Metrics of curves in shape optimization and analysis

Andrea C. G. Mennucci

Level set and PDE based reconstruction methods:
applications to inverse problems and image processing
CIME, Cetraro, 2008

Key words: Shape space, shape optimization, shape analysis, computer vision, Riemannian geometry, manifold of curves, hyperspace of compact sets, edge detection, image segmentation, visual tracking, curve evolution, gradient flow, active contour, sobolev active contour.

Introduction

In these lecture notes we will explore the mathematics of the space of immersed curves, as is nowadays used in applications in computer vision. In this field, the space of curves is employed as a *"shape space"*; for this reason, we will also define and study the space of *geometric curves*, that are *immersed curves up to reparameterizations*. To develop the usages of this space, we will consider the *space of curves* as an *infinite dimensional differentiable manifold*; we will then deploy an effective form of calculus and analysis, comprising tools such as a *Riemannian metric*, so as to be able to perform standard operations: minimizing a goal functional by gradient descent, or computing the distance between two curves. Along this path of mathematics, we will also present some current literature results. (Another common and interesting example of "shape spaces" is the space of all compact subsets of $\mathbb{R}^n$ — we will briefly discuss this option as well, and relate it to the aforementioned theory).

These lecture notes aim to be as self-contained as possible, so as to be accessible to young mathematicians and non-mathematicians as well. For this reason, many examples are intermixed with the definitions and proofs; in presenting advanced and complex mathematical ideas, the rigorous mathematical definitions and proofs were sometimes sacrificed and replaced with an intuitive description.

These lecture notes are organized as follows. Section 1 introduces the definitions and some basilar concepts related to immersed and geometric curves. Section 2 overviews the realm of applications for a shape space in computer vision, that we divide in the fields of "shape optimization" and "shape analysis"; and hilights features and problems of those theories as were studied up to a few years ago, so as to identify the needs and obstacles to further developments. Section 3 contains a summary of all mathematical concepts that are needed for the rest of the notes. Section 4 coalesces all the above in more precise definitions of spaces of curves to be used as "shape spaces", and sets mathematical requirements and goals for applications in computer vision. Section 5 indexes examples of "shape spaces" from the current literature, inserting it in a common paradigm of "representation of shape"; some of this literature is then elaborated upon in the following sections 6,7,8,9, containing two examples of metrics of compact subsets of $\mathbb{R}^n$, two examples of Finsler metrics of curves, two examples of Riemannian metrics of curves "up to pose", and four examples of Riemannian metrics of immersed curves. The last such example is the family of Sobolev-type Riemannian metrics of immersed curves, whose properties are studied in Section 10, with applications and numerical examples. Section 11 presents advanced mathematical topics regarding the Riemannian spaces of immersed and geometric curves.

I gratefully acknowledge that a part of the theory and many numerical experiments exposited were developed in joint work with Prof. Yezzi (GaTech) and Prof. Sundaramoorthi (UCLA); other numerical experiments were by A. Duci and myself. I also deeply thank the organizers for inviting me to Cetraro to give the lectures that were the basis for these lecture notes.

1 Shapes & curves

What is this course about? In the first two sections we begin by summarizing in a simpler form the definitions, reviewing the goals, and presenting some useful mathematical tools.

1.1 Shapes

A wide interest for the study of *shape spaces* arose in recent years, in particular inside the *computer vision* community. Some examples of shape spaces are as follows.

- The family of all collections of k points in $\mathbb{R}^n$.
- The family of all non empty compact subsets of $\mathbb{R}^n$. 
- The family of all closed curves in $\mathbb{R}^n$. 

There are two different (but interconnected) fields of applications for a good shape space in *computer vision*:

shape optimization where we want to find the shape that best satisfies a design goal; a topic of interest in engineering at large;

shape analysis where we study a family of shapes for purposes of statistics, (automatic) cataloging, probabilistic modeling, among others, and possibly create an a-priori model for a better shape optimization.

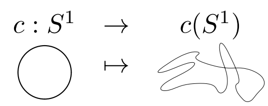
1.2 Curves

We will use formulæ of the form $A := B$ to mean that the formula A is defined by the formula B.

$S^1 = \{x \in \mathbb{R}^2 \mid |x| = 1\}$ is the circle in the plane. S^1 is the template for all possible **closed curves**. (*Open curves* will be called **paths**, to avoid confusion).

Definition 1.1 (Classes of curves)

- *A C^1* ***curve*** *is a continuously differentiable map $c : S^1 \to \mathbb{R}^n$ such that the derivative $c'(\theta) := \partial_\theta c(\theta)$ exists at all points $\theta \in S^1$.*
- *An* ***immersed curve*** *is a C^1 curve c such that $c'(\theta) \neq 0$ at all points $\theta \in S^1$.*



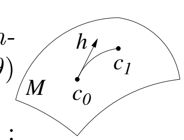
Note that, in our terminology , the "curve" is the function c, and not just the image $c(S^1)$ inside $\mathbb{R}^n$.

Most of the theory following will be developed for curves in $\mathbb{R}^n$, when this does not complicate the math. We will call **planar curves** those whose image is in $\mathbb{R}^2$.

The class of immersed curves is a differentiable manifold. For the purposes of this introduction, we present a simple, intuitive definition.

Definition 1.2 *The* ***manifold of (parametric) curves*** *M is the set of all closed immersed curves. Suppose that $c \in M$, $c : S^1 \to \mathbb{R}^n$ is a closed immersed curve.*

- *A deformation of c is a function $h : S^1 \to \mathbb{R}^n$.*
- *The set of all such h is the* ***tangent space*** *$T_c M$ of M at c.*
- *An infinitesimal deformation of the curve c_0 in "direction" h will yield (on first order) the curve $c_0(u) + \varepsilon h(u)$.*
- *A* ***homotopy*** *C connecting c_0 to c_1 is a continuously differentiable function $C : [0,1] \times S^1 \to \mathbb{R}^n$ such that $c_0(\theta) = C(0,\theta)$ and $c_1(\theta) = C(1,\theta)$.*
  By defining $\gamma(t) = C(t,\cdot)$ we can associate C to a path $\gamma : [0,1] \to M$ in the space of curves M, connecting c_0 to c_1.



When dealing with homotopies $C = C(t,\theta)$, we will use the "prime notation" C' for the derivative $\partial_\theta C$ in the "curve parameter" θ, and the "dot notation" $\dot{C}$ for the derivative $\partial_t C$ in the "time parameter" t.

1.3 Geometric curves and functionals

Shapes are usually considered to be *geometric objects*. Representing a curve using $c : S^1 \to \mathbb{R}^n$ forces a choice of parameterization, that is not really part of the concept of "shape". To get rid of this, we first summarily present what *reparameterizations* are.

Definition 1.3 *Let* $\mathrm{Diff}(S^1)$ *be the family of* ***diffeomorphisms*** *of* S^1*: all the maps* $\phi : S^1 \to S^1$ *that are* C^1 *and invertible, and the inverse* ϕ^{-1} *is* C^1*.*

$\mathrm{Diff}(S^1)$ enjoys some important mathematical properties.

- $\mathrm{Diff}(S^1)$ is a group, and its **group operation** is the composition $\phi, \psi \mapsto \phi \circ \psi$.
- $\mathrm{Diff}(S^1)$ is divided in two connected components, the family $\mathrm{Diff}^+(S^1)$ of diffeomorphisms with derivative $\phi' > 0$ at all points; and the family $\mathrm{Diff}^-(S^1)$ of diffeomorphisms with derivative $\phi' < 0$ at all points.

  $\mathrm{Diff}^+(S^1)$ is a subgroup of $\mathrm{Diff}(S^1)$.
- $\mathrm{Diff}(S^1)$ acts on M, the **action**[1] is the right function composition $c \circ \phi$. The resulting curve $c \circ \phi$ is a **reparameterization** of c.

  The action of the subgroup $\mathrm{Diff}^+(S^1)$ moreover does not change the orientation of the curve.

Definition 1.4 *The* ***quotient space***

$$B := M/\mathrm{Diff}(S^1)$$

is the space of curves up to reparameterization*, also called* ***geometric curves*** *in the following. Two parametric curves* $c_1, c_2 \in M$ *such that* $c_1 = c_2 \circ \phi$ *are the same geometric curve inside* B*.*

For some applications we may choose instead to consider the quotient w.r.to $\mathrm{Diff}^+(S^1)$*; the quotient space* $M/\mathrm{Diff}^+(S^1)$ *is the space of* ***geometric oriented curves****.*

B is mathematically defined as the set $B = \{[c]\}$ of all **equivalence classes** $[c]$ of curves that are equal but for reparameterization,

$$[c] := \{c \circ \phi \text{ for } \phi \in \mathrm{Diff}(S^1)\} \quad ;$$

and similarly for the quotient $M/\mathrm{Diff}^+(S^1)$. More properties of these quotients will be presented in Section 4.5.

We can now define the **geometric functional**s (that are, loosely speaking invariant w.r.to reparameterization of the curve).

Definition 1.5 *A functional* $F(c)$ *defined on curves will be called* ***geometric*** *if one of the two following alternative definitions holds.*

[1] We will provide more detailed definitions and properties of the "actions" in Section 3.8.

- *$F(c) = F(c \circ \phi)$ for all curves c and for all $\phi \in \mathrm{Diff}(S^1)$.*
- *For all curves c and all ϕ, if $\phi \in \mathrm{Diff}^+(S^1)$ then $F(c) = F(c \circ \phi)$, whereas if $\phi \in \mathrm{Diff}^-(S^1)$ then $F(c) = -F(c \circ \phi)$*

In the first case, F can be “projected” to B, that is, it may be considered as a function $F : B \to \mathbb{R}$. In the second case, F can be “projected” to $M/\mathrm{Diff}^+(S^1)$.

It is important to remark that “geometric” theories have often provided the best results in *computer vision*.

1.4 Curve–related quantities

A good way to specify the *design goal* for shape optimization is to define an **objective function** (a.k.a. **energy**) $F : M \to \mathbb{R}$ that is minimum in the curve that is most fit for the task.

When designing our F, we will want it to be *geometric*; this is easily accomplished if we use geometric quantities to start from. We now list the most important such quantities.

In the following, given $v, w \in \mathbb{R}^n$ we will write $|v|$ for the standard **Euclidean norm**, and $\langle v, w\rangle$ or $(v \cdot w)$ for the standard **scalar product**. We will again write $c'(\theta) := \partial_\theta c(\theta)$.

Definition 1.6 (Derivations) *If the curve c is* immersed*, we can define the **derivation with respect to the arc parameter***

$$\partial_s = \frac{1}{|c'|}\partial_\theta \ .$$

(We will sometimes also write D_s instead of ∂_s.)

Definition 1.7 (Tangent) *At all points where $c'(\theta) \neq 0$, we define the **tangent vector***

$$T(\theta) = \frac{c'(\theta)}{|c'(\theta)|} = \partial_s c(\theta) \quad .$$

(At the points where $c' = 0$ we may define $T = 0$).

It is easy to prove (and quite natural for our geometric intuition) that D_s and T are geometric quantities (according to the second Definition 1.5).

Definition 1.8 (Length) *The **length** of the curve c is*

$$\mathrm{len}(c) := \int_{S^1} |c'(\theta)| \, \mathrm{d}\theta \ . \tag{1.1}$$

We can define formally the **arc parameter** s by

$$\mathrm{d}s := |c'(\theta)| \, \mathrm{d}\theta \ ;$$

we use it only in integration, as follows.

Definition 1.9 (Integration by arc parameter) *We define the* ***integration by arc parameter*** *of a function $g : S^1 \to \mathbb{R}^k$ along the curve c by*

$$\int_c g(s)\,\mathrm{d}s := \int_{S^1} g(\theta)|c'(\theta)|\,\mathrm{d}\theta \ .$$

and the ***average integral***

$$⨍_c g(s)\,\mathrm{d}s := \frac{1}{\mathrm{len}(c)} \int_c g(s)\,\mathrm{d}s$$

and we will sometimes shorten this as $\boldsymbol{avg}_c(g)$.

The above integrations are geometric quantities (according to the first Definition 1.5).

1.4.1 Curvature

Suppose moreover that the curve is immersed and is C^2 regular (that means that it is twice differentiable, and the second derivative is continuous); in this case we may define *curvatures*, that indicate how much the curve *bends*. There are two different definitions of *curvature* of an immersed curve: *mean curvature* H and *signed curvature* κ, that is defined when c is planar. H and k are *extrinsic curvatures*, they are properties of the *embedding* of c into $\mathbb{R}^n$.

Definition 1.10 (H) *If c is C^2 regular and immersed, we can define the* ***mean curvature*** *H of c as*

$$H = \partial_s \partial_s c = \partial_s T$$

where $\partial_s = \frac{1}{|c'|}\partial_\theta$ is the *derivation with respect to the arc parameter*. It enjoys the following properties.

Properties 1.11

- *It is easy to prove that $H \perp T$.*
- *H is a geometric quantity. If $\tilde{c} = c \circ \phi$ and $\tilde{H}$ is its curvature, then $\tilde{H} = H \circ \phi$.*
- *$1/|H|$ is the radius of a tangent circle that best approximates the curve to second order.*

Definition 1.12 (N) *When the curve c is immersed and planar, we can define a* ***normal vector*** *N to the curve, by requiring that $|N| = 1$, $N \perp T$ and N is rotated $\pi/2$ radians anticlockwise with respect to T.*

Definition 1.13 (κ) *If c is planar and C^2 regular, then we can define a* ***signed scalar curvature*** *$\kappa = \langle H, N\rangle$, so that*

$$\partial_s T = \kappa N = H \quad \textit{and} \quad \partial_s N = -\kappa T \quad .$$

See fig. 1 on the following page. The above equality implies that κ is a geometric quantity, according to the second Definition 1.5.

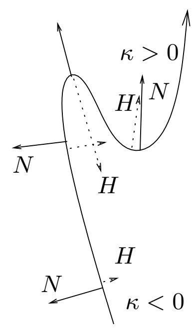


Figure 1: Example of a regular curve and its curvature.

2 Shapes in applications

A number of methods have been proposed in shape analysis to define distances between shapes, averages of shapes and statistical models of shapes. At the same time, there has been much previous work in shape optimization, for example image segmentation via active contours, 3D stereo reconstruction via deformable surfaces; in these later methods, many authors have defined energy functionals $F(c)$ on curves (or on surfaces), whose minima represent the desired segmentation/reconstruction; and then utilized the calculus of variations to derive curve evolutions to search minima of $F(c)$, often referring to these evolutions as gradient flows. The reference to these flows as gradient flows *implies a certain Riemannian metric on the space of curves; but this fact has been largely overlooked. We call this metric H^0, and properly define it in eqn. (2.4).*

2.1 Shape analysis

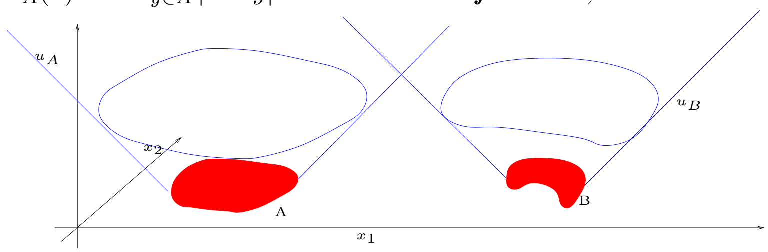
Many method and tools comprise the **shape analysis**. We may list

- distances between shapes,
- averages for shapes,
- principle component analysis for shapes and
- probabilistic models of shapes.

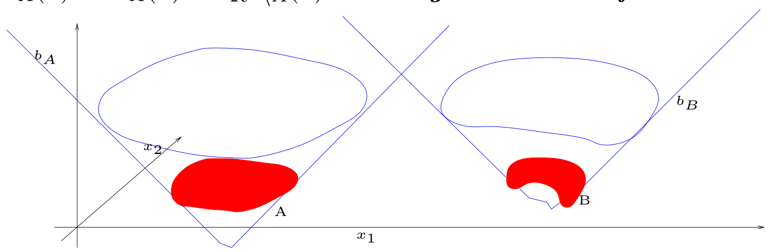
We will present a short overview of the above, in theory and in applications. We begin by defining the *distance function* and *signed distance function*, two tools that we will use often in this theory.

