligned}$$

and the above parameterizations are in arc parameter.

Proof. See Appendix C. ■

For the experiments in section 4.4, in order to obtain measurements, we will minimize an energy of the form

$$(64) \quad E_{ac}(c; I) = \int_R F(x; I) \, dx,$$

where R is the region enclosed by the simple curve c , $F : \Omega \rightarrow \mathbb{R}$ incorporates information from the image I , and dx is the area measure in Ω . Such energies have been introduced in a number of papers in the active contour literature (see, e.g., [7, 59, 42, 62]). It can be shown that the H^0 gradient of (64) is

$$(65) \quad \nabla_{H^0} E_{ac}(c; I) = F \mathcal{N},$$

where \mathcal{N} is the outward unit normal to the curve c . The \mathbb{H} gradient of E_{ac} in (64) may be computed from (65) using Proposition 4.6. Therefore, the measurement y_k is obtained by solving the following partial differential equation (PDE):

$$(66) \quad \partial_t C = -\nabla_{\mathbb{H}} E_{ac}(C; I_k),$$

$$(67) \quad C(0, \cdot) = \hat{\mu}_{k|k-1},$$

where $C : \mathbb{R}^+ \times \mathbb{S}^1 \rightarrow \mathbb{R}^2$. That is, the \mathbb{H} gradient descent is run until convergence using the predicted contour $\hat{\mu}_{k|k-1}$ at time k as the initialization.

Although the techniques constructed in this paper are built for tracking a single object, in order to solve the PDE (66) numerically, we use the level set method [40], which naturally allows for topological changes of the underlying curve C . The use of level set methods is beneficial even when the objects of interest do not exhibit topological changes, because a coarse initialization can undergo several topological changes before converging to a simple curve, which affords improved resistance to local minima. This is, of course, based on empirical evidence, as theoretical guarantees for convergence to local minima are hard to come by for the kind of functionals commonly used in image segmentation. When the method converges to multiply connected curves, we choose the component corresponding to a simple curve that has minimal energy according to (64) for the measurement y_k . Instead of evolving the curve C , the level set method evolves a Lipschitz function $\Psi : \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$ such that $\Psi(t, C(t, \cdot)) = 0$; i.e., the zero level set of Ψ is the curve C . The previous equation defines the evolution of Ψ along the level set:

$$\partial_t \Psi(t, x) = -\nabla \Psi(t, x) \cdot G(t, x) \quad \text{for } x \in C(t, \mathbb{S}^1),$$

where G is an extension of $-\nabla_{\mathbb{H}} E_{ac}$ in a narrowband of $C(t, \cdot)$:

$$(68) \quad G(t, C(t, s)) = -\nabla_{\mathbb{H}} E_{ac}(C; I_k)(s) \quad \text{for } s \in [0, L],$$

$$(69) \quad \nabla \Psi(t, x) \cdot \nabla G(t, x) = 0 \quad \text{for } x \notin C(t, \mathbb{S}^1).$$

The second equation above implies that the value of G at a point $x \notin C(t, \mathbb{S}^1)$ is equal to $G(t, C(t, s))$, where $C(t, s)$ is the point on $C(t, \mathbb{S}^1)$ closest to x . Note that this is the case for x with a small narrowband around $C(t, \mathbb{S}^1)$. More details for the numerical implementation can be found in [51].

4.4. Experiments. In this section we aim to illustrate the general qualitative behavior of the dynamical model that we have constructed and the ensuing filter. Therefore, we have chosen a very basic segmentation technique to obtain the measurements y_k by performing an active contour segmentation using the Chan–Vese model [7] and Sobolev active contours [51, 52] with the initialization being the previous state prediction, $\hat{\mu}_{k|k-1}$. At the initial time, the state contour, $\hat{\mu}_{0|0}$, is selected interactively and the state velocity $\hat{\nu}_{0|0}$ is set to zero. Furthermore, we have chosen the gain $K = 0.2$ unless specified otherwise. The red curve in the figures indicates the state prediction contour $\hat{\mu}_{k|k-1}$, the blue arrows indicate the state prediction velocity $\hat{\nu}_{k|k-1}$, and the green curves indicate the measurement y_k , all at frame k .

In the first experiment (Figure 3), we track a circle that continuously deforms (by a nonaffine deformation) into two joined blobs. The data is corrupted by a full occlusion in

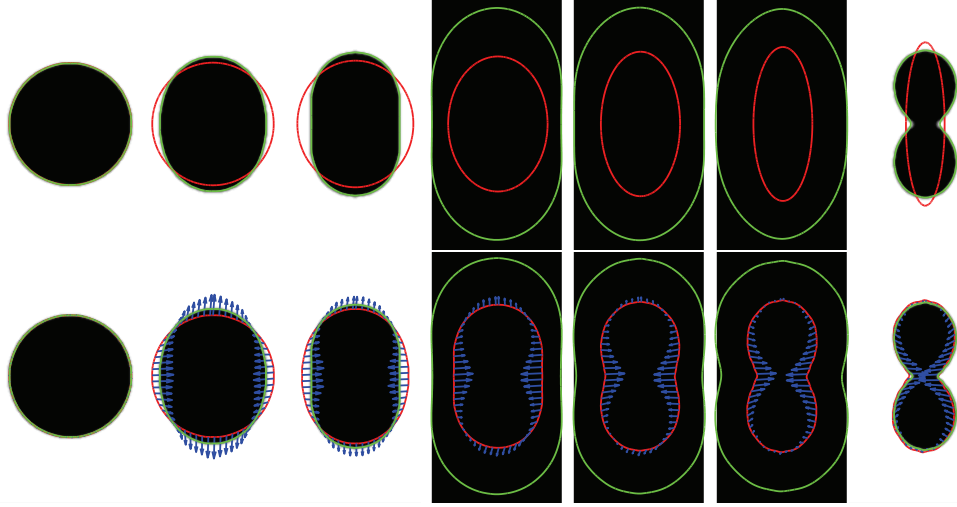


Figure 3. Tracking a synthetic deforming circle through a total occlusion. This experiment demonstrates the need for the dynamical model that extrapolates shape. In the first few frames, where there is no occlusion, the image segmentation (green) alone correctly follows the shape evolution. However, when the occlusion appears, the image segmentation is uninformative, and the dynamical model extrapolates the shape (red) and velocity (blue) of the contour (middle frames). A dynamical model with only affine motion [22] (top row) cannot extrapolate the deformation. The infinite-dimensional model (proposed work), on the other hand, correctly predicts the evolution towards a bilobate shape. Red: $\hat{\mu}_{k|k-1}$; blue: $\hat{\nu}_{k|k-1}$; and green: y_k .

frames 6–11 (the sequence ranges from 1–13). In frame 1, we choose the contour initialization to match the circle’s boundary. In the top row, we have used a dynamical model and filter on the affine motion parameters of the object, as is typical in prior work [22]. In the bottom row, we have used the proposed method, which defines a dynamical model and filters on the space of curves. At the moment of occlusion ($t = 6$), we set the gain $K = 0$, in which case the filter ignores the measurements y_k and moves according to state dynamics for $t \geq 6$ with the initial velocity $\hat{\nu}_{6|6}$. When only affine dynamics are considered, the shape of the object evolves towards an ellipse. The dynamical model on arbitrary deformations, on the other hand, correctly extrapolates during the occlusion and eventually converges to the bilobate shape.