Definition 2.1 *Let $A, B \subset \mathbb{R}^n$ be closed sets.*

- $u_A(x) := \inf_{y \in A} |x - y|$ *is the* ***distance function****,*



- $b_A(x) := u_A(x) - u_{\mathbb{R}^n \setminus A}(x)$ *is the* ***signed distance function****.*



2.1.1 Shape distances

A variety of distances have been proposed for measuring the difference between two given shapes. Two examples follows.

Definition 2.2 *The* ***Hausdorff distance***

$$d_H(A, B) := \Big(\sup_{x \in A} u_B(x) \Big) \vee \Big(\sup_{x \in B} u_A(x) \Big) \tag{2.1}$$

where $A, B \subset \mathbb{R}^n$ *are compact, and* $u_A(x)$ *is the* distance function*.*

The above can be used for curves; in this case a distance may be defined by associating to the two curves c_1, c_2 the Hausdorff distance of the image of the curves

$$d_H(c_1(S^1), c_2(S^1)) \ .$$

If c_1, c_2 are immersed and planar, we may otherwise use $d_H(\mathring{c}_1, \mathring{c}_2)$ where $\mathring{c}_1, \mathring{c}_2$ denote the closed region of the plane that is enclosed by c_1, c_2.

Definition 2.3 *Let* $A, B \subset \mathbb{R}^n$ *be two measurable sets, let*

$$A \Delta B := (A \setminus B) \cup (B \setminus A)$$

be the ***set symmetric difference****; let* $|A|$ *be the Lebesgue measure of* A*. We define the* ***set symmetric distance*** *as*

$$d(A, B) = \big| A \Delta B \big| \ ,$$

(where it is to be intended that two sets are considered "equal" if they differ by a measure zero set).

In the case of planar curves c_1, c_2, we can apply to above idea to define

$$d(c_1, c_2) = \left| \mathring{c}_1 \Delta \mathring{c}_2 \right| .$$

2.1.2 Shape averages

Many definitions have also been proposed for the average $\bar{c}$ of a finite set of shapes $c_1, \ldots, c_N$. There are methods based on the **arithmetic mean** (that are representation dependent).

- One such case is when the shapes are defined by a finite family of M parameters; so we can define the **parametric or landmark averaging**

  $$\bar{c}(p) = \frac{1}{N} \sum_{n=1}^{N} c_n(p)$$

  where p is a common parameter/landmark.

- More in general, we can define the **signed distance level set averaging** by using as a representative of a shape its signed distance function, and computing the average shape by

  $$\bar{c} = \left\{ x \mid \bar{b}(x) = 0 \right\}, \text{ where } \bar{b}(x) = \frac{1}{N} \sum_{n=1}^{N} b_{c_n}(x)$$

  where b_{c_n} is the signed distance function of c_n; or in case of planar curves, of the part of plane enclosed by c_n.

Then there are *non parametric, representation independent*, methods. The (possibly) most famous one is the

Definition 2.4 *The **distance-based average**.*[2] *Given a distance d_M between shapes, an average shape $\bar{c}$ may be found by minimizing the sum of its squared distances.*

$$\bar{c} = \arg\min_c \sum_{n=1}^{N} d_M(c, c_n)^2$$

It is interesting to note that in Euclidean spaces there is an unique minimum, that coincides with the arithmetic mean; while in general there may be no minimum, or more than one.

2.1.3 Principal component analysis (PCA)

Definition 2.5 *Suppose that X is a* random vector *taking values in $\mathbb{R}^n$; let $\overline{X} = \mathbb{E}(X)$ be the mean of X. The **principal component analysis** is the representation of X as*

$$X = \overline{X} + \sum_{i=1}^{n} Y_i S_i$$

[2]Due to Fréchet, 1948; but also attributed to Karcher Karcher [1977].

where S_i are constant vectors, and Y_i are uncorrelated real random variables with zero mean, and with decreasing variance. S_i is known as the ***i-th mode of principal variation****.*

The PCA is possible in general in any finite dimensional linear space equipped with an inner product. In infinite dimensional spaces, or equivalently in case of *random processes*, the PCA is obtained by the *Karhunen-Loève theorem*.

Given a finite number of samples, it is also possible to define an *empirical principal component analysis*, by estimating the expectation and variance of the data.

In many practical cases, the variances of Y_i decrease quite fast: it is then sensible to replace X by a simplified variable

$$\tilde{X} := \overline{X} + \sum_{i=1}^{k} Y_i S_i$$

with $k < n$.

So, if shapes are represented by some *"linear"* representation, the PCA is a valuable tool to study the statistics, and it has then become popular for imposing *shape priors* in various shape optimization problems. For more general manifold representations of shapes, we may use the *exponential map* (defined in 3.29), replacing S_n by *tangent vectors* and following *geodesics* to perform the summation.

We present here an example in applications.

Example 2.6 (PCA by signed distance representation) *In these pictures we see an example of synergy between analysis and optimization, from Tsai et al. [2001a]. The figure 2 contains a row of pictures that represent the (empirical) PCA of signed distance function of a family of plane shapes. The figure 3 on the following page contains a second row of images that represent: an image of a plane shape* (a) *occluded* (b) *and noise degraded* (c)*, an initialization for the shape optimizer* (d) *and the final optimal segmentation* (e)*.*

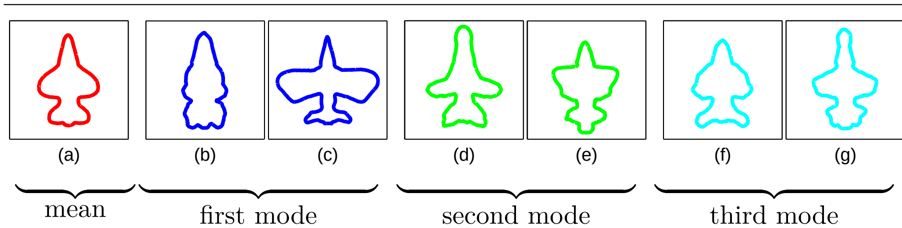


Figure 2: PCA of plane shapes.

(From Tsai et al. [2001a] © 2001 IEEE. Reproduced with permission).

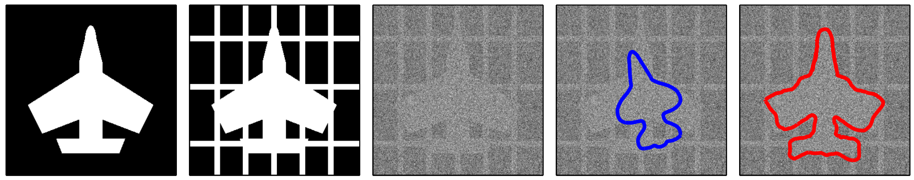


Figure 3: Segmentation of occluded plane shape.
(From Tsai et al. [2001a] © 2001 IEEE. Reproduced with permission).

2.2 Shape optimization & active contours

2.2.1 A short history of active contours

In the late 20th century, the most common approach to *image segmentation* was a combination of **edge detection** and **contour continuation**. With edge detection Canny [1986], small **edge elements** would be identified by a local analysis of the image; then a **continuation** method would be employed to reconstruct full contours. The methods were suffering from two drawbacks, edge detection being too sensitive to noise and to photometric features (such as, sharp shadows, reflections) that were not related to the physical structure of the image; and continuation was in essence a NP-complete algorithm.

Active contours were introduced by Kass et al. [1987] as a simpler approach to the segmentation problem. The idea is to minimize an energy $F(c)$ (where the variable c is a contour *i.e.* a curve), that contains an **edge-based** attraction term and a smoothness term, which becomes large when the curve is irregular. To search for the minimum of $F(c)$, a curve evolution method was derived from the calculus of variations. The evolving curve would then be attracted to the features of interest, while mantaining the property of being a closed smooth contour.

An unjustified feature of the model of [Kass et al., 1987] is that the evolution is dependent on the way the contour is parameterized. Hence Caselles et al. [1993]; Malladi et al. [1995] revisited the idea of [Kass et al., 1987] in the new form of a **geometric evolution**, which could be implemented by the **level set method** invented by Osher and Sethian [1988]. Thereafter, Kichenassamy et al. [1995] Caselles et al. [1995] presented the active contour approach to the edge detection problem in the form of the minimization of a geometric energy that is a generalization of Euclidean arc length. Again, the authors derived the **gradient descent flow** in order to minimize the geometric energy.

In contrast to the edge-based approaches for active contours (mentioned above), **region-based** energies for active contours have been proposed in Ronfard [1994] Zhu et al. [1995] Yezzi et al. [1999] Chan and Vese [2001]. In these approaches, an energy is designed to be minimized when the curve partitions the image into statistically distinct regions. This kind of energy has provided many desirable features; for example, it provides less sensitivity to noise, better ability to capture concavities of objects, more dependence on global features of the image, and less sensitivity to the initial contour placement.

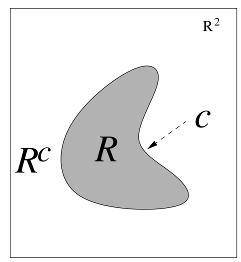
In Mumford and Shah [1985, 1989], the authors introduced and rigorously studied a region-based energy that was designed to both extract the boundary of distinct regions while also smoothing the image within these regions. Subsequently, Tsai et al. [2001b] Vese and Chan [2002] implemented a **level set method** for the curve evolution minimizing the energy functional considered by Mumford&Shah.

2.2.2 Energies in *computer vision*

A variational approach to solving shape optimization problems is to define and consequently minimize geometric energy functionals $F(c)$ where c is a planar curve. We may identify two main families of energies,

- the **region based energies**:

$$F(c) = \int_R f_{in}(x)\,dx + \int_{R^c} f_{out}(x)\,dx$$



  where $\int \cdots dx$ is area element integral, and R and R^c are the interior and exterior areas outlined by c; and

- the **boundary based energies**:

$$F(c) = \int_c \phi(c(s))\,ds$$

where ϕ is designed to be small on the salient features of the image, and $\int_C \cdots \mathrm{d}s$ is the *integration by arc parameter* defined in 1.9. Note that $\phi(x)\,\mathrm{d}s$ may be interpreted as a conformal metric on $\mathbb{R}^2$, so minima curves are geodesics w.r.to the conformal metric.

2.2.3 Examples of geometric energy functionals for segmentation

Suppose $I : \Omega \to [0,1]$ is the image (in black & white). We propose two examples of energies taken from the above families.

- The **Chan-Vese** segmentation energy Vese and Chan [2002]

$$F(c) \quad = \quad \int_R \left| I(x) - \mathrm{avg}_R(I) \right|^2 dx + \int_{R^c} \left| I(x) - \mathrm{avg}_{R^c}(I) \right|^2 dx$$

  where $\mathrm{avg}_R I = \frac{1}{|R|} \int_R I(x)dx$ is the average of I on the region R.

- The **geodesic active contour**, (Caselles et al. [1995],Kichenassamy et al. [1995])

$$F(c) = \int_c \phi(c(s))\,\mathrm{d}s \tag{2.2}$$

where ϕ may be chosen to be

$$\phi(x) = \frac{1}{1 + |\nabla I(x)|^2}$$

that is small on sharp discontinuities of I (∇I is the gradient of I w.r.to x). Since real images are noisy, the function ϕ in practice would be more influenced by the noise than by the image features; for this reason, usually the function ϕ is actually defined by

$$\phi(x) = \frac{1}{1 + |\nabla G \star I(x)|^2}$$

where G is a smoothing kernel, such as the *Gaussian*.

2.3 Geodesic active contour method

2.3.1 The "geodesic active contour" paradigm

The general procedure for geodesic active contours goes as follows.

1. Choose an appropriate *geometric* energy functional, E.
2. Compute the **directional derivative**

$$DE(c; h) = \left. \frac{\mathrm{d}}{\mathrm{d}t} E(c + th) \right|_{t=0}$$

   where c is a curve and h is an arbitrary perturbation of c.

3. Manipulate $DE(c; h)$ into the form

$$\int_c h(s) \cdot v(s) \, \mathrm{d}s \ .$$

4. Consider v to be the "*gradient*", the direction which increases E *fastest*.
5. Evolve $c = c(t, \theta)$ by the differential equation $\partial_t c = -v$; this is called the **gradient descent flow**.

(Note that the *directional derivative* is the basis for the definition 3.13 of the *Gâteaux differential*, that will be then discussed in Section 3.7 and 4.2.)

2.3.2 Example: geodesic active contour edge model

We propose an explicit computation starting from the classical active contour model.

1. Start from an energy that is minimal along image contours:

$$E(c) = \int_c \phi(c(s)) \, ds$$

   where ϕ is defined to extract relevant features; for examples as discussed after eqn. (2.2).

2. Calculate the directional derivative:

$$\begin{aligned} DE(c;h) &= \int_c \nabla\phi(c)\cdot h + \phi(c)(D_s h \cdot D_s c)\,\mathrm{d}s \\ &= \int_c \nabla\phi(c)\cdot h - \big(\nabla\phi(c)\cdot D_s c\big)(h\cdot D_s c) - \phi(c)(h\cdot D_{ss}c)\,\mathrm{d}s \\ &= \int_c h\cdot\big((\nabla\phi(c)\cdot N)N - \phi(c)\kappa N\big)\,\mathrm{d}s\ . \end{aligned} \tag{2.3}$$

3. Deduce the "gradient":

$$\nabla E = -\phi\kappa N + (\nabla\phi\cdot N)N\ .$$

4. Write the gradient flow

$$\frac{\partial c}{\partial t} = \phi\kappa N - (\nabla\phi\cdot N)N\ .$$

(Note that the flow of geometric energies moves only in orthogonal direction w.r.t the curve — this phenomenon will be explained in in Section 11.10.)

2.3.3 Implicit assumption of H^0 inner product

We have made a critical assumption in going from the directional derivative

$$DE(c;h) = \int_c h(s)\cdot v(s)\ \mathrm{d}s$$

to deducing that the gradient of E is $\nabla E(c) = v$. Namely, the definition of the gradient [3] is based on the following equality

$$\langle h(s), \nabla E\rangle = \int_c \underbrace{h(s)}_{h_1=h}\cdot\underbrace{v(s)}_{h_2=\nabla E}\ \mathrm{d}s,$$

that needs an inner-product structure.

This implies that we have been presuming all along that curves are equipped with a **H^0-type inner-product** defined as follows

$$\big\langle h_1, h_2\big\rangle_{H^0} := \int_c h_1(s)\cdot h_2(s)\,\mathrm{d}s\ . \tag{2.4}$$

2.4 Problems & goals

We will now briefly list some examples that show some limits of the usual *active contour* method.

[3] The precise definition of what the gradient is is in section 3.7.

2.4.1 Example: geometric heat flow

We first review one of the most mathematically studied gradient descent flows of curves. By direct computation, the *Gâteaux differential* of the length of a closed curve is

$$D\,\mathrm{len}(c; h) = \int_c \partial_s h \cdot T \,\mathrm{d}s = -\int_c h \cdot H \,\mathrm{d}s \tag{2.5}$$

Let $C = C(t, \theta)$ be an evolving family of curves. The **geometric heat flow** (also known as **motion by mean curvature**) is

$$\frac{\partial C}{\partial t} = H$$

where $H := \partial_s \partial_s C$ is the mean curvature. In the H^0 inner-product, this flow is the *gradient descent for curve length.*[4]

This flow has been studied deeply by the mathematical community, and is known to enjoy the following properties.

Properties 2.7

- *Embedded planar curves remain embedded.*
- *Embedded planar curves converge to a circular point.*
- Comparison principle*: if two embedded curves are used as initialization, and the first contains the second, then it will contain the second for all time of evolution. This is important for* level set method*s.*
- *The flow is well posed only for positive times, that is, for increasing t (similarly to the usual* heat equation*).*

For the proofs, see Gage and Hamilton [1986]; Grayson [1987].