In Figures 4 and 5, we track a deforming flatworm in the ocean. Figure 4 shows the proposed filtering technique applied to the sequence. The experiment demonstrates that the constant velocity plus perturbation model (whose trajectory is shown in red and blue arrows) does a good job at predicting and extrapolating the boundary and motion of the object. In Figure 4, we compare our proposed model to a simple frame-by-frame segmentation (e.g., no filtering in time) [37] and a filtering strategy that filters and models only on the affine motion [22]. As one can see from the figure, the proposed model yields a more accurate track than the affine motion model. In contrast, the affine model predicts a contour that is far enough from the desired local minimum that the measurement “leaks” into portions of the background with similar intensity.

In Figure 6, we track a contracting heart chamber from magnetic resonance imaging (MRI) and compare the results of frame-by-frame segmentation [37], tracking with an affine motion

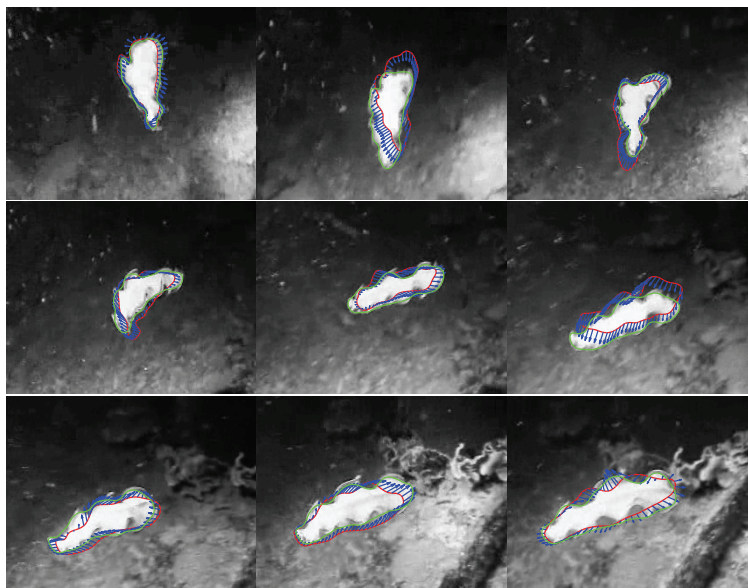


Figure 4. Tracking a flatworm (left to right, top to bottom) using the proposed filtering technique: the red curve is $\hat{\mu}_{k|k-1}$, the blue arrows are $\hat{v}_{k|k-1}$, and the green curve is the measurement y_k . This experiment demonstrates the dynamics of the contour and deformation under the constant velocity plus perturbation model, which correctly models the dynamics of the flatworm.

predictor/estimator [22], and tracking using the proposed prediction/estimation scheme on both shape and motion. As can be seen, a better prediction by the proposed model leads to more accurate measurements that prevent leaking (to a large extent) to the irrelevant chamber. Note that the gradient descent of the image-based energy was run for the same number of fixed iterations for all three tracking procedures.

5. Conclusion. We have introduced a new geometric metric in the space of closed planar curves that decomposes the space into three intuitive components. This decomposition, we have shown, has particular relevance in the tracking problem for computer vision. We have introduced a filtering and prediction scheme on the infinite-dimensional space of shapes, defined as simple, closed planar contours undergoing general diffeomorphisms. Previous work has either attempted to “separate” the “motion” (a finite-dimensional group) from the “deformation” and defined observers for dynamical models on the finite-dimensional motion parameters, or it has restricted the set of allowable deformations to finitely parametrized classes, for instance, obtained from manually obtained training data. The problem with the former approach is that it fails to predict deformations; as an object undergoes an occlusion, the tracker can extrapolate its affine motion but not its deformation. We have shown that predicting deformations allows us to significantly decrease prediction error. The problem with the latter approach is that it requires having training data available for the classes of objects and deformations that one wishes to track. While this is realistic for objects like humans walking, it becomes prohibitive when one wants to consider more gaits (limping, running, hopping) or more objects (flatworms, jellyfish, hurricanes), for which training data may not be available.

Deriving a dynamic observer on the space of curves entails the use of differential and

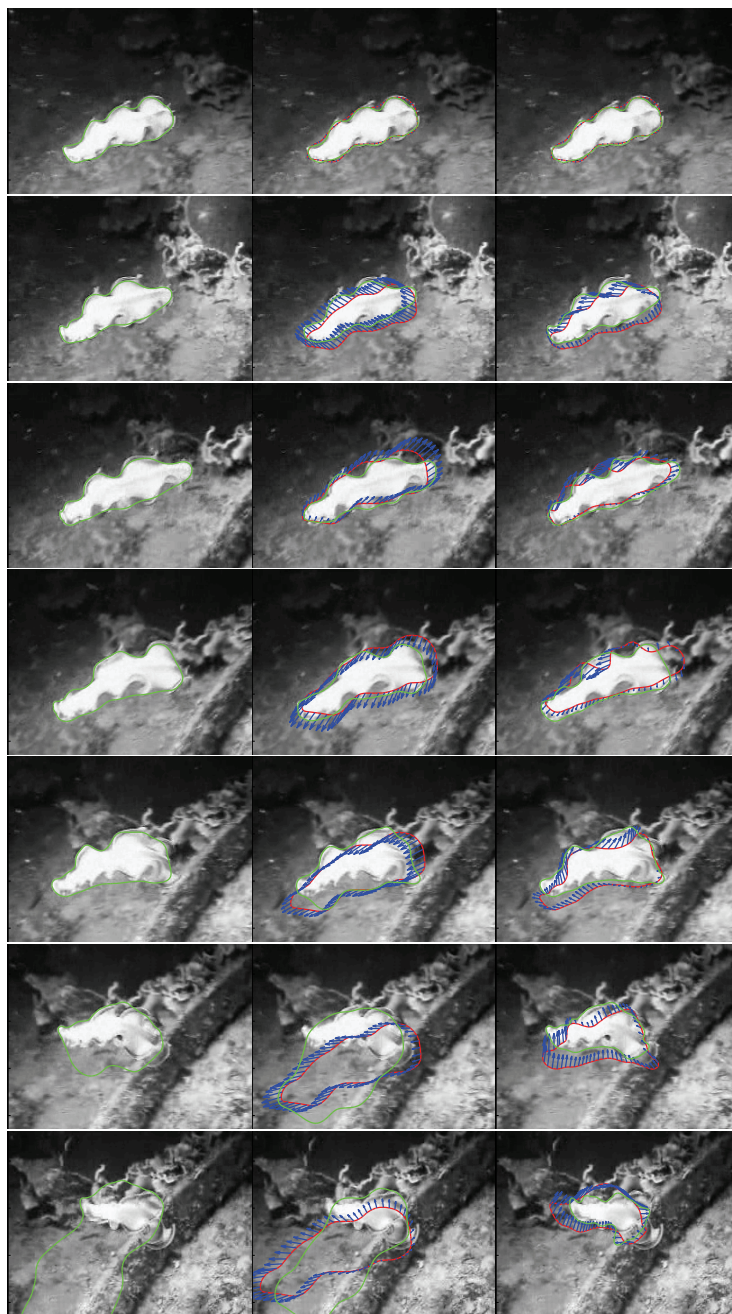


Figure 5. Comparison between frame-by-frame segmentation, i.e., no dynamics [37] (left), dynamical model only on the motion (scales and translation) [22] in the middle column, and dynamical model on both the motion and deformation (proposed work, right column). In the left column, it can be seen that the predicted shape fails to adapt to the newly deformed object; in the right column, where both the motion and deformation are extrapolated, the object is predicted with far greater accuracy. Red: $\hat{\mu}_{k|k-1}$; blue: $\hat{\nu}_{k|k-1}$; and green: y_k .

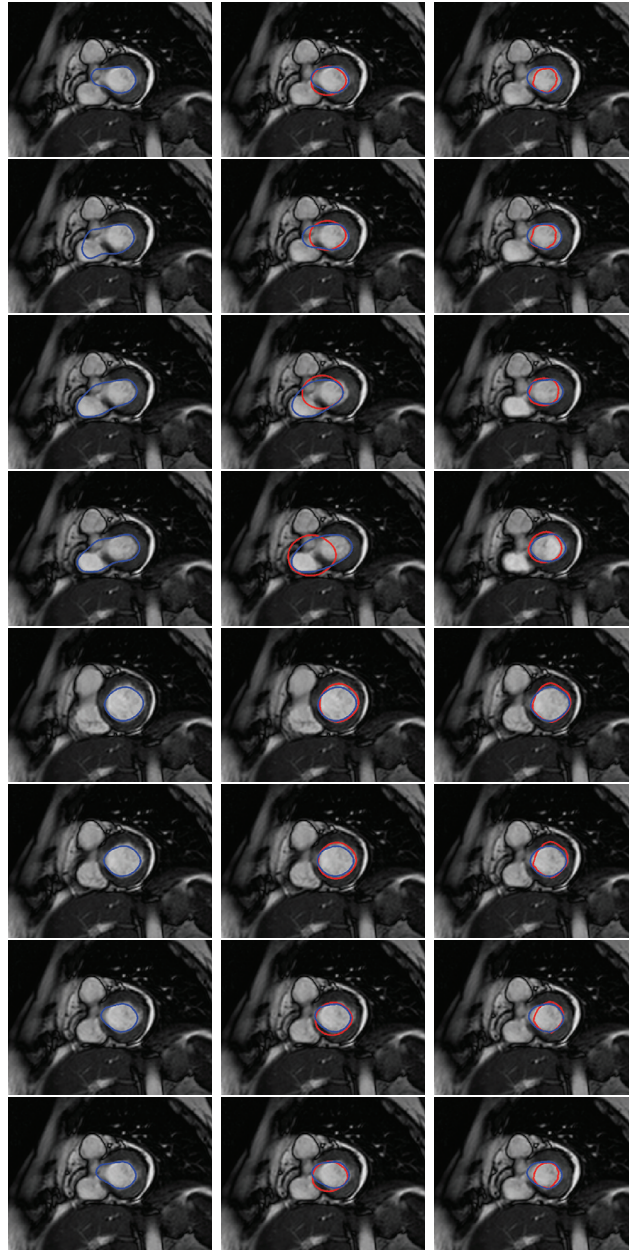


Figure 6. Tracking a ventricle in a contracting heart from MRI. Left column: frame-by-frame segmentation (no dynamics) [37]; middle column: dynamical model only on the motion (scales and translation) [22]; and right column: dynamical model on both the motion and deformation (proposed work). In both the left and middle columns, the contour leaks into an irrelevant chamber. On the right, because the deformation is predicted, the contour is predicted with greater accuracy (although not perfect) and thus results in better measurements (preventing leaking to a large extent). Red: $\hat{\mu}_{k|k-1}$; blue: y_k .

Riemannian geometries and extends classical results in prediction and filtering theory. We have illustrated the case of (first-order) random walk dynamics, but our approach can be easily extended to any linear dynamics, for instance, autoregressive moving-average models of higher order. This is made possible by the fact that the stochastic processes driving the dynamics are defined on the space tangent to the state space, which is linear, and therefore standard tools from systems theory can be applied, albeit with care because these linear spaces are still infinite-dimensional.

While one may wish to bypass the significant mathematical burden by discretizing the objects of interest at the outset, for instance, by using a piecewise linear contour, or a spline or Bezier curve, this introduces difficulties later. In fact, the location of control points or vertices can move while keeping the data unchanged, which results in an unobservable model, and therefore causes spurious dynamics in the observer. Our approach avoids these representational issues by modeling the native objects—closed simple planar contours—directly in the space where they belong, leaving the discretization to the last stage of computation, which is the numerical integration of the PDEs implementing the observer. Our approach has been demonstrated on real and synthetic sequences of deforming objects and shows improvement over the state of the art choice.

Appendix A. Proofs. We first rewrite the metric \mathbb{H} in a form that will be more convenient for the following proofs. Let P_ch be the projector linear operator

$$(70) \quad P_ch \stackrel{\text{def}}{=} h - (D_sc) \oint_c (h \cdot D_sc) ds.$$

Then the three terms of the metric \mathbb{H} may be rewritten as

$$(71) \quad \|h\|_{\mathbb{H}-t}^2 = \left| \oint p(h) ds \right|^2 = |D\bar{c}(c; h)|^2,$$

$$(72) \quad \|h\|_{\mathbb{H}-1}^2 = \left| \oint_c D_sh \cdot D_sc ds \right|^2 = |D(\log(L(c)))(c; h)|^2,$$

$$(73) \quad \begin{aligned} \|h\|_{\mathbb{H}-d}^2 &= L(c)^2 \oint_c |D_sh^d|^2 ds \\ &= \oint_c |P_c(D_sh)|^2 ds = \oint_c |D_sh|^2 ds - \left(\oint_c D_sh \cdot D_sc ds \right)^2, \end{aligned}$$

so that

$$\|h\|_{\mathbb{H}}^2 = \|h\|_{\mathbb{H}-t}^2 + \lambda_t \|h\|_{\mathbb{H}-1}^2 + \lambda_d \|h\|_{\mathbb{H}-d}^2.$$

A.1. Proof of Theorem 3.2. We now prove Theorem 3.2.