2.4.2 Short length bias

The usual edge-based active contour energy is of the form

$$E(c) = \int_c g(c(s)) \,\mathrm{d}s$$

supposing that g is reasonably constant in the region where c is running, then $E(c) \simeq \overline{g}\,\mathrm{len}(c)$, due to the integral being performed w.r.to the arc parameter $\mathrm{d}s$. So one way to reduce E is to shrink the curve. Consequently, when the image is smooth and featureless (as in medical imaging), the usual edge based energies would often drive the curve to a point.

To avoid it, an **inflationary term** νN (with $\nu > 0$) was usually added to the curve evolution, to obtain

$$\frac{\partial c}{\partial t} = \phi \kappa N - \nabla \phi + \nu N \ ,$$

see Caselles et al. [1995]; Kichenassamy et al. [1995]. Unfortunately, this adds a parameter that has to be tuned to match the characteristics of each specific image.

[4] A different gradient descent flow for curve length will be discussed in Remark 10.32.

2.4.3 Average weighted length

As an alternative approach we may normalize the energy to obtain a more length-independent energy, the **average weighted length**

$$\operatorname{avg}_c(g) := \frac{1}{\operatorname{len}(c)} \int_c g(c(s))\, \mathrm{d}s \tag{2.6}$$

but this generates an **ill-posed** H^0-flow, as we will now show.

2.4.4 Flow computations

Here we write $L := \operatorname{len}(c)$. Let $g : \mathbb{R}^n \to \mathbb{R}^k$; let

$$\operatorname{avg}_c(g(c)) := \frac{1}{L} \int_c g(c(s))\, \mathrm{d}s = \frac{1}{L} \int_{S^1} g(c(\theta))|c'(\theta)|\, \mathrm{d}\theta \quad ;$$

then the *Gâteaux differential* is

$$\begin{aligned} D[\operatorname{avg}_c(g(c))](c;h) &= \frac{1}{L} \int_c \nabla g(c) \cdot h + g(c)(\partial_s h \cdot T)\, \mathrm{d}s - \\ &\quad - \frac{1}{L^2} \int_c g(c)\, \mathrm{d}s \int_c \partial_s h \cdot T\, \mathrm{d}s = \\ &= \fint_c \nabla g(c) \cdot h + \big(g(c) - \operatorname{avg}_c(g(c))\big)(\partial_s h \cdot T)\, \mathrm{d}s \quad . \end{aligned} \tag{2.7}$$

In the above, we omit the argument (s) for brevity; ∇g is the gradient of g.

If the curve is planar, we define the normal and tangent vectors N and T as by 1.12 and 1.13; then, integrating by parts, the above becomes

$$\begin{aligned} D[\operatorname{avg}_c(g(c))](c;h) &= \frac{1}{L} \int_c \nabla g(c) \cdot h - (\nabla g(c) \cdot T)(h \cdot T) - \\ &\quad - \big(g(c) - \operatorname{avg}_c(g(c))\big)(h \cdot D_s^2 c)\, \mathrm{d}s = \\ &= \fint_c \Big(\nabla g(c) \cdot N - \kappa\big(g(c) - \operatorname{avg}_c(g(c))\big)\Big)(h \cdot N)\, \mathrm{d}s \quad . \end{aligned} \tag{2.8}$$

Suppose now that we have a shape optimization functional E including a term of the form $\operatorname{avg}_c(g(c))$; let $C = C(t,\theta)$ be an evolving family of curves trying to minimize E; this flow would contain a term of the form

$$\frac{\partial C}{\partial t} = \ldots \big(g(c(s)) - \operatorname{avg}_c(g)\big)\kappa N \ldots$$

unfortunately the above flow is ill defined: it is a backward-running *geometric heat flow* on roughly half of the curve. We will come back to this example in Section 10.8.1; more examples and methods for computing Gâteaux differentials are in Section 4.2.

2.4.5 Centroid energy

We will now propose another simple example where the above phenomenon is again evident.

Example 2.8 (Centroid-based energy) *Let us fix a target point $v \in \mathbb{R}^n$. We recall that $avg_c(c)$ is the* ***center of mass*** *of the curve. The energy*

$$E(c) := \frac{1}{2}|avg_c(c) - v|^2 \tag{2.9}$$

penalizes the distance from the center of mass to v. Let in the following $\bar{c} = avg_c(c)$ for simplicity. The directional derivative *of E in direction h is*

$$DE(c; h) = \Big\langle \bar{c} - v, D(\bar{c})(c; h) \Big\rangle$$

where in turn (by eqn. (2.7) and (2.8))

$$\begin{aligned} D(\bar{c})(c; h) &= \fint_c h + (c - \bar{c})(D_s h \cdot D_s c)\, \mathrm{d}s = \\ &= \fint_c h - D_s c\,(h \cdot D_s c) - (c - \bar{c})(h \cdot D_s^2 c)\, \mathrm{d}s \end{aligned} \tag{2.10}$$

supposing that the curve is planar, then

$$h - D_s c\,(h \cdot D_s c) = N(h \cdot N)$$

so

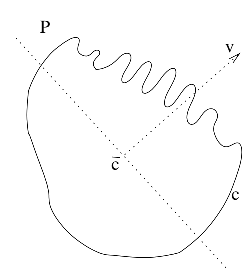
$$DE(c; h) = \fint \langle \bar{c} - v, N\rangle (h \cdot N) - \langle \bar{c} - v, c - \bar{c}\rangle \kappa (h \cdot N)\, \mathrm{d}s\ .$$

The H^0 gradient descent flow is then

$$\frac{\partial c}{\partial t} = -\nabla_{H^0} E(c) = \langle (v - \bar{c}), N\rangle N - \kappa N \Big\langle (c - \bar{c}), (v - \bar{c}) \Big\rangle\ .$$

The first term $\langle (v - \bar{c}) \cdot N\rangle N$ in this gradient descent flow moves the whole curve towards v.

Let $P := \{w : \big\langle (w - \bar{c}) \cdot (v - \bar{c}) \big\rangle \geq 0\}$ be the half plane "on the v side". The second term $-\kappa N \big\langle (c - \bar{c}) \cdot (v - \bar{c}) \big\rangle$ in this gradient descent flow tries to decrease the curve length out of P and increase the curve length in P, ***and this is ill posed****.*



We will come back to this example in Proposition 10.20.

2.4.6 Conclusions

More recent works that use active contours for segmentation are not only based on information from the image to be segmented (edge-based or region-based), but also **prior information**, that is information known about the *shape* of the desired object to be segmented. The work of Leventon et al. [2000] showed how to incorporate *prior information* into the active contour paradigm. Subsequently, there have been a number of works, for example Tsai et al. [2001a]; Rousson and Paragios [2002]; Chen et al. [2002]; Cremers and Soatto [2003]; Raviv et al. [2004], which design energy functionals that incorporate prior shape information of the desired object. In these works, the main idea is designing a novel term of the energy that is small when the curve is *close*, in some sense, to a pre-specified shape.

The need for *prior information* terms arose from several factors such as

- the fact that some images contain limited information,
- the energies functions considered previously could not incorporate complex information,
- the energies had too many extraneous local minima, and *the gradient flows* to minimize these energies allowed for arbitrary deformations that gave rise to unlikely shapes; as in the example in Fig. 4 of segmentation using the *Chan-Vese* energy, where the flowing contour gets trapped into noise.

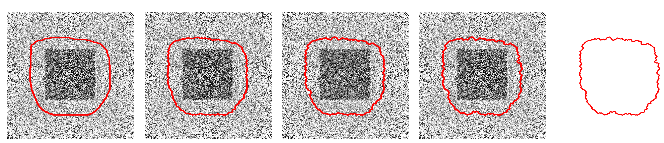


Figure 4: H^0 gradient descent flow of Chan–Vese energy, to segment a noisy square

Works on incorporating prior shape knowledge into active contours have led to a fundamental question on how to define distance between two curves or shapes. Many works, for example Younes [1998]; Soatto and Yezzi [2002]; Mio and Srivastava [2004a]; Charpiat et al. [2004]; Michor and Mumford [2006]; Yezzi and Mennucci [2004b, 2005], in the *shape analysis* literature have proposed different ways of defining this distance.

However, Michor and Mumford [2006]; Yezzi and Mennucci [2005] observed that all previous works on geometric active contours that derive *gradient flows* to minimize energies, which were described earlier, imply a natural notion of Riemannian metric, given by the geometric inner product H^0 (that we just "discovered" in eqn. (2.4)). Subsequently, Michor and Mumford [2006]; Yezzi and Mennucci [2004b] have shown a surprising property: the H^0 Riemannian metric on the space of curves is **pathological**, since the "distance" between *any* two curves is zero.

In addition to the pathologies of the Riemannian structure induced by H^0, there are also undesirable features of H^0 gradient flows. Some of these features are listed below.

1. There are no regularity terms in the definition of the H^0 inner product. That is, there is nothing in the definition of H^0 that discourages flows that are not smooth in the space of curves. Thus, when energies are designed to depend on the image that is to be segmented, the H^0 gradient is very sensitive to noise in the image.

   Therefore, in *geometric active contour* models, a penalty on the curve's length is added to keep the curve smooth. However, this changes the energy that is being optimized and ends up solving a different problem.

2. H^0 gradients, evaluated at a particular point on the curve, depend locally on derivatives of the curve. Therefore, as the curve becomes non-smooth, as mentioned above, the derivative estimates become inaccurate, and thus, the curve evolution becomes inaccurate. Moreover, for region-based and edge-based active contours, the H^0 gradient at a particular point on the curve depends locally on image data at the particular point. Although region-based energies may depend on global statistics, such as mean values of regions, the H^0 gradient still depends on local data. These facts imply that the H^0 gradient descent flow is sensitive to noise and local features.

3. The H^0 norm gives *non-preferential* treatment to arbitrary deformations regardless of whether the deformations are global motions (not changing the shape of the curve) such as translations, rotations and scales; or whether they are more local deformations.

4. Many *geometric active contour*s (such as edge and region-based active contours) require that the unit normal to the evolving curve be defined. As such, the evolution does not make sense for polygons. Moreover, since in general, an H^0 active contour does not remain smooth, one needs special numerical schemes based on viscosity theory in a level set framework to define the flow.

5. Many simple and apparently sound energies cannot be implemented for *shape optimization* tasks;

   - some energies generate *ill-posed* H^0 flows;
   - if an energy integrand uses derivatives of the curve of degree up to d, then the PDE driving the flow has degree $2d$; but derivatives of the curves of high degree are noise sensitive and are difficult to implement numerically, and
   - the active contours method works best when implemented using the *level set method*, but this is difficult for flows PDEs of degree higher than 2.

In conclusion, if one wishes to have a consistent view of the geometry of the space of curves in both shape optimization and shape analysis, then one should use the H^0 metric when computing distances, averages and morphs between shapes. Unfortunately, H^0 does not yield a well define metric structure, since the associated distance is identically zero.

So to achieve our goal, we will need to devise new metrics.

3 Basic mathematical notions

In this section we provide the mathematical theory that will be needed in the rest of the course. (Some of the definitions are usually known to mathematics' students; we will present them nonetheless as a chance to remark less known facts.) We will though avoid technicalities in the definitions, and for the most part just provide a base intuition of the concepts. The interested reader may obtain more details from a books in analysis, such as Apostol [1974], in functional analysis, such as Rudin [1973], and in differential and Riemannian geometry, such as do Carmo [1992], Klingenberg [1982] or Lang [1999].

We start with a basic notion, in a quite simplified form.

Definition 3.1 (Topological spaces) *A **topological space** is a set M with associated a **topology** τ of subsets, that are the **open sets** in M.*

The topology is the simplest and most general way to define what are the "convergent sequences of points" and the "continuous functions". We will not provide details regarding topological spaces, since in the following we will mostly deal with *normed spaces* and *metric spaces*, where the topology is induced by a norm or a metric. We just recall this definition.

Definition 3.2 *A **homeomorphism** is an invertible continuous function $\phi : M \to N$ between two topological spaces M, N, whose inverse ϕ^{-1} is again continuous.*

3.1 Distance and metric space

Definition 3.3 *Given a set M, a **distance** $d = d_M$ is a function*

$$d : M \times M \to [0, \infty]$$

such that

1. $d(x, x) = 0$,
2. *if* $d(x, y) = 0$ *then* $x = y$,
3. $d(x, y) = d(y, x)$ (d is symmetric)
4. $d(x, z) \le d(x, y) + d(y, z)$ (the triangular inequality).

There are some possible generalizations.

- The second request may be waived, in this case d is a **semidistance**.
- The third request may be waived: then d would be an **asymmetric distance**. Most theorems we will see can be generalized to asymmetric distances; see Mennucci [2004].

3.1.1 Metric space

The pair (M, d) is a **metric space**. A metric space has a distinguished topology, generated by balls of the form $B(x, r) := \{y \mid d(x, y) < r\}$; according to this topology, $x_n \to_n x$ iff $d(x_n, x) \to_n 0$; and functions are continuous if they map convergent sequences to convergent sequences. We will assume in the following that the reader is acquainted with the concepts of "*open, closed, compact sets*" and "*continuous functions*" in metric spaces. We recall though the definition of "*completeness*".

Definition 3.4 *A metric space (M, d) is* ***complete*** *iff, for any given sequence $(c_n)_n \subset M$, the fact that*

$$\lim_{m,n\to\infty} d(c_m, c_n) = 0$$

implies that c_n converges.

Example 3.5

- *$\mathbb{R}^n$ is usually equipped with the* ***Euclidean distance***

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \;;$$

  and $(\mathbb{R}^n, d)$ is a complete metric space.
- *Any closed subset of a complete space is complete.*
- *If we cut away an accumulation point out of a space, the resulting space is not complete.*

A complete metric space is a space without "missing points". This is important in optimization: if a space is not complete, any optimization method that moves the variable and searches for the optimal solution may fail since the solution may, in a sense, be "missing" from the space.

3.2 Banach, Hilbert and Fréchet spaces

Definition 3.6 *Given a vector space E, a* ***norm*** *$\|\cdot\|$ is a function*

$$\|\cdot\| : E \to [0, \infty]$$

such that

1. $\|\cdot\|$ *is convex*
2. *if* $\|x\| = 0$ *then* $x = 0$.
3. $\|\lambda x\| = |\lambda| \, \|x\|$ *for* $\lambda \in \mathbb{R}$

Again, there are some possible generalizations.

- If the second request is waived, then $\|\cdot\|$ is a **seminorm**.
- If the third request holds only for $\lambda \geq 0$, then the norm is **asymmetric**; in this case, it may happen that $\|x\| \neq \| - x\|$.

Each (semi/asymmetric)norm $\|\cdot\|$ defines a (semi/asymmetric)distance

$$d(x, y) := \|x - y\|.$$

So a norm induces a topology.

3.2.1 Examples of spaces of functions

We present some examples and definitions.

Definition 3.7 *A* ***locally-convex topological vector space*** E *(shortened as* ***l.c.t.v.s.*** *in the following) is a vector space equipped with a collection of seminorms* $\|\cdot\|_k$ *(with* $k \in K$ *an index set); the seminorms induce a topology, such that* $c_n \to c$ *iff* $\|c_n - c\|_k \to_n 0$ *for all* k*; and all vector space operations are continuous w.r.to this topology.*

The simplest example of l.c.t.v.s. is obtained when there is only one norm; this gives raise to two renowned examples of spaces.

Definition 3.8 (Banach and Hilbert spaces)

- *A* ***Banach space*** *is a vector space* E *with a norm* $\|\cdot\|$ *defining a distance* $d(x, y) := \|x - y\|$ *such that* E *is metrically complete.*
- *A* ***Hilbert space*** *is a space with an inner product* $\langle f, g \rangle$*, that defines a norm* $\|f\| := \sqrt{\langle f, f \rangle}$ *such that* E *is metrically complete.*

(Note that a Hilbert space is also a Banach space).