Proof. Using Hölder's inequality and the fact that $P_c h$ is a projection operator for the metric H^0 , we obtain the following three inequalities for the three terms that compose \mathbb{H} :

$$\begin{aligned}\|h\|_{\mathbb{H}-t}^2 &= \left| \oint_c h + (c - \bar{c})(D_s h \cdot D_s c) \, ds \right|^2 \\ &\leq 2 \left| \oint_c h \, ds \right|^2 + 2L(c)^2 \oint_c |D_s h|^2 \, ds \leq \left(2 + \frac{2}{\lambda}\right) \|h\|_{\tilde{H}^1}^2, \\ \|h\|_{\mathbb{H}-1}^2 &\leq \oint_c |D_s h|^2 \, ds \leq \frac{1}{\lambda L(c)^2} \|h\|_{\tilde{H}^1}^2, \\ \|h\|_{\mathbb{H}-d}^2 &\leq \oint_c |D_s h|^2 \, ds \leq \frac{1}{\lambda L(c)^2} \|h\|_{\tilde{H}^1}^2.\end{aligned}$$

We then multiply the second by λ_l and the third by λ_d and sum up so that we obtain the inequality

$$\|h\|_{\mathbb{H}}^2 \leq \|h\|_{\tilde{H}^1}^2 \left(\frac{\lambda_d + \lambda_l}{\lambda L(c)^2} + \left(2 + \frac{2}{\lambda}\right) \right)$$

that entails

$$\|h\|_{\mathbb{H}} \leq \|h\|_{\tilde{H}^1} \frac{1}{a_1} \left(\frac{1}{L(c)} + 1 \right)$$

for an appropriate small $a_1 > 0$, from which the leftmost thesis of Theorem 3.2 follows. Conversely,

$$(74) \quad \oint_c |D_s h|^2 \, ds = \oint_c |P_c(D_s h)|^2 \, ds + \left(\oint_c (D_s h \cdot D_s c) \, ds \right)^2 = \|h\|_{\mathbb{H}-d}^2 + \|h\|_{\mathbb{H}-1}^2$$

and

$$(75) \quad \begin{aligned} \left| \oint_c h \, ds \right|^2 &\leq 2 \left| \oint_c h + (c - \bar{c})(D_s h \cdot D_s c) \, ds \right|^2 + 2 \left| \oint_c (c - \bar{c})(D_s h \cdot D_s c) \, ds \right|^2 \\ &\leq 2 \|h\|_{\mathbb{H}-t}^2 + 2L(c)^2 \|h\|_{\mathbb{H}-1}^2, \end{aligned}$$

so multiplying (74) by $\lambda L(c)^2$ and summing with (75) yields

$$\|h\|_{\tilde{H}^1}^2 \leq \lambda L(c)^2 \|h\|_{\mathbb{H}-d}^2 + (\lambda + 2)L(c)^2 \|h\|_{\mathbb{H}-1}^2 + 2\|h\|_{\mathbb{H}-t}^2;$$

hence

$$\|h\|_{\tilde{H}^1} \leq \|h\|_{\mathbb{H}} \left(\frac{\lambda L(c)^2}{\lambda_d} + \frac{(\lambda + 2)L(c)^2}{\lambda_l} + 2 \right)^{1/2}. \quad \blacksquare$$

A.2. Proof of Theorem 3.4.

We now prove Theorem 3.4.

Proof. The tangent map of (18) is

$$(76) \quad \begin{aligned} \mathbb{R}^2 \times \mathbb{R} \times T_{\bar{c}} M_d &\rightarrow T_c M, \\ h^t, h^l, h^d &\mapsto h = h^t + h^l e^l \tilde{c} + e^l h^d \\ &= h^t + h^l(c - \bar{c}) + L(c)h^d; \end{aligned}$$

and the tangent to the inverse (19) is

$$\begin{aligned}
 T_c M &\rightarrow \mathbb{R}^2 \times \mathbb{R} \times T_{\bar{c}} M_d, \\
 (77) \quad h &\mapsto h^t = \oint_c h + (c - \bar{c})(D_s h \cdot D_s c) \, ds, \\
 h^l &= \oint_c (D_s h \cdot D_s c) \, ds, \\
 h^d &= L(c)^{-1}(h - h^t - (c - \bar{c})h^l).
 \end{aligned}$$

We now use the equalities shown at the beginning of the appendix, and the definitions (10) and (11), and write

$$\begin{aligned}
 \|h\|_{\mathbb{H}-t}^2 &= \left| \oint_c h + (c - \bar{c})(D_s h \cdot D_s c) \, ds \right|^2 = |h^t|^2, \\
 \|h\|_{\mathbb{H}-1}^2 &= \left(\oint_c (D_s h \cdot D_s c) \, ds \right)^2 = (h^l)^2.
 \end{aligned}$$

We now carefully recall that, in the definition of the metric (17) on M_d , we are using arc parameter derivatives with respect to the curve \tilde{c} and not the curve c ; so

$$\begin{aligned}
 (D_s h^d) &= \frac{\frac{\partial}{\partial \theta} h^d}{\left| \frac{\partial}{\partial \theta} \tilde{c} \right|} = \frac{\frac{\partial}{\partial \theta} (h - h^t - (c - \bar{c})h^l)}{\left| \frac{\partial}{\partial \theta} c \right|} \\
 (78) \quad &= \frac{\frac{\partial}{\partial \theta} h - \frac{\partial}{\partial \theta} c h^l}{\left| \frac{\partial}{\partial \theta} c \right|} = D_s h - D_s c (\overline{D_s h \cdot D_s c}) = P_c(D_s h)
 \end{aligned}$$

(where P was defined in (70)), so that

$$\|h^d\|_{M_d}^2 = \int_{\tilde{c}} |D_s h^d|^2 \, ds = \oint_c |P_c(D_s h)|^2 \, ds = \|h\|_{\mathbb{H}-d}^2.$$

We conclude that (18) is an isometry. \blacksquare

A.3. Proof of Proposition 3.6. We now prove Proposition 3.6. The following proof is based on classical methods, first applied to Riemannian geometries of immersed curves in [35].

Proof. Let \mathcal{G} be a group that acts on M . Given a curve $c \in M$ and a tangent vector $\xi \in T_e \mathcal{G}$ (where e is the identity in \mathcal{G} , and $T_e \mathcal{G}$ is the *Lie algebra* of \mathcal{G}), we derive the action for fixed c and e “moving” in direction ξ ; the result of this derivative is a tangent vector $\zeta = \zeta_{\xi, c} \in T_c M$ (depending linearly on ξ). By the Emmy Noether theorem, if the metric is invariant with respect to the action of \mathcal{G} , and if $\gamma(t)$ is a geodesic, then

$$(79) \quad \left\langle \zeta_{\xi, \gamma(t)}, \dot{\gamma}(t) \right\rangle_{\gamma(t)}$$

is constant in t for any choice of $\xi \in T_e \mathcal{G}$.

- For the translation group, $\zeta = \xi \in \mathbb{R}^2$, by (71),

$$\left\langle \xi, \frac{\partial}{\partial t} C \right\rangle_{\mathbb{H}} = \xi \cdot \frac{\partial}{\partial t} \overline{C}$$

is constant for any ξ ; hence $\frac{\partial}{\partial t} \overline{C}$ is constant. Alternatively, we can use the isometry in Theorem 3.4.

- The *rotation* group is represented by orthonormal matrices; the tangent $T_e \mathcal{G}$ is the set of the antisymmetric matrices $B \in \mathbb{R}^{2 \times 2}$; then $\zeta = BC$. We compute

$$\begin{aligned} \left\langle BC, \frac{\partial}{\partial t} C \right\rangle_{\mathbb{H-t}} &= B \overline{C} \cdot \left(\frac{\partial}{\partial t} \overline{C} \right), \\ \left\langle BC, \frac{\partial}{\partial t} C \right\rangle_{\mathbb{H-l}} &= 0, \\ \left\langle BC, \frac{\partial}{\partial t} C \right\rangle_{\mathbb{H-d}} &= \int_C (BD_s C) \cdot \left(D_s \frac{\partial}{\partial t} C \right) ds. \end{aligned}$$

We also know that $\frac{\partial}{\partial t} \overline{C}$ is a constant and call it v ; so

$$(B \overline{C}) \cdot \left(\frac{\partial}{\partial t} \overline{C} \right) = (B \overline{C}_0 + t B v) \cdot v = (B \overline{C}_0) \cdot v,$$

which is constant. By direct computation,

$$\frac{\partial}{\partial t} (D_s C) = - \left(D_s \frac{\partial}{\partial t} C \cdot D_s C \right) (D_s C) + D_s \left(\frac{\partial}{\partial t} C \right),$$

so we conclude that

$$\frac{1}{\lambda_d} \left\langle BC, \frac{\partial}{\partial t} C \right\rangle_{\mathbb{H}} = \int_C (BD_s C) \cdot \left(D_s \frac{\partial}{\partial t} C \right) ds = \int_C (BD_s C) \cdot \left(\frac{\partial}{\partial t} D_s C \right) ds$$

and that these are constant in t .

- The *reparameterization* group is $\mathcal{G} = \text{Diff}(S^1)$; the action is the composition $\phi, c \mapsto c \circ \phi$; a tangent vector in $T_e \mathcal{G}$ is a scalar field $\xi : S^1 \rightarrow \mathbb{R}$; we in the end have that

$$\zeta(\theta) = \xi(\theta) c'(\theta)$$

(where $c' = \frac{\partial}{\partial \theta} c$) or

$$\zeta(\theta) = f(\theta) D_s c(\theta);$$

that is, ζ is a generic vector field parallel to the curve. We compute

$$\left\langle f D_s C, \frac{\partial}{\partial t} C \right\rangle_{\mathbb{H-t}} = 0, \quad \left\langle f D_s C, \frac{\partial}{\partial t} C \right\rangle_{\mathbb{H-l}} = 0$$

from (71) and (72), while

$$\begin{aligned}
 \left\langle f D_s C, \frac{\partial}{\partial t} C \right\rangle_{\mathbb{H}\text{-d}} &= \oint_C (D_s(f D_s C)) \cdot \left(D_s \frac{\partial}{\partial t} C \right) ds \\
 &\quad - \oint (D_s(f D_s C)) \cdot D_s C \, ds \oint (D_s C) \cdot \left(D_s \frac{\partial}{\partial t} C \right) ds \\
 &= - \oint_C (f D_s C) \cdot \left(D_s^2 \frac{\partial}{\partial t} C \right) ds \\
 &\quad - \oint (f D_s C) \cdot D_s^2 C \, ds \oint (D_s^2 C) \cdot \left(\frac{\partial}{\partial t} C \right) ds \\
 &= - \oint_C f(D_s C) \cdot \left(D_s^2 \frac{\partial}{\partial t} C \right) ds,
 \end{aligned}$$

since $(D_s C) \cdot (D_s^2 C) = 0$.

For the rescaling group, we cannot use Emmy Noether's theorem directly, since the metric \mathbb{H} is not rescaling-invariant as a whole. The result is then a consequence of Theorem 3.4. ■

Appendix B. Fast algorithm for matrix exponential derivatives. In this appendix, we derive a fast method for computing

$$D(\exp X)(X; Z) = \int_0^1 \exp(tX) Z \exp((1-t)X) dt.$$

Since $D(\exp X)(X; \cdot)$ is a linear operator acting on Z , we have

$$D(\exp X)(X; Z) = \sum_{ij} z_{ij} \int_0^1 \exp(tX) \Delta_{ij} \exp((1-t)X) dt,$$

where $z_{i,j}$ are the components of the matrix Z , and Δ_{ij} is the $n \times n$ -matrix of components

$$(\Delta_{ij})_{kl} = \begin{cases} 1, & (k, l) = (i, j), \\ 0, & (k, l) \neq (i, j). \end{cases}$$

Noting that $B \Delta_{ij} C = B_{\cdot, i} C_{j, \cdot}$, where $B_{\cdot, i}$ (column vector) denotes column i of B , and $C_{j, \cdot}$ (row vector) denotes row j of C , we have that

$$D(\exp X)(X; Z) = \sum_{ij} z_{ij} \int_0^1 (\exp(tX))_{\cdot, i} (\exp((1-t)X))_{j, \cdot} dt.$$

Note that $(B_{\cdot, i} C_{j, \cdot})_{kl} = (B \otimes C)_{j+n(k-1), (i-1)n+l}$, where $i, j, k, l \in \{1, \dots, n\}$, n is the dimension of X , and \otimes denotes the Kronecker product. Therefore,

$$(80) \quad (D(\exp X)(X; Z))_{kl} = \sum_{ij} z_{ij} \left(\int_0^1 \exp(tX) \otimes \exp((1-t)X) dt \right)_{j+n(k-1), (i-1)n+l}.$$