Example 3.9 *Let* $I \subset \mathbb{R}^k$ *be open; let* $p \in [1, \infty]$*. A standard example of Banach space is the* L^p *space of functions* $f : I \to \mathbb{R}^n$*. If* $p \in [1, \infty)$*, the* L^p *space contains all functions such that* $|f|^p$ *is Lebesgue integrable, and is equipped with the norm*

$$\|f\|_{L^p} := \sqrt[p]{\int_I |f(x)|^p \, \mathrm{d}x} \; .$$

For the case $p = \infty$*,* L^∞ *contains all Lebesgue measurable functions* $f : I \to \mathbb{R}^n$ *such that there is a constant* $c \geq 0$ *for which* $|f(x)| \leq c$ *for all* $x \in I \setminus N$*, where* N *is a set of measure zero; and* L^∞ *is equipped with the norm*

$$\|f\|_{L^\infty} := \operatorname{supess}_I |f(x)|$$

that is the lowest possible constant c.

If $p = 2$, L^2 is a Hilbert space by inner product

$$\langle f, g\rangle := \int_I f(x) \cdot g(x)\, \mathrm{d}x \ .$$

Note that, in these spaces, by definition, $f = g$ iff the set $\{f \neq g\}$ has *Lebesgue measure* zero.

3.2.2 Fréchet space

The following citations Hamilton [1982] are referred to the first part of Hamilton's 1982 survey on the Nash&Moser theorem.

Definition 3.10 *A* ***Fréchet space*** *E is a* complete Hausdorff metrizable *l.c.t.v.s.; where we define that the l.c.t.v.s. E is*

complete *when, for any sequence (c_n), the fact that*

$$\lim_{m,n\to\infty} \|c_m - c_n\|_k = 0$$

for all k implies that c_n converges;

Hausdorff *when, for any given c, if $\|c\|_k = 0$ for all k then $c = 0$;*

metrizable *when there are countably many seminorms associated to E.*

The reason for the last definition is that, if E is metrizable, then we can define a distance

$$d(x, y) := \sum_{k=0}^{\infty} \frac{2^{-k}\ \|x - y\|_k}{1 + \|x - y\|_k}$$

that generates the same topology as the family of seminorms $\|\cdot\|_k$; and the vice versa is true as well, see Rudin [1973].

3.2.3 More examples of spaces of functions

Example 3.11 *Let $I \subset \mathbb{R}^m$ be open and non empty.*

- *The* ***Banach space*** *$C^j(I \to \mathbb{R}^n)$, with associated norm*

$$\|f\| := \sup_{i\le j} \sup_{t\in I} |f^{(i)}(t)| \ ;$$

where $f^{(i)}$ is the i-th derivative. In this space $f_n \to f$ iff $f_n^{(i)} \to f^{(i)}$ uniformly for all $i \le j$.

- *The **Sobolev space** $H^j(I \to \mathbb{R}^n)$, with scalar product*

$$\langle f, g\rangle_{H^j} := \int_I f(t)\cdot g(t) + \cdots + f^{(j)}(t)\cdot g^{(j)}(t)\,\mathrm{d}t$$

  where $f^{(j)}$ is the j-th derivative[5].

- *The **Fréchet space** of **smooth functions** $C^\infty(I \to \mathbb{R}^n)$.*
  The seminorms are
$$\|f\|_k = \sup_{x\in I} |f^{(k)}(x)|$$
  where $f^{(k)}$ is the k-th derivative. In this space, $f_n \to f$ iff all the derivatives $f_n^{(k)}(x)$ converge to derivatives $f^{(k)}(x)$, and for any fixed k the convergence is uniform w.r.to $x \in I$.

This last is the strongest topology between the topologies usually associated to spaces of functions.

3.2.4 Dual spaces.

Definition 3.12 *Given a* l.c.t.v.s. *E, the **dual space** E^* is the space of all linear functions $L : E \to \mathbb{R}$.*

If E is a Banach space, it is easy to see that E^* is again a Banach space, with norm
$$\|L\|_{E^*} := \sup_{\|x\|_E \le 1} |Lx| \ .$$
The biggest problem when dealing with Fréchet spaces, is that the dual of a Fréchet space is not in general a Fréchet space, since it often fails to be metrizable. (In most cases, the duals of Fréchet spaces are "quite wide" spaces; a classical example being the dual elements of smooth functions, that are the **distributions**). So given F, G Fréchet spaces, we cannot easily work with *"the space $\mathcal{L}(F,G)$ of linear functions between F and G"*.

As a workaround, given an auxiliary space H, we will consider *"indexed families of linear maps"* $L : F \times H \to G$, where $L(\cdot, h)$ is linear, and L is jointly continuous; but we will <u>not</u> consider L as a map

$$\begin{array}{ll} h \mapsto & (f \mapsto L(f,h)) \\ H \to & \mathcal{L}(F,G) \end{array} \tag{3.1}$$

3.2.5 Gâteaux differential

An example is the *Gâteaux differential.*

[5] The derivatives are computed in distributional sense, and must exists as Lebesgue integrable functions.

Definition 3.13 *We say that a continuous map $P : U \to G$, where F, G are Fréchet spaces and $U \subset F$ is open, is* ***Gâteaux differentiable*** *if for any $h \in F$ the limit*

$$DP(f; h) := \lim_{t \to 0} \frac{P(f + th) - P(f)}{t} = [\partial_t P(f + th)]|_{t=0} \tag{3.2}$$

exists. The map $DP(f; \cdot) : F \to G$ is the ***Gâteaux differential****.*

Note that the Gâteaux differential may fail to be linear in h; indeed, any 1-homogeneous function P is Gâteaux differentiable at $f = 0$.

Definition 3.14 *We say that P is C^1 if $DP : U \times F \to G$ exists and is jointly continuous.*

(This is weaker than what is usually required in Banach spaces).

The basics of the usual calculus hold.

Theorem 3.15 ([Hamilton, 1982, Thm. 3.2.5]) *If P is C^1 then $DP(f; h)$ is linear in h.*

Theorem 3.16 (Chain rule [Hamilton, 1982, Thm. 3.3.4]) *If P, Q are C^1 then $P \circ Q$ is C^1 and*

$$D(P \circ Q)(f; h) = DP\big(Q(f); DQ(f; h)\big) \ .$$

Also, the **implicit function theorem** holds, in the form due to Nash&Moser: see again Hamilton [1982] for details.

3.2.6 Troubles in calculus

But some important parts of calculus are instead lost.

Example 3.17 *Suppose $P : U \subset F \to F$ is smooth, and consider the O.D.E.*

$$\frac{d}{dt} f = Pf$$

to be solved for a solution $f : (-\varepsilon, \varepsilon) \to F$.

- *if F is a Banach space, then, given initial condition $f(0) = x$, the solution will exist and be unique (for $\varepsilon > 0$ small enough);*
- *but if F is a Fréchet space, then f may fail to exist or be unique.*

See [Hamilton, 1982, Sec. 5.6]

A consequence (that we will discuss more later on) is that the exponential map in Riemannian manifolds may fail to be locally surjective. See [Hamilton, 1982, Sec. 5.5.2].

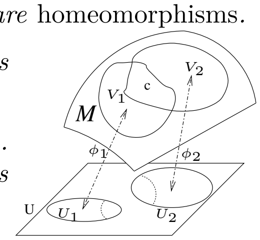
3.3 Manifold

To build a manifold of curves, we have ahead two main definitions of "manifold" to choose from.

Definition 3.18 (Differentiable Manifold) *An **abstract differentiable manifold** is a topological space M associated to a model* l.c.t.v.s. *U. It is equipped with an **atlas** of **charts** $\phi_k : U_k \to V_k$, where $U_k \subset U$ are open sets, and $V_k \subset M$ are open sets that cover M. The maps ϕ_k are* homeomorphisms.

The composition $\phi_1^{-1} \circ \phi_2$ restricted to $\phi_2^{-1}(V_1 \cap V_2)$ is usually required to be a smooth diffeomorphism.

*The **dimension** $\dim(M)$ of M is the dimension of U. When M is finite-dimensional, the model space (in this book) is always $U = \mathbb{R}^n$.*



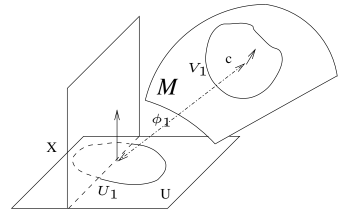
Since ϕ_k are homeomorphisms, then the topology of M is "identical" to the topology of U. The rôle of the charts is to define the differentiable structure of M mimicking that of U. See for example the definition of *directional derivative* in eqn. (3.4).

3.3.1 Submanifold

Definition 3.19 *Suppose A, B are open subsets of two linear spaces. A **diffeomorphism** is an invertible C^1 function $\phi : A \to B$, whose inverse ϕ^{-1} is again C^1.*

Let U be a fixed closed linear subspace of a *l.c.t.v.s.* X.

Definition 3.20 *A **submanifold** is a subset M of X, such that, at any point $c \in M$ we may find a **chart** $\phi_k : U_k \to V_k$, with $V_k, U_k \subset X$ open sets, $c \in V_k$. The maps ϕ_k are diffeomorphisms, and ϕ_k maps $U \cap U_k$ onto $M \cap V_k$.*



Most often, $M = \{\Phi(c) = 0\}$ where $\Phi : X \to Y$; so, to prove that M is a submanifold, we will use the *implicit function theorem*.

Note that M is itself an abstract manifold, and the model space is U; so $\dim(M) = \dim(U) \leq \dim(X)$. Vice versa any abstract manifold is a submanifold of some large X (by well known *embedding theorems*).

3.4 Tangent space and tangent bundle

Let us fix a differentiable manifold M, and $c \in M$. We want to define the **tangent space** T_cM of M at c.

Definition 3.21
- *The **tangent space** T_cM is more easily described for submanifolds; in this case, we choose a chart ϕ_k and a point $x \in U_k$ s.t.*

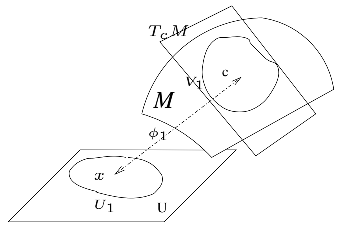


Figure 5: Tangent space

$\phi_k(x) = c$; T_cM is the image of the linear space U under the derivative $D_x\phi_k$. T_cM is itself a linear subspace in X. In figure 5, we graphically represent T_cM though as an affine subspace, by translating it to the point c.

- *The* ***tangent bundle*** *TM is the collection of all tangent spaces. If M is a submanifold of the vector space X, then*

$$TM := \{(c, h) \mid c \in M, h \in T_cM\} \subset X \times X \ .$$

The tangent bundle TM is itself a differentiable manifold; its charts are of the form $(\phi_k, D\phi_k)$, where ϕ_k are the charts for M.

3.5 Fréchet Manifold

When studying the space of all curves, we will deal with **Fréchet manifolds**, where the model space E will be a Fréchet space, and the composition of local charts $\phi_1^{-1} \circ \phi_2$ is smooth.

Some objects that we may find useful in the following are Fréchet manifolds.

Example 3.22 (Examples of Fréchet manifolds) *Given two finite-dimensional manifolds S, R, with S compact and n-dimensional,*

- *[Hamilton, 1982, Example 4.1.3]. the space $C^\infty(S; R)$ of smooth maps $f : S \to R$ is a Fréchet manifold. It is modeled on $U = C^\infty(\mathbb{R}^n; R)$.*
- *[Hamilton, 1982, Example 4.4.6]. The space of smooth diffeomorphisms* $\mathrm{Diff}(S)$ *of S onto itself is a Fréchet manifold. The group operation $\phi, \psi \to \phi \circ \psi$ is smooth.*
- *But if we try to model* $\mathrm{Diff}(S)$ *on $C^k(S; S)$, then the group operation is not even differentiable.*
- *[Hamilton, 1982, Example 4.6.6]. The quotient of the two above*

$$C^\infty(S; R)/\mathrm{Diff}(S)$$

is a Fréchet manifold. It contains "all smooth maps from S to R, up to a diffeomorphism of S".

So the theory of Fréchet space seems apt to define and operate on the *manifold of geometric curves.*

3.6 Riemann & Finsler geometry

We first define *Riemannian geometries*, and then we generalize to *Finsler geometries.*

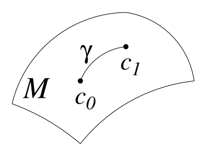
3.6.1 Riemann metric, length

Definition 3.23 • *A* ***Riemannian metric*** *G on a differentiable manifold M defines a* ***scalar product*** *$\langle h_1, h_2 \rangle_{G_{|c}}$ on $h_1, h_2 \in T_c M$, dependent on the point $c \in M$. We assume that the scalar product varies smoothly w.r.to c.*

- *The scalar product defines the* ***norm*** *$|h|_c = |h|_{G_{|c}} = \sqrt{\langle h, h \rangle}_{G_{|c}}$.*

Suppose $\gamma : [0,1] \to M$ is a path connecting c_0 to c_1.

- The **length** is $\quad \mathrm{Len}(\gamma) := \int_0^1 |\dot\gamma(v)|_{\gamma(v)} dv$
  where $\dot\gamma(v) := \partial_v \gamma(v)$.



- The **energy** (or **action**) is $\mathbb{E}(\gamma) := \int_0^1 |\dot\gamma(v)|^2_{\gamma(v)} \, \mathrm{d}v$

3.6.2 Finsler metric, length

Definition 3.24 *We define a* ***Finsler metric*** *to be a function $F : TM \to \mathbb{R}^+$, such that*

- *F is continuous and,*
- *for all $c \in M$, $v \mapsto F(c, v)$ is a norm on $T_c M$.*

We will sometimes write $|v|_c := F(c, v)$.

(Sometimes F is called a "Minkowsky norm").

As for the case of norms, a *Finsler metric* may be asymmetric; but, for sake of simplicity, we will only consider the symmetric case.

Using the norm $|v|_c$ we can then again define the length of paths by

$$\mathrm{Len}^F(\gamma) = \int_0^1 |\dot\gamma(t)|_{\gamma(t)} \, \mathrm{d}t = \int_0^1 F(\gamma(t), \dot\gamma(t)) \, \mathrm{d}t$$

and similarly the action.

3.6.3 Distance

Definition 3.25 *The* ***distance*** $d(c_0, c_1)$ *is the infimum of* $\mathrm{Len}(\gamma)$ *between all* C^1 *paths* γ *connecting* $c_0, c_1 \in M$.

Remark 3.26 *In the following chapter, we will define some differentiable manifolds* M *of curves, and add a Riemann (or Finsler) metric* G *on those; there are two different choices for the model space,*

- *suppose we model the differentiable manifold* M *on a Hilbert space* U*, with scalar product* $\langle , \rangle_U$*; this implies that* M *has a topology* τ *associated to it, and this topology, through the charts* ϕ*, is the same as that of* U*. Let now* G *be a Riemannian metric; since the derivative of a chart* $D\phi(c)$ *maps* U *onto* T_cM*, one natural hypothesis will be to assume that* $\langle , \rangle_U$ *and* $\langle , \rangle_{G,c}$ *be locally equivalent (uniformly w.r.to small movements of c); as a consequence, the topology generated by the Riemannian distance* d *will coincide with the original topology* τ*. A similar request will hold for the case of a Finsler metric* G*, in this case* U *will be a Banach space with a norm equivalent to that defined by* G *on* T_cM*.*

- *We will though find out that, for technical reasons, we will initially model the spaces of curves on the Fréchet space* C^∞*; but in this case there cannot be a norm on* T_cM *that generates the same original topology (for the proof, see I.1.46 in Rudin [1973]).*

3.6.4 Minimal geodesics

Definition 3.27 *If there is a path* γ^* *providing the minimum of* $\mathrm{Len}(\gamma)$ *between all paths connecting* $c_0, c_1 \in M$*, then* γ^* *is called a* ***minimal geodesic****.*

The *minimal geodesic* is also the minimum of the action (up to reparameterization).

Proposition 3.28 *Let* ξ^* *provide the minimum* $\min_\gamma \mathbb{E}(\gamma)$ *in the class of all paths* γ *in* M *connecting* x *to* y*. Then* ξ^* *is a* minimal geodesic *and its speed* $|\dot{\xi}^*|$ *is constant.*

Vice versa, let γ^* *be a minimal geodesic, then there is a reparameterization* $\xi^* = \gamma^* \circ \phi$ *s.t.* ξ^* *provides the minimum* $\min_\gamma \mathbb{E}(\gamma)$*.*

A proof may be found in Mennucci [2004].

3.6.5 Exponential map

The *action* $\mathbb{E}$ is a smooth integral, quadratic in $\dot{\gamma}$, and we can compute the Euler-Lagrange equations; its minima are more regular, since they are guaranteed to have constant speed; consequently, when trying to find geodesics, we will try to minimize the *action* and not the *length*. This also related to the idea of the exponential map.

Definition 3.29 *Let $\ddot\gamma = \Gamma(\dot\gamma, \gamma)$ be the Euler-Lagrange ODE of the action $\mathbb{E}(\gamma) = \int_0^1 |\dot\gamma(v)|^2 \, \mathrm{d}v$. Any solution of this ODE is a* ***critical geodesic****. Note that any* minimal geodesic *is a* critical geodesic*.*

Define the ***exponential map*** $\exp_c : T_cM \to M$ *as* $\exp_c(\eta) = \gamma(1)$*, where* γ *is the solution of*

$$\begin{cases} \ddot\gamma(v) = \Gamma(\dot\gamma(v), \gamma(v)), \quad \gamma(0) = c, \quad \dot\gamma(0) = \eta \end{cases} \tag{3.3}$$

Solving the above ODE (3.3) is informally known as **shooting geodesics**.

The exponential map is often used as a chart, since it is the least possibly deforming map at the origin.

Theorem 3.30 *Suppose that M is a smooth differentiable manifold modeled on a Hilbert space with a smooth Riemannian metric. The derivative of the exponential map* $\exp_c : T_cM \to M$ *at the origin is an isometry, hence* $\exp_c$ *is a local diffeomorphism between a neighborhood of* $0 \in T_cM$ *and a neighborhood of* $c \in M$*.*

(See Lang [1999], VIII §5). The exponential map can then "linearize" small portions of M, and so it will enable us to use linear methods such as the *principal component analysis*. Unfortunately, the above result does not hold if M is modeled on a Fréchet space.

3.6.6 Hopf–Rinow theorem

Theorem 3.31 (Hopf–Rinow) *Suppose M is a* ***finite*** *dimensional Riemannian or Finsler manifold. The following statements are equivalent:*

- (M, d) *is metrically complete;*
- *the O.D.E. (3.3) can be solved for all $c \in M$, $\eta \in T_cM$ and $v \in \mathbb{R}$;*
- *the map* $\exp_c$ *is surjective;*

and all those imply that, $\forall x, y \in M$ there exists a minimal geodesic connecting x to y.

3.6.7 Drawbacks in infinite dimensions

In a certain sense, infinite dimensional manifolds are simpler than their corresponding finite-dimensional counterparts: indeed, by Eells and Elworthy [1970],

Theorem 3.32 (Eells–Elworthy) *Any smooth differentiable manifold M modeled on an infinite dimensional separable Hilbert space H may be embedded as an open subset of that Hilbert space.*

In other words, it is always possible to express M using one single chart. (But note that this may not be the best way for computations/applications).

When M is an infinite dimensional Riemannian manifold, though, only a small part of the Hopf–Rinow theorem still holds.

Proposition 3.33 *Suppose M is **infinite** dimensional, modeled on a Hilbert space, and (M, d) is complete, then the O.D.E. (3.3) of a* critical geodesic *can be solved for all $v \in \mathbb{R}$.*

But other implications fails.

Example 3.34 (Atkin [1975]) *There exists an infinite dimensional complete Hilbert smooth manifold M and $x, y \in M$ such that there is no* critical geodesic *connecting x to y.*

That is,

- (M, d) is complete $\not\Rightarrow$ $\exp_c$ is surjective,
- and (M, d) is complete $\not\Rightarrow$ minimal geodesics exist.

It is then, in general, quite difficult to prove that an infinite dimensional manifold admits minimal geodesics (even when it is known to be metrically complete). There are though some positive results, such as Ekeland [1978] (that we cannot properly discuss for lack of space); or the following.

Theorem 3.35 (Cartan–Hadamard) *Suppose that M is connected, simply connected and has seminegative sectional curvature; then these are equivalent:*

- (M, d) *is complete*
- *there exists a $c \in M$ such that the map $\eta \to \exp_c(\eta)$ is well defined*

and then there exists an unique minimal geodesic connecting any two points.

For the proof, see Corollary 3.9 and 3.11 in sec. IX.§3 in Lang [1999].

3.7 Gradient in abstract differentiable manifold

Let M be a differentiable manifold, and U the model l.c.t.v.s. . Choose $c \in M$, and $\phi_1 : U_1 \to V_1$ a chart with $c \in V_1$, where V_1 is an open subset of V.

Definition 3.36 *Let $E : M \to \mathbb{R}$ be a functional; we say that it is **Gâteaux differentiable** at c if for all $h \in T_c M$ there exists the **directional derivative** at c in direction h*

$$DE(c; h) := \lim_{t \to 0} \frac{E(\phi_1(x + tk)) - E(c)}{t} \quad . \tag{3.4}$$

where $c = \phi_1(x)$, $h = D\phi_1(x; k)$.

We suppose that E is C^1 near c; this is expressed using the chart, imposing that $DE\big(\phi_1(x_1), D\phi_1(x_1; h)\big)$ is continuous, for $x_1 \in U_1$ and $h \in U$. By theorem 3.15, this implies that $DE(c; \cdot)$ is a linear function from $T_c M$ into $\mathbb{R}$; for this reason it is considered an element of the **cotangent space** $T_c^* M$ (that is the dual space of $T_c M$ — and we will not discuss here further).

Suppose now that we add a Riemannian geometry to M; this defines an inner product $\langle \cdot, \cdot \rangle_c$ on $T_c M$, so we can then define the gradient.

Definition 3.37 (Gradient) *The* ***gradient*** $\nabla E(c)$ *of* E *at* c *is the unique vector* $v \in T_c M$ *such that*

$$\langle v, h \rangle_c = DE(c; h) \quad \forall h \in T_c M \ .$$

If M is modeled on a Hilbert space H, and the inner product $\langle \cdot, \cdot \rangle_c$ used on $T_c M$ is equivalent to the inner product in H (as we discussed in 3.26), then the above equation uniquely defines what the gradient is. When M is modeled on a Fréchet space, though, there is no choice of inner product that is "compatible" with M; and there are pathological situations where the gradient does not exist.

Example 3.38 *Let* $M = C^\infty([-1, 1] \to \mathbb{R})$*; since* M *is a vector space, then it is trivially a manifold, with a single identity chart; and* $T_c M = M$*. Let* $E : M \to \mathbb{R}$ *be the* ***evaluation functional*** $E(f) = f(0)$*, then* $DE(f; h) = h(0)$*; let*

$$\langle f, g \rangle = \int_{-1}^{1} f(t) g(t) \, \mathrm{d}t \ ;$$

then the gradient of E *would be* δ_0 *(the* Dirac's delta*), that is a* distribution *(or more simply a* probability measure*) but not an element of* M*.*

3.8 Group actions

Definition 3.39 *Let* $\mathcal{G}$ *be a group,* M *a set. We say that* $\mathcal{G}$ ***acts on*** M *if there is a map*

$$\begin{array}{ccc} \mathcal{G} \times M & \to & M \\ g, m & \mapsto & g \cdot m \end{array}$$

that respects the group operations:

- *if* e *is the identity element then* $e \cdot m = m$*, and*
- *for any* $g, h \in \mathcal{G}, m \in M$

$$h \cdot (g \cdot m) = (hg) \cdot m \ . \tag{3.5}$$

(where hg *is the product of the two elements in the group* G*).*

If $\mathcal{G}$ is topological group, that is, a group with a topology such that the group operation is continuous, and M has a topology as well, we will require that the action be continuous. Similarly for smooth actions between smooth manifolds.

The action of a group $\mathcal{G}$ on M induces an **equivalence relation** $\sim$, defined by

$$m_1 \sim m_2 \text{ iff there exists } g \text{ such that } m_1 = g \cdot m_2 .$$

We can then define the following objects.

Definition 3.40

- *The **orbit** $[m]$ of m is the family of all m_1 equivalent to m.*
- *The **quotient** $M/\mathcal{G}$ is the space $M/\sim$ of equivalence classes.*
- *There is a **projection** $\pi : M \to M/\mathcal{G}$, sending each m to its orbit $\pi(m) = [m]$.*

We will see many examples of actions in Example 4.8.

3.8.1 Distances and groups

Let $d_M(x, y)$ be a distance on a space M, and $\mathcal{G}$ a group acting on M; we may think of defining a distance on $B = M/\mathcal{G}$ by

$$d_B([x], [y]) := \inf_{x\in[x],y\in[y]} d_M(x, y) = \inf_{g,h\in\mathcal{G}} d_M(g \cdot x, h \cdot y) \tag{3.6}$$

that is the lowest distance between two orbits. Unfortunately, this definition does not in general satisfy the triangle inequality.

Proposition 3.41 *If d_M is **invariant w.r.to the action of the group** $\mathcal{G}$, that is*

$$d_M(g \cdot x, g \cdot y) = d_M(x, y) \quad \forall g \in \mathcal{G} \ ,$$

then the above (3.6) can be simplified to

$$d_B([x], [y]) = \inf_{g\in\mathcal{G}} d_M(g \cdot x, y) \ . \tag{3.7}$$

and d_B is a semidistance.

Proof. We write $d_B(x, y)$ instead of $d_B([x], [y])$ for simplicity. It is easy to check that

$$d_B(x, y) = \inf_{g\in\mathcal{G}} d_M(g \cdot x, y) = \inf_{g\in\mathcal{G}} d_M(y, g \cdot x) = \inf_{g\in\mathcal{G}} d_M(g^{-1} \cdot y, x) = d_B(y, x) \ .$$

For the triangle inequality,

$$d_B(x, z) + d_B(z, y) = \inf_{g\in\mathcal{G}} d_M(g \cdot x, z) + \inf_{h\in\mathcal{G}} d_M(z, h \cdot y) =$$
$$= \inf_{g,h\in\mathcal{G}} d_M(g \cdot x, z) + d_M(z, h \cdot y) \geq \inf_{g,h\in\mathcal{G}} d_M(g \cdot x, h \cdot y) = d_B(x, y)$$

□

Remark 3.42 *There is a main problem: is $d_B([x], [y]) > 0$ when $[x] \neq [y]$? That is, there is no guarantee, in line of principle, that the above definition (3.7) won't simply result in $d_B \equiv 0$.*

4 Spaces and metrics of curves

In this section we review the mathematical definitions regarding the space of curves in full detail, and express a set of mathematical goals for the theory.

4.1 Classes of curves

Remember that $S^1 = \{x \in \mathbb{R}^2 \mid |x| = 1\}$ is the circle in the plane.

We will often associate $\mathbb{R}^2 = \mathbb{C}$, for convenience. Consequently,

- sometimes S^1 will be identified with $\mathbb{R}/(2\pi)$ (that is $\mathbb{R}$ modulus 2π translations).
- but other times (and in particular if the curve is planar) we will associate $S^1 = \{e^{it}, t \in \mathbb{R}\} = \{z, |z| = 1\} \subset \mathbb{C}$.

We recall that a **curve** is a map $c : S^1 \to \mathbb{R}^n$. The **image**, a.k.a. **trace**, of the curve is $c(S^1) = \{c(\theta), \theta \in S^1\}$.

Definition 4.1 (Classes of Curves)

- $\mathrm{Imm}(S^1, \mathbb{R}^n)$ *is the class of* ***immersed curves*** *c, such that $c' \neq 0$ at all points.*

- $\mathrm{Imm}_f(S^1, \mathbb{R}^n)$ *is the class of* ***freely immersed curve****, the immersed curves c such that, moreover, if $\phi : S^1 \to S^1$ is a diffeomorphism and $c(\phi(t)) = c(t)$ for all t, then $\phi =$Id. So, in a sense, the curve is "completely" characterized by its image.*
- $\mathrm{Emb}(S^1, \mathbb{R}^n)$ *are the* ***embedded curves****, maps c that are diffeomorphic onto their image $c(S^1)$; and the image is an embedded submanifold of $\mathbb{R}^n$ of dimension 1.*

Each class contains the one following it. The following is an example of a non-freely immersed curve.

Example 4.2 *We define the* ***doubly traversed circle*** *using the complex notation $c(z) = z^2$ for $z \in S^1 \subset \mathbb{C}$; or otherwise identifying $S^1 = \mathbb{R}/(2\pi)$, and in this case*

$$c(\theta) = (\cos(2\theta), \sin(2\theta))$$

for $\theta \in \mathbb{R}/(2\pi)$. Setting $\phi(t) = t + \pi$, we have that $c = c \circ \phi$, so c is not freely immersed.

Vice versa, the following result is a sufficient condition to assert that a curve is freely immersed.

Proposition 4.3 (Michor and Mumford [2006]) *If c is immersed and there is a $x \in \mathbb{R}^n$ s.t. $c(t) = x$ for one and only one t, then c is freely immersed.*

Remark 4.4 $\operatorname{Imm}(S^1, \mathbb{R}^n)$ *,* $\operatorname{Imm}_f(S^1, \mathbb{R}^n)$ *,* $\operatorname{Emb}(S^1, \mathbb{R}^n)$ *, are open subset of the Banach space* $C^r(S^1 \to \mathbb{R}^n)$ *when* $r \geq 1$*, so they are trivially submanifolds (the charts are identity maps); and similarly for the Fréchet space* $C^\infty(S^1 \to \mathbb{R}^n)$*.*

$\operatorname{Imm}(S^1, \mathbb{R}^n)$ *,* $\operatorname{Imm}_f(S^1, \mathbb{R}^n)$ *,* $\operatorname{Emb}(S^1, \mathbb{R}^n)$ *are connected iff* $n \geq 3$*; whereas in the case* $n = 2$ *of planar curves, they are divided in connected components each containing curves with the same* rotation index *(see Prop. 8.1 for the definition).*

4.2 Gâteaux differentials in the space of immersed curves

Let once again M be the manifold of smooth immersed curves. We will mainly be interested in the Gâteaux derivation of operators O and functions $E : M \to \mathbb{R}^k$. By **operator** we mean any one of these possible operations.

- An operator
$$O : M \to T_c M$$
such that $O(c)$ is a smooth a vector field; examples are the operations of computing the tangent field T, and the curvature H.

- An operator
$$O : M \to M$$
such that $O(c)$ is an immersed curve; examples are all the group actions listed in Section 4.3

- An operator
$$O : M \to (T_c M \to T_c M)$$
such that $O(c)$ is itself an operator on vector fields along c; an example is the derivative D_s;

- or sometimes
$$O : M \to (T_c M \to \mathbb{R}^n)$$
such as the average along the curve $O(c) = \operatorname{avg}_c(\cdot)$.

In all above cases, the evaluated operator $O(c)$ is itself a function; for this reason, to avoid confusion in the notation, in this section we write the Gâteaux derivation as $D_{c,h}O$ instead of $DO(c; h)$.

The charts in M are trivial (as noted in 4.4), so the definition (3.4) of *Gâteaux differential* simplifies to

$$D_{c,h}E := \partial_t E(c + th)\big|_{t=0} \tag{4.1}$$

and similarly for an operator.

These rules following are the building blocks that are used (by means of standard calculus rules for derivatives) to compute the derivation of all other geometric operators usually found in computer vision applications.