We now give a fast method to compute the integral in (80). Note that

$$\begin{aligned} \int_0^1 \exp(tX) \otimes \exp((1-t)X) dt &= \int_0^{\frac{1}{2}} \exp(tX) \otimes [\exp((1/2-t)X) \exp(X/2)] dt \\ &\quad + \int_{\frac{1}{2}}^1 [\exp(X/2) \exp((t-1/2)X)] \otimes \exp((1-t)X) dt \\ &= \mathcal{I}_{1/2}(\text{Id}_{n \times n} \otimes \exp(X/2)) + (\exp(X/2) \otimes \text{Id}_{n \times n}) \mathcal{I}_{1/2}, \end{aligned}$$

where we define

$$(81) \quad \mathcal{I}_{1/2^k} = \int_0^{\frac{1}{2^k}} \exp(tX) \otimes \exp((1/2^k - t)X) dt.$$

Analogous to the computation above, we find that

$$(82) \quad \mathcal{I}_{1/2^k} = \mathcal{I}_{1/2^{k+1}}(\text{Id}_{n \times n} \otimes \exp(X/2^{k+1})) + (\exp(X/2^{k+1}) \otimes \text{Id}_{n \times n}) \mathcal{I}_{1/2^{k+1}}.$$

Therefore, given an integer m , we may compute \mathcal{I}_1 recursively using the formula (82) m times, and $\mathcal{I}_{1/2^m}$ can be approximated using a Riemannian sum or any other numerical integration method.

The technique above resembles the fast Fourier transform algorithm and is similar to a technique found in [55].

Appendix C. Computing the \mathbb{H} gradient of energies. In this appendix, we compute the gradient of an energy $E_{ac} : M \rightarrow \mathbb{R}$ with respect to the \mathbb{H} metric in terms of the H^0 metric. By definition of gradient, we have that

$$(83) \quad DE_{ac}(c; h) = \langle h, \nabla_{\mathbb{H}} E \rangle_{\mathbb{H}} = \langle h, \nabla_{H^0} E \rangle_{H^0} \quad \forall h \in T_c M.$$

Therefore, we solve

$$\oint_c h \cdot f ds = h^t \cdot g^t + \lambda_l h^l g^l + \lambda_d L(c)^2 \oint_c D_s h^d \cdot D_s g^d ds,$$

where $f \stackrel{\text{def}}{=} \nabla_{H^0} E$ and $g \stackrel{\text{def}}{=} \nabla_{\mathbb{H}} E$. We set for convenience

$$A = (D_s c)(c - \bar{c})^T, \text{ and}$$

$A = A(s)$ is a 2×2 matrix field along c ; note that

$$h^t = D(\bar{c})(c; h) = \oint h - (D_s A^T) h ds.$$

One can show that

$$\begin{aligned} h^t \cdot g^t &= \oint_c h \cdot [g^t - (D_s A) g^t] ds, \\ h^l g^l &= - \oint_c h \cdot (g^l D_s^2 c) ds, \\ \oint_c D_s h^d \cdot D_s g^d ds &= - \frac{1}{L(c)} \oint_c h \cdot D_s^2 g^d ds. \end{aligned}$$

So we must solve for g (i.e., the components g^t, g^l , and g^d of g) in

$$f = [\text{Id} - D_s A] g^t - \lambda_l g^l D_s^2 c - \lambda_d L(c) D_s^2 g^d.$$

It is clear that $g^t = \bar{f}$, and noting that g^d does not change the length we have that

$$(84) \quad \overline{f \cdot (c - \bar{c})} = -\overline{[(D_s A) g^t] \cdot (c - \bar{c})} + \lambda_l g^l = \lambda_l g^l,$$

which verifies (62). Eventually we obtain the equation

$$(85) \quad -\lambda_l g^l D_s^2 c - (f - \bar{f}) - (D_s A) \bar{f} = \lambda_d L(c) D_s^2 g^d.$$

We note that the left-hand side (LHS) has zero integral along the curve, so we integrate both sides to obtain

$$(86) \quad -\lambda_l g^l D_s c - \hat{f} - A \bar{f} + v = \lambda_d L(c) D_s g^d,$$

where $\hat{f}(\sigma) \stackrel{\text{def}}{=} \int_0^\sigma (f(s) - \bar{f}) ds$; the above identity is determined up to the constant $v \in \mathbb{R}^2$. Since the right-hand side (RHS) of (86) has zero average on the curve, this forces the choice of v , which will be

$$(87) \quad v \stackrel{\text{def}}{=} \oint_c \hat{f} + A \bar{f} ds.$$

We then let $w \stackrel{\text{def}}{=} \hat{f} + A \bar{f} - v$ and rewrite (86) as

$$(88) \quad -\lambda_l g^l D_s c - w = \lambda_d L(c) D_s g^d,$$

where both the LHS and RHS have zero average.

To solve (88) for g^d , we now integrate once again and obtain

$$(89) \quad -\lambda_l g^l (c - \bar{c}) - \hat{w} + k = \lambda_d L(c) g^d,$$

where $\hat{w}(\sigma) \stackrel{\text{def}}{=} \int_0^\sigma w(s) ds$, and once again the identity is determined up to the constant $k \in \mathbb{R}^2$. The constant k is obtained by imposing that $(g^d)^t = 0$; that is, g^d does not move the centroid of the curve.⁹

Appendix D. Discretization of (42). We discretize (42) to solve for the vertical component of $(\delta e, \delta f) \in T_{(e,f)} \mathbf{St}(2, C^\infty)$. Let $e_i, f_i, \delta e_i, \delta f_i$ ($1 \leq i \leq N$) be uniform samplings of $e, f, \delta e, \delta f : \mathbb{S}^1 \rightarrow \mathbb{R}$. Set

$$\begin{aligned} e_0 &= \pm e_N, & f_0 &= \pm f_N, & \delta e_0 &= \pm \delta e_N, & \delta f_0 &= \pm \delta f_N, \\ e_{N+1} &= \pm e_1, & f_{N+1} &= \pm f_1, & \delta e_{N+1} &= \pm \delta e_1, & \delta f_{N+1} &= \pm \delta f_1; \end{aligned}$$

⁹The other identity $(g^d)^l = \int_c D_s g^d \cdot D^s c ds = 0$ is true regardless of the choice of k and v and is verified by the identity (86).

one chooses $+$ or $-$ if the curve is of even or odd winding number, respectively. Define b_i, c_i, d_i, g_i ($1 \leq i \leq N$) as

$$\begin{aligned} b_i &= -\frac{1}{2}[\delta e_i(e_{i+1} - e_{i-1}) - e_i(\delta e_{i+1} - \delta e_{i-1}) + \delta f_i(f_{i+1} - f_{i-1}) - f_i(\delta f_{i+1} - \delta f_{i-1})], \\ c_i &= \frac{1}{2}(e_i^2 + f_i^2), \\ d_i &= \frac{1}{2}[e_i(e_{i+1} - e_{i-1}) + f_i(f_{i+1} - f_{i-1})], \\ g_i &= \frac{1}{2}\left[e_i(e_{i+1} - 2e_i + e_{i-1}) - \frac{1}{2}(e_{i+1} - e_{i-1})^2 \right. \\ &\quad \left. + f_i(f_{i+1} - 2f_i + f_{i-1}) - \frac{1}{2}(f_{i+1} - f_{i-1})^2\right]. \end{aligned}$$

Let $A = (a_{ij})$, where $1 \leq i, j \leq N$. Set, for $1 \leq i, j \leq N$,

$$a_{ii} = -c_i + g_i, \quad a_{i(i-1)} = c_i - d_i, \quad a_{i(i+1)} = c_i + d_i, \quad a_{ij} = 0 \quad \text{for } j-1 > i > j+1,$$

and $B = (b_i)$. Note that $a_{i0} := a_{iN}$, $a_{i(N+1)} := a_{i0}$. Then, if $x = (\beta_i)$ is a sampling of β in (42), $Ax = B$.