Proposition 4.5

$$
\begin{aligned}
D_{c,h}\,\mathrm{len}(c) &= \int_c (D_s h \cdot D_s c)\,\mathrm{d}s = -\int_c (h \cdot D_s^2 c)\,\mathrm{d}s \quad , && (4.2)\\
D_{c,h}(D_s O) &= -(D_s h \cdot D_s c)(D_s O) + D_s(D_{c,h} O) \quad , && (4.3)\\
D_{c,h}\int_c O\,\mathrm{d}s &= \int_c D_{c,h} O + O.\,(D_s h \cdot D_s c)\,\mathrm{d}s \quad , && (4.4)\\
D_{c,h}\fint_c O\,\mathrm{d}s &= \fint_c D_{c,h} O + O.(D_s h \cdot D_s c)\,\mathrm{d}s - \fint_c O\,\mathrm{d}s \fint_c (D_s h \cdot D_s c)\,\mathrm{d}s && (4.5)
\end{aligned}
$$

The proof is by direct computation.

For example, from (4.5) we easily obtain (2.7), that is

$$
D_{c,h}\mathrm{avg}_c(c) = \fint_c h + (c - \mathrm{avg}_c(c))(D_s h \cdot D_s c)\,\mathrm{d}s \ .
$$

If $C(t,\theta)$ is a homotopy, we can obtain a different interpretation of all previous equalities substituting formally C for O, ∂_t for $D_{c,h}$ and eventually $\partial_t C$ for h.

The above proposition helps in formalizing and speeding up calculus on geometric quantities, as in this example.

Example 4.6 *We can Gâteaux-derive the "second arc parameter derivative operator D_{ss}" by repeated applications of (4.3), as follows*

$$
\begin{aligned}
D_{c,h}(D_{ss} g) &= -(D_s h \cdot D_s c) D_{ss} g + D_s[D_{c,h}(D_s g)] = \\
&= -(D_s h \cdot D_s c) D_{ss} g + D_s[-(D_s h \cdot D_s c)(D_s g) + D_s(D_{c,h} g)] = \\
&= D_{ss}(D_{c,h} g) - 2(D_s h \cdot D_s c) D_{ss} g - [D_s(D_s h \cdot D_s c)](D_s g)
\end{aligned}
$$

and hence obtain ***the Gâteaux differential of the curvature***

$$
D_{c,h} D_{ss} c = D_{ss} h - 2(D_s h \cdot D_s c) D_{ss} c - [D_s(D_s h \cdot D_s c)] D_s c
$$

and eventually obtain ***the Gâteaux differential of the elastica energy*** *by substituting this last equality in (4.4)*

$$
\begin{aligned}
D_{c,h}\int_c |D_{ss} c|^2\,\mathrm{d}s &= \int_c 2 D_{ss} c \cdot (D_{c,h} D_{ss} c) + |D_{ss} c|^2.\,(D_s h \cdot D_s c)\,\mathrm{d}s = \\
&= \int_c 2 D_{ss} c \cdot D_{ss} h - 3(D_s h \cdot D_s c)|D_{ss} c|^2\,\mathrm{d}s
\end{aligned}
$$

since $(D_{ss} c \cdot D_s c) = 0$.

Remark 4.7 *Most objects we deal with share an important property: they are* reparameterization invariant*. In the case of operators, in all cases in which*

$O(c)$ is a curve or a vector field, then this means that for any diffeomorphism $\varphi \in \mathrm{Diff}^+(S^1)$ there holds

$$O(c \circ \phi)(\theta) = O(c)(\phi(\theta)) \ . \tag{4.6}$$

(whereas for $\varphi \in \mathrm{Diff}^-(S^1)$ there is a possible sign choice as explained in Definition 1.5).

By considering an infinitesimal perturbation $a : S^1 \to \mathbb{R}$ of the identity in Diff(S^1), we obtain the formula

$$D_{c,h}O(c)(\theta) = a(\theta) \,.\, \partial_\theta O(c)(\theta) \ \ \textit{for any} \ \ h(\theta) = a(\theta).\partial_\theta c(\theta) \tag{4.7}$$

and this last may be rewritten in arc parameter

$$D_{c,h}O(c)(s) = a(s) \,.\, D_s O(c)(s) \ \ \textit{for any} \ \ h(s) = a(s).D_s c(s) \ . \tag{4.8}$$

Similarly if $E : M \to \mathbb{R}$ is a reparameterization invariant *energy, then*

$$D_{c,h}E(c) = 0 \ \ \textit{for any} \ \ h = a.D_s c \ . \tag{4.9}$$

It is also possible to prove that the condition (4.8) is equivalent to saying that the operator is reparameterization invariant; and similarly for E.

4.3 Group actions on curves

Let M again be the manifold of immersed curves. An example of group action that we saw in Definition 1.3 is obtained when $\mathcal{G} = \mathrm{Diff}(S^1)$; $\mathcal{G}$ acts on curves by right composition

$$c, \phi \mapsto c \circ \phi$$

and this action is the *reparameterization*.

In general, all groups that act on $\mathbb{R}^n$ also act on M; many are of interest in *computer vision*. The action of these groups on curves is always of the form $c, A \mapsto A \circ c$, where $A : \mathbb{R}^n \to \mathbb{R}^n$ is the group action on $\mathbb{R}^n$.

Example 4.8

- *$O(n)$ is the group of* ***rotation**s in $\mathbb{R}^n$. It is represented by the group of orthogonal matrixes*

  $$O(n) := \{A \in \mathbb{R}^{n \times n} \mid AA^t = A^t A = Id\}$$

  The action on $v \in \mathbb{R}^n$ is the matrix·vector multiplication Av.

- *$SO(n)$ is the group of* ***special rotation**s in $\mathbb{R}^n$ with $\det(A) = 1$.*

- *$\mathbb{R}^n$ is the group of* ***translation**s in $\mathbb{R}^n$. The action on $v \in \mathbb{R}^n$ is the vector sum $v \mapsto v + T$.*

- *$E(n)$ is the* ***Euclidean group***, *generated by rotations and translations.*

- *$\mathbb{R}^+$ is the group of* ***rescaling**s.*

The *quotient spaces $M/\mathcal{G}$* are the spaces of *curves up to rotation and/or translations and/or scaling (...).*

4.4 Two categories of *shape spaces*

Let M be a manifold of curves. We should distinguish two different ideas of shape spaces of curves.

	geometric curves	**geometric curves up to pose**
The space of shapes	*curves up to reparameterization,*	*curves up to reparameterization, rotation, translation, scaling,*
is modeled as	$\boxed{M/\mathrm{Diff}(S^1)}$,	$\boxed{M/\mathrm{Diff}(S^1)/E(n)/\mathbb{R}^+}$,
is well-suited to	shape optimization.	shape analysis.
References:	*Michor and Mumford [2006], Yezzi-M.-Sundaramoorthi Sundaramoorthi et al. [2005, 2006, 2007a,b, 2008a,b] Mennucci et al. [2008] Glaunès et al. [2005]; Trouvé and Younes [2005]*	*Srivastava et al. [2005]; Mio and Srivastava [2004b]; Klassen et al. [2004], Younes [1998], Younes et al. [2008]*

The term *preshape space* is sometimes used for the leftmost space, when both spaces are studied in the same paper.

4.5 Geometric curves

Unfortunately the quotient

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

of *immersed curves up to reparameterization* is not a Fréchet manifold.

We (re)define the space of **geometric curves**.

Definition 4.9

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

is the quotient of $\mathrm{Imm}_f(S^1, \mathbb{R}^n)$ *(the free immersions) by the diffeomorphisms* $\mathrm{Diff}(S^1)$ *(that act as reparameterizations).*

The good news is that

Proposition 4.10 (§2.4.3 in Michor and Mumford [2006]) *If Imm_f and $Diff(S^1)$ have the topology of the Fréchet space of C^∞ functions, then $B_{i,f}$ is a Fréchet manifold modeled on C^∞.*

The bad news is that

- when we add a simple Riemannian metric to $B_{i,f}$, the resulting metric space is not metrically complete; indeed, there cannot be any norm on C^∞ that generates the same topology of the Fréchet space C^∞ (as we discussed in 3.26);

- by modeling $B_{i,f}$ as a Fréchet manifold, some calculus is lost, as we saw in Section 3.2.5.

Remark 4.11 *It seems that this is the only way to properly define the manifold. If otherwise we choose $M = C^k(S^1 \to \mathbb{R}^n)$ to be the manifold of curves, then if $c \in M$, $c' \notin T_c M$. Hence we (must?) model M on C^∞ functions.*

4.5.1 Research path

Following Michor and Mumford [2006] we so obtained a possible program of math research:

- define
$$B = B_{i,f}(S^1, \mathbb{R}^n) = \mathrm{Imm}_f(S^1, \mathbb{R}^n)/\mathrm{Diff}(S^1)$$
and consider B as a Fréchet manifold modeled on C^∞,
- define a Riemann/Finsler geometry on it, study its properties,
- metrically complete the space.

In the last step, we would hope to obtain a differentiable manifold; unfortunately, this is sometimes not true, as we will see in the overview of the literature.

4.6 Goals (revisited)

We formulate an abstract set of goal properties on a metric $\langle h_1, h_2 \rangle_{G_{|c}}$ on spaces of curves.

1. [**rescaling invariance**] For any $\lambda > 0$, if we rescale c to λc, then
$$\langle h_1, h_2 \rangle_{G_{|\lambda c}} = \lambda^a \langle h_1, h_2 \rangle_{G_{|c}}$$
(where $a \in \mathbb{R}$ is an universal constant);
2. [**Euclidean invariance**] Suppose that A is an Euclidean transformation, and R is its rotational part; if we apply A to c and R to h_0, h_1, then
$$\langle Rh_1, Rh_2 \rangle_{G_{|Ac}} = \langle h_1, h_2 \rangle_{G_{|c}} \ ;$$
3. [**parameterization invariance**]
the metric does not depend on the parameterization of the curve, that is $\|\tilde{h}\|_{\tilde{c}} = \|h\|_c$ when $\tilde{c}(t) = c(\varphi(t))$ and $\tilde{h}(t) = h(\varphi(t))$.

If a metric satisfies the above three properties, we say that it is a **geometric metric**.

4.6.1 Well posedness

We also add a more basic set of requirements.

0. [**well-posedness and existence of minimal geodesics**]

- The metric induces a *good* distance d: that is, the distance between different curves is positive, and d generates the same topology that the atlas of the manifold M induces;
- (M, d) is complete;
- for any two curves in M, there exists a minimal geodesic connecting them.

4.6.2 Are the goal properties consistent?

So we state the abstract problem:

Problem 4.12 *Consider the space of curves M, and the family of all Riemannian (or, Finsler) Geometries F on M.*

Does there exist a metric F satisfying the above properties 0,1,2,3?

Consider metrics F that may be written in integral form

$$F(c,h) = \int_c f\big(c(s), \partial_s c(s), \dots, \partial^j_{s^j} c(s),\, h(s), \dots, \partial^i_{s^i} h(s)\big)\, \mathrm{d}s$$

what is the relationship between the degrees i, j and the properties in this section?

All this boils down to a fundamental question: *can we design metrics to satisfy our needs?*

5 Representation/embedding/quotient in the current literature

5.1 The representation/embedding/quotient paradigm

A common way to model shapes is by **representation/embedding**:

- we **represent** the shape A by a function u_A
- and then we **embed** this representation in a space E, so that we can operate on the shapes A by operating on the representations u_A.

Most often, representation/embedding alone do not directly provide a satisfactory *shape space*. In particular, in many cases it happens that the representation is "redundant", that is, the same shape has many different possible representations. An appropriate **quotient** is then introduced.

There are many examples of *shape spaces* in the literature that are studied by means of the *representation/embedding/quotient* scheme. Understanding the basic math properties of these three operations is then a key step in understanding *shape space*s and designing/improving them.

We now present a rapid overview of how this scheme is exploited in the current literature of shape spaces; then we will come back to some of them to explain more in depth.

5.2 Current literature

Example 5.1 (The family of all non empty compact subsets of $\mathbb{R}^N$) *A standard representation is obtained by associating each closed subset A to the* distance function u_A *or the* signed distance function b_A *(that were defined in Definition 2.1). We may then define a* topology of shapes *by deciding that $A_n \to A$ when $u_{A_n} \to u_A$ uniformly on compact sets. This convergence coincides with the Kuratowski topology of closed sets; if we limit shapes to be compact sets, the Kuratowski topology is induced by the Hausdorff distance. See section 6.2.*

Example 5.2 *Trouvé-Younes* et al *(see Glaunès et al. [2005], Trouvé and Younes [2005] and references therein) modeled the motion of shapes by studying a left invariant Riemannian metric on the family $\mathcal{G}$ of diffeomorphisms of the space $\mathbb{R}^N$; to recover a true metric of shapes, a quotient is then performed w.r.to all diffeomorphisms $\mathcal{G}_0$ that do not move a template shape.*

But the *representation/embedding/quotient* scheme is also found when dealing with spaces of curves:

Example 5.3 (Representation by angle function) *In the work of Klassen et al. [2004]; Srivastava et al. [2005]; Mio and Srivastava [2004b], smooth planar closed curves $c : S^1 \to \mathbb{R}^2$ of length 2π are represented by a pair of angle-velocity functions $c'(u) = \exp(\phi(u) + i\alpha(u))$ (identifying $\mathbb{R}^2 = \mathbb{C}$) then (ϕ, α) are embedded as a subset N in $L^2(0, 2\pi)$ or $W^{1,2}(0, 2\pi)$. Since the goal is to obtain a* shape space *representation for shape analysis purposes, a quotient is then introduced on N. See Section 8.2.*

Example 5.4 *Another representation of planar curves for shape analysis is found in Younes [1998]. In this case the angle function is considered mod(π). This representation is both simple and very powerful at the same time. Indeed, it is possible to prove that **geodesics do exist and to explicitly show examples of geodesics**. See Section 8.3.*

Example 5.5 (Harmonic representation) *A. Duci* et al *(see Duci et al. [2003, 2006]) represent a closed planar contour as the zero level of a harmonic function. This novel representation for contours is explicitly designed to possess a linear structure, which greatly simplifies linear operations such as averaging, principal component analysis or differentiation in the space of shapes.*

And, of course, we have in this list the spaces of embedded curves.

Example 5.6 *When studying embedded curves, usually, for the sake of mathematical analysis, the curves are modeled as immersed parametric curves; a quotient w.r.to the group of possible reparameterizations of the curve c (that coincides with the group of diffeomorphisms* $\mathrm{Diff}(S^1)$*) is applied afterward to all the mathematical structures that are defined (such as the manifold of curves, the Riemannian metric, the induced distance, etc.).*

It is interesting to note this fact.

Remark 5.7 (A remark on the quotienting order) *When we started talking of geometric curves, we proposed the quotient $B = M/\mathrm{Diff}(S^1)$ (the space of* curves up to reparameterization*); and this had to be followed by other quotients w.r.to Euclidean motions and/or scaling, to obtain $M/\mathrm{Diff}(S^1)/E(n)/\mathbb{R}^+$. In practice, though, the space B happens to be more difficult to study; hence most* shape space *theories that deal with curves prefer a different order: the quotient $M/E(n)/\mathbb{R}^+$ is modeled and studied first; then the quotient $M/E(n)/\mathbb{R}^+/\mathrm{Diff}(S^1)$ is performed (often, only numerically).*

6 Metrics of sets

We now present two examples of Shape Theories where a shape may be a generic subset of the plane; with particular attention to how they behave w.r.to curves.

6.1 Some more math on distance and geodesics

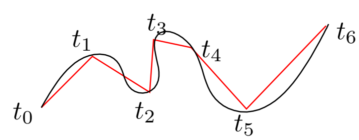
We start by reviewing some basic results in abstract metric spaces theory.

6.1.1 Length induced by a distance

In this subsection (M, d) is a generic metric space.

Definition 6.1 (Length by total variation) *Define the **length** of a continuous curve $\gamma : [\alpha, \beta] \to M$, by using the total variation*

$$\mathrm{Len}^d \gamma := \sup_T \sum_{i=1}^{n} d\big(\gamma(t_{i-1}), \gamma(t_i)\big) \ .$$



where the supremum is carried out over all finite subsets $T = \{t_0, \cdots, t_n\} \subset [\alpha, \beta]$ and $t_0 \le \cdots \le t_n$.

Definition 6.2 (The induced geodesic distance)

$$d^g(x, y) := \inf_\gamma \mathrm{Len}^d \gamma \tag{6.1}$$

the infimum is in the class of all continuous γ in M connecting x to y.

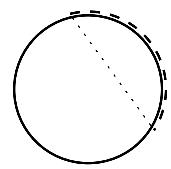
Note that $d^g(x, y) < \infty$ iff x, y may be connected by a Lipschitz path.

6.1.2 Minimal geodesics

Definition 6.3 *If in the definition (6.1) there is a curve γ^* providing the minimum, then γ^* is called a* ***minimal geodesic*** *connecting x to y.*

Proposition 6.4 *In Riemann and Finsler manifolds, the* integral length $\mathrm{Len}(\gamma)$ *that we defined in Section 3.6.1 coincides with the* total variation length $\mathrm{Len}^d\,\gamma$ *that we defined in Definition 6.1 on the previous page. As a consequence, $d = d^g$: we say that d is* ***path-metric****.*

In general, though, it is easy to devise examples where $d \neq d^g$.



Example 6.5 *The set $M = S^1 \subset \mathbb{R}^2$ is a metric space, if associated with $d(x,y) = |x-y|$ (that is represented as a dotted segment in the picture); in this case, $d^g(x,y) = |\arg(x) - \arg(y)|$ (that is represented as a dashed arc in the picture).*

6.1.3 Existence of geodesics and of average

Proposition 6.6 *If for a $\rho > 0$ the closed ball*

$$\mathbb{D}^g(x,\rho) := \{y \mid d^g(x,y) \leq \rho\}$$

is compact, then x and any $y \in \mathbb{D}^g$ may be connected by a geodesic.

Proposition 6.7 *If for a $x \in M$ and all $\rho > 0$, $\mathbb{D}^g(x,\rho)$ is compact, then the* distance-based average *(that was defined in 2.4) exists.*

The proofs may be found in Duci and Mennucci [2007].

6.1.4 Geodesic rays

More in general,

Definition 6.8 *a continuous curve $\gamma : I \to M$ (where $I \subset \mathbb{R}$ is an interval) is a* ***geodesic ray*** *if for each $t \in I$ there is a $\varepsilon > 0$ s.t. $J = [t-\varepsilon, t+\varepsilon] \subset I$ and γ restricted to J is the geodesic between $\gamma(t-\varepsilon)$ and $\gamma(t+\varepsilon)$.*

A critical geodesic in a Riemann smooth manifold (modeled on a Hilbert space) is always a *geodesic ray*, as in this example.

Example 6.9 *The multiply traversed full circle $\gamma(t) = (0, \cos(t), \sin(t))$ with $t \in \mathbb{R}$ is a* geodesic ray *in the sphere S^2.*

6.1.5 Hopf–Rinow , Cohn-Vossen Theorem

Definition 6.10

- *We recall that (M,d) is* ***path-metric*** *if $d = d^g$.*
- *Moreover, (M,d) is* ***locally compact*** *if small closed balls are compact.*

Theorem 6.11 (Hopf–Rinow , Cohn-Vossen) *Suppose* (M,d) *is locally compact and path-metric; the following statements are equivalent:*

- *for all* $x \in M, \rho > 0$*, the closed ball*

$$\mathbb{D}(x,\rho) := \{y \mid d(x,y) \le \rho\}$$

  is compact;

- (M,d) *is complete;*

- *any geodesic ray* $\gamma : [0,1) \to M$ *may be extended on* $[0,1]$*;*

and all imply that any two points may be connected by a minimal geodesic.

Note that the above theorem cannot be used in infinite dimensional differentiable manifolds, since the the closed balls are never compact in that case (see Theorems 1.21, 1.22 in Chapter I in Rudin [1973]).

6.2 Hausdorff metric

Let Ξ be the collection of all compact subsets of $\mathbb{R}^N$. (This is sometimes called the "*topological hyperspace*" of $\mathbb{R}^N$). Let $A, B \subset \mathbb{R}^N$ compact non-empty. We already defined the **distance function** $u_A(x) := \inf_{y\in A} |x-y|$, and the Hausdorff distance of two sets A, B as

$$d_H(A,B) := \Big(\sup_{x\in A} u_B(x)\Big) \vee \Big(\sup_{x\in B} u_A(x)\Big) \quad .$$

We now list some known properties.

Properties 6.12

- $d_H(A,B) = \sup_{x\in\mathbb{R}^N} |u_A(x) - u_B(x)| \quad .$

- (Ξ, d_H) *is path-metric;*

- *each family of compact sets that are contained in a fixed large closed ball in* $\mathbb{R}^N$ *is compact in* (Ξ, d_H)*; so*

- *any closed ball* $\mathbb{D}(A,\rho) := \{B \mid d_H(A,B) \le \rho\}$ *is compact in* (Ξ, d_H)*, and moreover*

- *by Theorem 6.11, minimal geodesics exist in* (Ξ, d_H)*.*

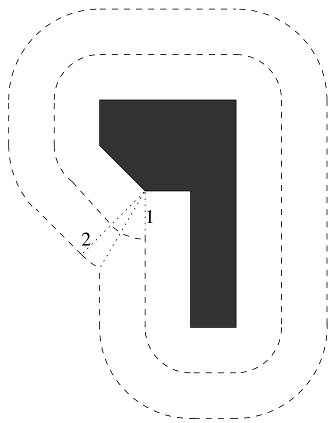


Figure 6: Fattening of a set

6.2.1 An alternative definition

Let $D_r(x)$ be the closed ball of center x and radius $r > 0$ in $\mathbb{R}^N$, and $D_r = D_r(0)$. We define the **fattened set** to be

$$A + D_r := \{x + y \mid x \in A, |y| \le r\} = \bigcup_{x \in A} D_r(x) = \{y \mid u_A(y) \le r\}.$$

Note that the fattened set is always closed, (since the distance function $u_A(x)$ is continuous).

Example 6.13 *In figure 6 we see an example of a set A fattened to $r = 1, 2$; the set A is the black polygon (and is filled in), whereas the dashed lines in the drawing are the contours of the fattened sets.* [6]

We can then state the following equivalent definition of the *Hausdorff distance*:

$$d_H(A, B) = \min\{\delta > 0 \mid A \subset (B + D_\delta), B \subset (A + D_\delta)\} \quad .$$

6.2.2 Uncountable many geodesics

Unfortunately d_H is quite “unsmooth”. There are choices of A, B compact that may be joined by an uncountable number of geodesics. In fact we can consider this simple example.

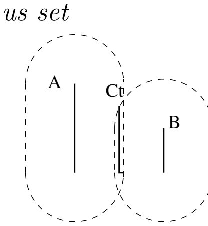
Example 6.14 (Duci and Mennucci [2007]) *Let us set*

$$A = \{x = 0, 0 \le y \le 2\}$$

$$B = \{x = 2, 0 \le y \le 1\}$$

$$C_t = \left\{x = 1, 0 \le y \le \tfrac{3}{2}\right\} \cup \{y = 0, 1 \le x \le t\}$$

$$\textit{with } 1 \le t \le \sqrt{5}/2;$$



[6]The fattened sets are not drawn filled — otherwise they would cover A.

and in the picture we represent (dashed) the fattened sets $A + D_{\sqrt{5}/2}$ and $B + D_{\sqrt{5}/2}$. Note that $d_H(A, B) = \sqrt{5}$ while $d_H(A, C_t) = d_H(B, C_t) = \sqrt{5}/2$: so C_t are all midpoints that are on different geodesics between A and B.

6.2.3 Curves and connected sets

Let again Ξ be the collection of all compact subsets of $\mathbb{R}^N$. Let Ξ_c be the subclass of compact *connected* subsets of $\mathbb{R}^N$. We now relate the space of curves to this metric space, by listing these properties and remarks.

- Ξ_c is a closed subset of (Ξ, d_H);
- Ξ_c is the closure in (Ξ, d_H) of the class of (images of) all embedded curves.
- Ξ_c is Lipschitz-path-connected [7];
- for all above reasons, it is possible to connect any two $A, B \in \Xi_c$ by a minimal geodesic moving in Ξ_c.
- So, if we try to find a minimal geodesic connecting two curves using the metric d_H, we will end up finding a geodesic in (Ξ_c, d_H); and similarly if we try to optimize an energy defined on curves.
- But note that Ξ_c is not geodesically convex in Ξ, that is, there exist two points $A, B \in \Xi_c$ and a minimal geodesics ξ connecting A to B in the metric space (Ξ, d), such that the image of ξ is not contained inside Ξ_c.
- We don't know if (Ξ_c, d_H) is path-metric.

6.2.4 Applications in *computer vision*

Charpiat et al. [2004] propose an approximation method to compute $\mathrm{len}^H(\xi)$ by means of a family of energies defined using a smooth integrand; the approximation is mainly based on the property $\|f\|_{L^p} \to_p \|f\|_{L^\infty}$, for any measurable function f defined on a bounded domain; they successively devise a method to find approximation of geodesics.

6.3 A Hausdorff-like distance of compact sets

In Duci and Mennucci [2007] a L^p-like distance on the compact subsets of $\mathbb{R}^N$ was proposed. (Here $p \in [1, \infty)$.)

To this end, we fix $\varphi : [0, \infty) \to (0, \infty)$, decreasing, C^1, with $\varphi(|x|) \in L^p$. We then define $v_A(x) := \varphi(u_A(x))$, where u_A is the *distance function*. We eventually define the distance

$$d(A, B) := \|v_A - v_B\|_{L^p} \quad . \tag{6.2}$$

[7] That is, any $A, B \in \Xi_c$ can be connected by a Lipschitz arc $\gamma : [0, 1] \to \Xi_c$

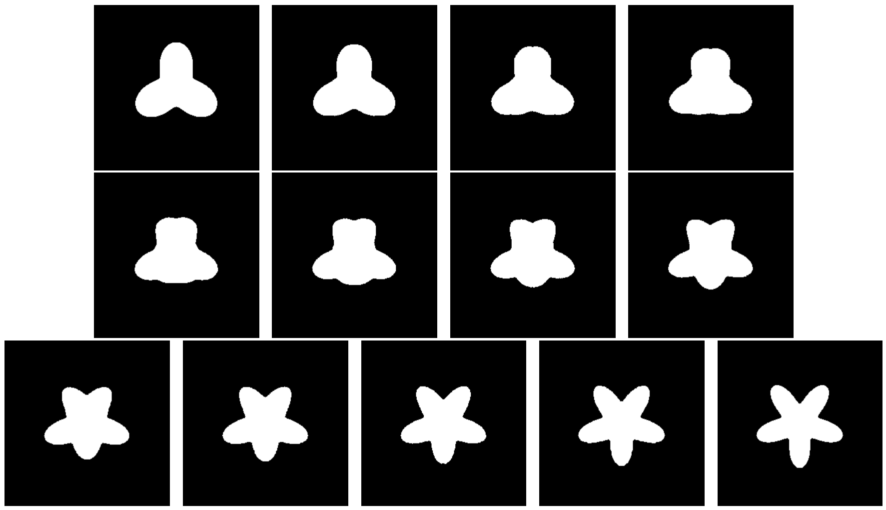


Figure 7: Example of a minimal geodesic

Remark 6.15 *This* shape space *is a perfect example of the representation/ embedding/quotient scheme. Indeed, this* shape space *is represented as* $N_c = \{v_A \mid A \text{ compact}\}$ *and embedded in* L^p*. Given* $v \in N_c$*, we recover the shape* $A = \{v = \varphi^{-1}(0)\}$ *(that is a* level set *of* v*).*

Example 6.16 *A simple example (that works for all* p*) is given by* $\varphi(t) = e^{-t}$*, so that* $v_A(x) = \exp(-u_A(x))$*; in this case,* $A = \{v = 1\}$*.*

The distance d of eqn. (6.2) enjoys the following properties.

Properties 6.17

- *It is Euclidean invariant;*
- *it is locally compact but not path-metric;*
- *the topology induced is the same as that induced by* d_H*;*
- *minimal geodesics do exist, since* $\mathbb{D}^g := \{B \mid d^g(A, B) \le \rho\}$ *is compact.*

The proofs are in Duci and Mennucci [2007]. We present a numerical computation of a minimal geodesic (by A. Duci) in Figure 7.

6.3.1 Analogy with the Hausdorff metric, L^p *vs* L^∞

We recall that $d_H(A, B) = \|u_A(x) - u_B(x)\|_{L^\infty}$; whereas instead now we are proposing $d(A, B) := \|v_A - v_B\|_{L^p}$. The idea being that this distance of compact sets is modeled on L^p, whereas the Hausdorff distance is "modeled" on L^∞. (Note that the Hausdorff distance is not really obtained by embedding, since $u_A \notin L^\infty$). L^p is more regular than L^∞, as shown by this remark.

Remark 6.18 *Given any $f, g \in L^p$ with $p \in (1, \infty)$, the segment connecting f to g is the unique minimal geodesic connecting them. Suppose now that the dimension of $L^\infty(\Omega, \mathcal{A}, \mu)$ is greater than 1. Given generically $f, g \in L^\infty$, there is an uncountable number of minimal geodesics connecting them.*[8]

(For the proof see 2.11 & 2.13 in Duci and Mennucci [2007]). The above result suggests that it may be possible to shoot geodesics in the "Riemannian metric" associated to this distance.

6.3.2 Contingent cone

Let again $N_c = \{v_A \mid A \text{ compact}\}$ be the family of all representations. Given a $v \in N_c$, let $T_v N_c \subset L^p$ be the **contingent cone**

$$\begin{aligned} T_v N_c \quad &:= \quad \{\lim_n t_n(v_n - v) \mid t_n > 0, v_n \in N_c, v_n \to v\} \\ &= \quad \left\{ \lambda \lim_n \frac{v_n - v}{\|v_n - v\|_{L^p}} \mid \lambda \geq 0, v_n \to v \right\} \quad , \end{aligned}$$

where it is intended that the above limits are in the sense of strong convergence in L^p.

$T_v N_c$ contains all directions in which it is possible *"in the L^p sense"* to infinitesimally deform a compact set. For example, if $\Phi(x,t) : \mathbb{R}^N \times (-\varepsilon, \varepsilon) \to \mathbb{R}^N$ is smooth diffeomorphical motion of $\mathbb{R}^N$, and $A_t := \Phi(A, t)$, then v_{A_t} is (locally) Lipschitz, so it is differentiable for almost all t, and the derivative is in TN_c. This includes all perspective, affine, and Euclidean deformations. Unfortunately the contingent cone is not capable of expressing some shape deformations.

Example 6.19 *We consider the* ***removing*** *motion; let A be compact, and suppose that x is in the internal part of A; let $A_t := A \setminus B(x,t)$ be the removal of a small ball from A. The motion v_{A_t} inside N_c is Lipschitz, but the limit*

$$\lim_{t \to 0+} \frac{v_{A_t} - v_A}{\|v_{A_t} - v_A\|_{L^p}}$$

does not exist in L^p. (Morally, if $p = 1$, the limit would be the measure δ_x).

6.3.3 Riemannian metric

Let now $p = 2$. The set N_c may fail to be a smooth submanifold of L^2; yet we will, as much as possible, pretend that it is, in order to induce a sort of "Riemannian metric" on N_c from the standard L^2 metric.

Definition 6.20 *We define the "Riemannian metric" on N_c simply by*

$$\langle h, k \rangle := \langle h, k \rangle_{L^2}$$

[8] "Generically" is meant in the Baire sense: the set of exceptions is of first category.

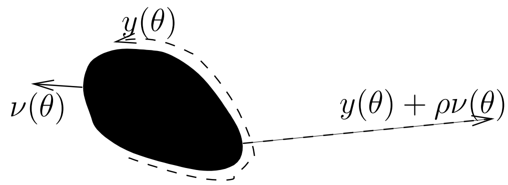


Figure 8: Polar coordinates around a convex set

for $h, k \in T_v N_c$ *and correspondingly a norm by*

$$|h| := \sqrt{\langle h, h \rangle}$$

where $T_v N_c$ *is the* contingent cone*.*

Proposition 6.21 *The distance induced by this "Riemannian metric" coincides with the geodesically induce distance* d^g*.*

The proof is in 3.22 in Duci and Mennucci [2007].