Acknowledgment. The authors wish to thank Stefano Deiala for his very detailed reading of the manuscript during the preparation of the final draft. His corrections, comments, and many valuable suggestions are greatly appreciated.

REFERENCES

- [1] M. F. BEG, M. I. MILLER, A. TROUVÉ, AND L. YOUNES, *Computing large deformation metric mappings via geodesic flows of diffeomorphisms*, Int. J. Comput. Vis., 61 (2005), pp. 139–157.
- [2] A. BLAKE AND R. BROCKETT, *On snakes and estimation theory*, in Proceedings of the 33rd IEEE Conference on Decision and Control, 1994.
- [3] A. BRONSTEIN, M. BRONSTEIN, AND R. KIMMEL, *Calculus of non-rigid surfaces for geometry and texture manipulation*, IEEE Trans. Visual Comput. Graphics, 13 (2007), pp. 902–913.
- [4] V. CASELLES, F. CATTE, T. COLL, AND F. DIBOS, *A geometric model for edge detection*, Numer. Math., 66 (1993), pp. 1–31.
- [5] V. CASELLES, R. KIMMEL, AND G. SAPIRO, *Geodesic active contours*, in Proceedings of the 5th IEEE International Conference on Computer Vision, Cambridge, MA, 1995, pp. 694–699.
- [6] V. CERVERA, F. MASCARÓ, AND P. W. MICHOR, *The action of the diffeomorphism group on the space of immersions*, Differential Geom. Appl., 1 (1991), pp. 391–401.
- [7] T. CHAN AND L. VESE, *Active contours without edges*, IEEE Trans. Image Process., 10 (2001), pp. 266–277.
- [8] G. CHARPIAT, O. D. FAUGERAS, AND R. KERIVEN, *Approximations of shape metrics and application to shape warping and empirical shape statistics*, Found. Comput. Math., 5 (2005), pp. 1–58.
- [9] G. CHARPIAT, R. KERIVEN, J. PONS, AND O. FAUGERAS, *Designing spatially coherent minimizing flows for variational problems based on active contours*, in Proceedings of the 10th IEEE International Conference on Computer Vision, 2005, pp. 1403–1408.
- [10] G. CHARPIAT, P. MAUREL, J.-P. PONS, R. KERIVEN, AND O. FAUGERAS, *Generalized gradients: Priors on minimization flows*, Int. J. Comput. Vis., 73 (2007), pp. 325–344.
- [11] D. CREMERS, *Nonlinear dynamical shape priors for level set segmentation*, J. Sci. Comput., 35 (2008), pp. 132–143.

- [12] D. CREMERS AND S. SOATTO, *Motion competition: A variational approach to piecewise parametric motion segmentation*, Int. J. Comput. Vis., 62 (2004), pp. 249–265.
- [13] M. DELFOUR AND J. P. ZOLÉSIO, *Shape identification via metrics constructed from the oriented distance function*, Control Cybernet., 34 (2005), pp. 137–164.
- [14] M. DO CARMO, *Riemannian Geometry*, Birkhäuser Boston, Boston, MA, 1992.
- [15] A. EDELMAN, T. A. ARIAS, AND S. T. SMITH, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 303–353.
- [16] R. GOLDENBERG, R. KIMMEL, E. RIVLIN, AND M. RUDZSKY, *Fast geodesic active contours*, IEEE Trans. Image Process., 10 (2001), pp. 1467–1475.
- [17] C. GOSCH, K. FUNDANA, A. HEYDEN, AND C. SCHNÖRR, *View point tracking of rigid objects based on shape sub-manifolds*, in Proceedings of the 10th European Conference on Computer Vision, Vol. 3, 2008, pp. 251–263.
- [18] J. GRABOWSKI, *Derivative of the exponential mapping for infinite-dimensional Lie groups*, Ann. Global Anal. Geom., 11 (1993), pp. 213–220.
- [19] R. S. HAMILTON, *The inverse function theorem of Nash and Moser*, Bull. Amer. Math. Soc. (N.S.), 7 (1982), pp. 65–222.
- [20] B. K. P. HORN AND B. G. SCHUNCK, *Determining optical flow*, Artif. Intell., 17 (1981), pp. 185–203.
- [21] M. ISARD AND A. BLAKE, *CONDENSATION—Conditional density propagation for visual tracking*, Int. J. Comput. Vis., 29 (1998), pp. 5–28.
- [22] J. JACKSON, A. YEZZI, AND S. SOATTO, *Tracking deformable moving objects under severe occlusions*, in Proceedings of the 43rd IEEE Conference on Decision and Control, 2004.
- [23] M. KASS, A. WITKIN, AND D. TERZOPOULOS, *Snakes: Active contour models*, Int. J. Comput. Vis., 1 (1987), pp. 321–331.
- [24] D. G. KENDALL, *Shape manifolds, Procrustean metrics, and complex projective spaces*, Bull. London Math. Soc., 16 (1984), pp. 81–121.
- [25] S. KICHENASSAMY, A. KUMAR, P. OLIVER, A. TANNENBAUM, AND A. YEZZI, *Gradient flows and geometric active contour models*, in Proceedings of the 5th IEEE International Conference on Computer Vision, Cambridge, MA, 1995, pp. 810–815.
- [26] M. KILIAN, N. MITRA, AND H. POTTSMANN, *Geometric modeling in shape space*, ACM Trans. Graphics, 26 (2007), article 64.
- [27] S. LANG, *Fundamentals of Differential Geometry*, Springer-Verlag, New York, 1999.
- [28] D. LUENBERGER, *Observing the state of a linear system*, IEEE Trans. Military Electron., 8 (1964), pp. 74–80.
- [29] R. MALLADI, J. SETHIAN, AND B. VEMURI, *Shape modeling with front propagation: A level set approach*, IEEE Trans. Pattern Anal. Mach. Intell., 17 (1995), pp. 158–175.
- [30] K. V. MARDIA AND I. L. DRYDEN, *Shape distributions for landmark data*, Adv. in Appl. Prob., 21 (1989), pp. 742–755.
- [31] R. MATHIAS, *Evaluating the Fréchet derivative of the matrix exponential*, Numer. Math., 63 (1992), pp. 213–226.
- [32] A. MENNUCCI, A. YEZZI, AND G. SUNDARAMOORTHY, *Properties of Sobolev type metrics in the space of curves*, Interfaces Free Bound., 10 (2008), pp. 423–445.
- [33] P. MICHOR AND D. MUMFORD, *Riemannian geometries on the space of plane curves*, J. Eur. Math. Soc. (JEMS), 8 (2006), pp. 1–48.
- [34] P. W. MICHOR AND D. MUMFORD, *Vanishing geodesic distance on spaces of submanifolds and diffeomorphisms*, Doc. Math., 10 (2005), pp. 217–245.
- [35] P. W. MICHOR AND D. MUMFORD, *An overview of the Riemannian metrics on spaces of curves using the Hamiltonian approach*, Appl. Comput. Harmon. Anal., 23 (2007), pp. 76–113.
- [36] W. MIO AND A. SRIVASTAVA, *Elastic-string models for representation and analysis of planar shapes*, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, 2004, pp. 10–15.
- [37] M. MOELICH AND T. CHAN, *Tracking Objects with the Chan-Vese Algorithm*, Technical report, UCLA, Los Angeles, CA, 2003.
- [38] D. MUMFORD AND J. SHAH, *Optimal approximations by piecewise smooth functions and associated variational problems*, Comm. Pure Appl. Math., 42 (1989), pp. 577–685.

- [39] M. NIETHAMMER, P. A. VELA, AND A. TANNENBAUM, *Geometric observers for dynamically evolving curves*, IEEE Trans. Pattern Anal. Mach. Intell., 30 (2008), pp. 1093–1108.
- [40] S. OSHER AND J. SETHIAN, *Fronts propagating with curvature-dependent speed: Algorithms based on the Hamilton-Jacobi equations*, J. Comput. Phys., 79 (1988), pp. 12–49.
- [41] N. PAPADAKIS AND E. MÉMIN, *A variational technique for time consistent tracking of curves and motion*, J. Math. Imaging Vision, 31 (2008), pp. 81–103.
- [42] N. PARAGIOS AND R. DERICHE, *Geodesic active regions: A new paradigm to deal with frame partition problems in computer vision*, J. Vis. Comm. Image Represent., 13 (2002), pp. 249–268.
- [43] N. PETERFREUND, *The velocity snake: Deformable contour for tracking in spatio-velocity space*, Comput. Vis. Image Understand., 73 (1999), pp. 346–356.
- [44] Y. RATHI, N. VASWANI, A. TANNENBAUM, AND A. YEZZI, *Particle filtering for geometric active contours and application to tracking deforming objects*, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, pp. 2–9.
- [45] R. RONFARD, *Region based strategies for active contour models*, Int. J. Comput. Vis., 13 (1994), pp. 229–251.
- [46] M. RUMPF AND B. WIRTH, *A nonlinear elastic shape averaging approach*, SIAM J. Imaging Sci., 2 (2009), pp. 800–833.
- [47] S. SOATTO AND A. J. YEZZI, *DEFORMATION: Deforming motion, shape average and the joint registration and segmentation of images*, in Proceedings of the 7th European Conference on Computer Vision, Vol. 3, 2002, pp. 32–57.
- [48] G. SUNDARAMOORTHY, A. MENNUCCI, S. SOATTO, AND A. YEZZI, *Tracking deforming objects by filtering and prediction in the space of curves*, in Proceedings of the 48th IEEE Conference on Decision and Control, 2009, pp. 2395–2401.
- [49] G. SUNDARAMOORTHY, A. YEZZI, AND A. MENNUCCI, *Sobolev active contours*, in Variational, Geometric, and Level Set Methods in Computer Vision, Springer-Verlag, Berlin, 2005, pp. 109–120.
- [50] G. SUNDARAMOORTHY, A. YEZZI, A. MENNUCCI, AND G. SAPIRO, *New possibilities with Sobolev active contours*, Int. J. Comput. Vis., 84 (2009), pp. 113–129.
- [51] G. SUNDARAMOORTHY, A. J. YEZZI, AND A. MENNUCCI, *Sobolev active contours*, Int. J. Comput. Vis., 73 (2007), pp. 345–366.
- [52] G. SUNDARAMOORTHY, A. J. YEZZI, AND A. MENNUCCI, *Coarse-to-fine segmentation and tracking using Sobolev active contours*, IEEE Trans. Pattern Anal. Mach. Intell., 30 (2008), pp. 851–864.
- [53] D. TERZOPOULOS AND R. SZELISKI, *Tracking with Kalman snakes*, in Active Vision, MIT Press, Cambridge, MA, 1992.
- [54] N. VASWANI, Y. RATHI, A. YEZZI, AND A. TANNENBAUM, *Deform PF-MT: Particle filter with mode tracker for tracking nonaffine contour deformations*, IEEE Trans. Image Process., 19 (2010), pp. 841–857.
- [55] C. WASSHUBER, *Computational Single-Electronics*, Springer-Verlag, Berlin, 2001.
- [56] B. WIRTH, L. BAR, M. RUMPF, AND G. SAPIRO, *Geodesics in shape space via variational time discretization*, in Proceedings of the 7th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition, 2009, pp. 288–302.
- [57] A. YEZZI AND A. MENNUCCI, *Metrics in the Space of Curves*, <http://arxiv.org/abs/arXiv:math.DG/0412454> (2004).
- [58] A. YEZZI AND A. MENNUCCI, *Conformal metrics and true “gradient flows” for curves*, in Proceedings of the 10th International Conference on Computer Vision, 2005, pp. 913–919.
- [59] A. YEZZI, A. TSAI, AND A. WILLSKY, *A statistical approach to snakes for bimodal and trimodal imagery*, in Proceedings of the International Conference on Computer Vision, Vol. 2, 1999, pp. 898–903.
- [60] L. YOUNES, *Computable elastic distances between shapes*, SIAM J. Appl. Math., 58 (1998), pp. 565–586.
- [61] L. YOUNES, P. W. MICHOR, J. SHAH, AND D. MUMFORD, *A metric on shape space with explicit geodesics*, Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl., 19 (2008), pp. 25–57.
- [62] S. C. ZHU, T. S. LEE, AND A. L. YUILLE, *Region competition: Unifying snakes, region growing, energy/bayes/MDL for multi-band image segmentation*, in Proceedings of the 5th IEEE International Conference on Computer Vision, Cambridge, MA, 1995, pp. 416–423.