To conclude, we propose an explicit computation of the Riemannian Metric for the case of compact sets in the plane with smooth boundaries; we then pull back the metric to obtain a metric of closed embedded planar curves. We start with the case of convex sets.We fix $p = 2$, $N = 2$.

6.3.4 Polar coordinates of smooth convex sets

Let $\Omega \subset \mathbb{R}^2$ be a convex set with smooth boundary; let $y(\theta) : [0, L] \to \partial\Omega$ be a parameterization of the boundary (by arc parameter), $\nu(\theta)$ the unit vector normal to $\partial\Omega$ and pointing external to Ω. The following ***"polar"* change of coordinates** ψ holds:

$$\psi : \mathbb{R}^+ \times [0, L] \to \mathbb{R}^2 \setminus \Omega \quad , \quad \psi(\rho, \theta) = y(\theta) + \rho\nu(\theta) \tag{6.3}$$

see figure 8. We suppose that $y(\theta)$ moves on $\partial\Omega$ in anticlockwise direction; so $\nu = J\partial_s y$, $\partial_{ss} y = -\kappa\nu$; where J is the rotation matrix (of angle $-\pi/2$), κ is the curvature, and $\partial_s y$ is the tangent vector.

We can then express a generic integral through this change of coordinates as

$$\int_{\mathbb{R}^2 \setminus \Omega} f(x)\, \mathrm{d}x = \int_{\mathbb{R}^+} \int_{\partial\Omega} f(\psi(\rho, s)) |1 + \rho\kappa(s)|\, \mathrm{d}\rho\, \mathrm{d}s$$

where s is arc parameter, and ds is integration in arc parameter.

6.3.5 Smooth deformations of a convex set

We want to study a smooth deformation of Ω, that we call Ω_t; then the border $y(\theta, t)$ depends on a time parameter t. Suppose also that $\kappa(\theta) > 0$, that is, that the set is strictly convex: then for small smooth deformations, the set Ω_t will still be strictly convex. We suppose that the border of Ω_t moves with orthogonal speed α; more precisely, we assume that $(\partial_t y) \perp \partial_s y$, that is, $(\partial_t y) = \alpha\nu$ with $\alpha = \alpha(t,\theta) \in \mathbb{R}$. Since this deformation is smooth, we expect that it will be associated to a vector $h_\alpha \in T_v N_c$, defined by $h_\alpha := \partial_t v_{\Omega_t}$. We now show briefly how to explicitly compute it.

Suppose that x is a fixed point in the plane, $x \notin \Omega_t$, and express it using polar coordinates $x = \psi(\rho,\theta)$, with $\rho = \rho(t), \theta = \theta(t)$. With some computations, $\rho' = -\alpha$. Now, for $x \notin \Omega_t$, $u_{\Omega_t}(x) = \rho(t)$ hence we obtain the explicit formula for h_α

$$h_\alpha := \partial_t v_{\Omega_t}(x) = -\varphi'(u_{\Omega_t}(x))\alpha \ ;$$

whereas $h_\alpha(x) = 0$ for $x \in \mathring{\Omega}_t$.

6.3.6 Pullback of the metric on convex boundaries

Let us fix two orthogonal smooth vector fields $\alpha(s)\nu(s), \beta(s)\nu(s)$, that represent two possible deformations of $\partial\Omega$; those correspond to two vectors $h_\alpha, h_\beta \in T_v N_c$; so the Riemannian Metric that we defined in 6.20 can be pulled back on $\partial\Omega$, to provide the metric

$$\begin{aligned} \langle \alpha,\beta\rangle &:= \int_{\mathbb{R}^2} h_\alpha(x)h_\beta(x)dx = \int_{\mathbb{R}^2\setminus\Omega} h_\alpha(x)h_\beta(x)dx = \\ &= \int_{\partial\Omega}\left[\int_{\mathbb{R}^+} (\varphi'(\rho))^2(1+\rho\kappa(s))\,\mathrm{d}\rho\right]\alpha(s)\beta(s)ds \end{aligned}$$

that is,

$$\langle \alpha,\beta\rangle = \int_{\partial\Omega} (a + b\kappa(s))\alpha(s)\beta(s)ds \tag{6.4}$$

with

$$a := \int_{\mathbb{R}^+} (\varphi'(\rho))^2\,\mathrm{d}\rho \quad , \quad b := \int_{\mathbb{R}^+} (\varphi'(\rho))^2 \rho\,\mathrm{d}\rho \ .$$

6.3.7 Pullback of the metric on smooth contours

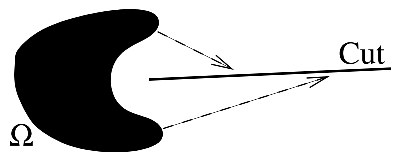
If Ω is smooth but not convex, then the above formula holds up to the cutlocus.

Definition 6.22 *The* ***cutlocus*** *(* a.k.a. *the* ***external skeleton****) is the set of points $x \notin \Omega$ such that there are two (or more) different points $y_1, y_2 \in \Omega$ of minimum distance from x to Ω, that is,*

$$|x - y_1| = |x - y_2| = u_\Omega(x) \ .$$

We define a function $R(s) : [0, L] \to [0,\infty]$ that measures the distance from $\partial\Omega$ to the cutlocus Cut; in turn, we can parameterize Cut as

$$\mathrm{Cut} = \{\psi(R(s), s) \mid s \in [0, L], R(s) < \infty\} \ ;$$



(The arrows represent the distance $R(s)$ from $y(s) \in \partial\Omega$ to the cutlocus).

Figure 9: Smooth contour and cutlocus.

$R(s)$ is locally Lipschitz where it is finite (by results in Itoh and Tanaka [2000], Li and Nirenberg [2005]). The *"polar"* change of coordinates ψ (defined in (6.3)) is a diffeomorphism between the sets

$$\{(\rho, s)\ s \in [0, L], 0 < \rho < R(s)\} \leftrightarrow \mathbb{R}^2 \setminus (\Omega \cup \mathrm{Cut}) \ .$$

See fig. 9.

The pullback metric for deformations of the contour $\partial\Omega$ is

$$\langle h, k\rangle = \int_{\partial\Omega} \left[\int_0^{R(s)} (\varphi'(\rho))^2 (1 + \rho\kappa(s))\, \mathrm{d}\rho \right] \alpha(s)\beta(s)\, \mathrm{d}s$$

6.3.8 Conclusion

In this case we have found (a posteriori) a Riemannian metric of closed embedded planar curves, and we know the structure of the completion, and the completion admits minimal geodesics. On the down side, the completion is not really "a smooth Riemannian manifold". For example, it is difficult to study the minimal geodesics, and to prove any property about them. Anyway, the fact that N_c is locally compact and the regularity property 6.18 of L^p spaces suggest that it may be possible to "shoot geodesics" (in some weak form). Unfortunately the topology has too few open sets to be used for shape optimization: many simple energy functionals are not continuous.

7 Finsler geometries of curves

We present two examples of Finsler geometries of curves that have been used (sometimes covertly) in the literature.

7.1 Tangent and normal

Let $v, T \in \mathbb{R}^N$, let

$$N = T^\perp = \{w \in \mathbb{R}^N : w \cdot T = 0\}$$

be the space orthogonal to T. Usually T will be the tangent to a curve c.[9]

[9]There is a slight abuse of notation here, since in the definition $N = T^\perp$ given for planar curves in 1.12, we defined N to be a "vector" and not the "vector space orthogonal to T".

Definition 7.1 *We define the projection onto the normal space $N = T^\perp$*

$$\pi_N : \mathbb{R}^N \to \mathbb{R}^N \quad , \quad \pi_N v = v - \langle v, T\rangle T$$

and the projection on the tangent T

$$\pi_T : \mathbb{R}^N \to \mathbb{R}^N \quad , \quad \pi_T v = \langle v, T\rangle T$$

so $\pi_N v + \pi_T v = v$ and $|\pi_N v|^2 + |\pi_T v|^2 = |v|^2$.

7.2 L^∞ metric and Fréchet distance

If we wish to define a norm on $T_c M$ that is modeled on the norm of the Banach space $L^\infty(S^1 \to \mathbb{R}^N)$, we define

$$F^\infty(c, h) := \|\pi_N h\|_{L^\infty} = \sup_\theta |\pi_N h(\theta)|$$

where $N = (D_s c)^\perp$. This metric weights the maximum normal motion of the curve. *(The rôle of π_N will be properly explained in §11.10).* This Finsler metric is geometric. The length of a homotopy is

$$\mathrm{Len}_\infty(C) := \int_0^1 \sup_{\theta \in S^1} |\pi_N \partial_t C(t, \theta)| \,\mathrm{d}t \ .$$

Definition 7.2 (Fréchet distance)

$$d_f(c_0, c_1) := \inf_\phi \sup_u |c_1(\phi(u)) - c_0(u)|$$

where $u \in S^1$ and ϕ is chosen in the class of diffeomorphisms of S^1.

This distance is induced by the Finsler metric $F^\infty(c, h)$.

Theorem 7.3 *$d_f(c_0, c_1)$ coincides with the infimum of the length* $\mathrm{Len}_\infty(C)$ *for C connecting c_0 to (a reparameterization of) c_1.*

For the proof, see theorem 15 in Mennucci et al. [2008].

7.3 L^1 metric and Plateau problem

If we wish to define a geometric norm on $T_c M$ that is modeled on the norm of the Banach space $L^1(S^1 \to \mathbb{R}^N)$, we may define the metric

$$F^1(c, h) = \|\pi_N h\|_{L^1} = \int |\pi_N h(\theta)||c'(\theta)| d\theta \ ;$$

the length of a homotopy is then

$$\mathrm{Len}(C) = \iint |\pi_N \partial_t C(t, \theta)||C'(t, \theta)| \,\mathrm{d}t \,\mathrm{d}\theta$$

Figure 10: Examples of curves of different rotation index.

which coincides with

$$\mathrm{Len}(C) = \iint |\partial_t C(t,\theta) \times \partial_\theta C(t,\theta)| \, \mathrm{d}\theta \, \mathrm{d}t$$

This last is easily recognizable as the surface area of the homotopy (up to multiplicity); the problem of finding a minimal geodesic connecting c_0 and c_1 in the F^1 metric may be reconducted to the Plateau problem of finding a surface which is an immersion of $I = S^1 \times [0,1]$ and which has fixed borders to the curves c_0 and c_1. The Plateau problem is a wide and well studied subject upon which Fomenko expounds in the monograph Fomenko [1990].

8 Riemannian metrics of curves up to pose

A particular approach to the study of shapes is to define a shape to be a *curve up to reparameterization, rotation, translation, scaling*; to abbreviate, we call this the *shape space of curves up to pose.* We present two examples of Riemannian metrics.

8.1 Representation by angle function

We consider in the following planar curves $\xi : S^1 \to \mathbb{R}^2$ of length 2π and parameterized by arc length.

Proposition 8.1 (Continuous lifting, rotation index) *If $\xi \in C^1$ and parameterized by arc parameter, then ξ' is continuous and $|\xi'| = 1$, so there exists a continuous function $\alpha : \mathbb{R} \to \mathbb{R}$ satisfying*

$$\xi'(s) = \big(\cos(\alpha(s)),\ \sin(\alpha(s))\big) \tag{8.1}$$

and $\alpha(s)$ is unique, up to adding the constant $k2\pi$ with $k \in \mathbb{Z}$.

Moreover $\alpha(s+2\pi) - \alpha(s) = 2\pi I$, where I is an integer, known as ***rotation index*** *of ξ. This number is unaltered if ξ is homotopically deformed in a smooth way.*

See the examples in Fig. 10.

The addition of a real constant to $\alpha(s)$ is equivalent to a rotation of ξ. We then understand that we may represent arc parameterized curves $\xi(s)$, up to translation, scaling, and rotations, by considering a suitable class of liftings $\alpha(s)$ for $s \in [0, 2\pi]$.

The use of the angle function for representing (possibly non-closed) planar curves dates back to Zahn and Roskies [1972].

8.2 Shape Representation using direction functions

Klassen, Mio, Srivastava et al in Klassen et al. [2004]; Mio and Srivastava [2004b] represent a close planar curve c by a pair of velocity-angle functions (ϕ, α) through the identity

$$c'(u) = \exp(\phi(u) + i\alpha(u))$$

(identifying $\mathbb{R}^2 = \mathbb{C}$), and then defining a metric on the velocity-angle function space. They propose models of spaces of curves where the metrics involve higher order derivatives in Klassen et al. [2004].

We review here the simplest such model, where $\phi = 0$.

In Klassen et al. [2004] two spaces of closed planar curves are defined; we present the case of *"Shape Representation using Direction Functions"*;

Definition 8.2 *let* $L^2 := L^2([0, 2\pi] \to \mathbb{R})$*;* $\Phi : L^2 \to \mathbb{R}^3$ *is defined by*

$$\Phi_1(\alpha) := \int_0^{2\pi} \alpha(s)\,\mathrm{d}s\ ,\ \Phi_2(\alpha) := \int_0^{2\pi} \cos\alpha(s)\,\mathrm{d}s\ ,\ \Phi_3(\alpha) := \int_0^{2\pi} \sin\alpha(s)\,\mathrm{d}s\ .$$

and the space of (pre)-shapes is defined as the closed subset S *of* L^2*,*

$$S := \left\{\alpha \in L^2 \mid \Phi(\alpha) = (2\pi^2, 0, 0)\right\}\ ;$$

the condition $\Phi_1(\alpha) = 2\pi^2$ forces a choice of rotation, while $\Phi_2 = 0, \Phi_3 = 0$ ensure that α represents a closed curve.

For any $\alpha \in S$, it is possible to reconstruct the curve by integrating

$$\xi(s) = \int_0^s \big(\cos(\alpha(t)), \sin(\alpha(t))\big)dt \tag{8.2}$$

This means that α identifies an unique curve (of length 2π, and arc parameterized) up to rotations and translations, and to the choice of the base point $\xi(0)$; for this last reason, S is called in Klassen et al. [2004] a **preshape space**.

Inside the family of arc parameterized curves, the reparameterization[10] of ξ is restricted to the transformation $\hat{\xi}(u) = \xi(u - s_0)$, that is a different choice of base point for the parameterization along the curve. Inside the *preshape space* S, the same operation is encoded by the relation $\hat{\alpha}(s) = \alpha(s - s_0)$.

To obtain a model of *immersed curves up to reparameterization, rotation, translation, scaling* we need to quotient S by the relation $\alpha \sim \hat{\alpha}$, for all possible $s_0 \in \mathbb{R}$. We do not discuss this quotient here.

We now prove that $S \setminus Z$ is a smooth manifold. We first define Z.

Definition 8.3 (Flat curves) *Let* Z *be the set of all* $\alpha \in L^2([0, 2\pi])$ *such that* $\alpha(s) = a + k(s)\pi$ *where* $k(s) \in \mathbb{Z}$ *and* k *is measurable,* $a = 2\pi - \int k \in \mathbb{R}$*, and*

$$|\{k(s) = 0 \bmod 2\}| = |\{k(s) = 1 \bmod 2\}| = \pi$$

[10] We are considering only reparameterizations in $\mathrm{Diff}^+(S^1)$.

- Z is closed (by thm. 4.9 in Brezis [1986]).
- $S \setminus Z$ contains the (representation α by continuous lifting of) all smooth immersed curves.
- Z contains the (representations α of) flat curves ξ, that is, curves ξ whose image is contained in a line.

Example 8.4 *An example of a flat curve is*

$$\xi_1(s) = \xi_2(s) = \begin{cases} s/\sqrt{2} & s \in [0, \pi] \\ (2\pi - s)/\sqrt{2} & s \in (\pi, 2\pi] \end{cases}, \qquad \alpha = \begin{cases} \pi/2 & s \in [0, \pi] \\ 3\pi/2 & s \in (\pi, 2\pi] \end{cases}$$

Proposition 8.5 *$S \setminus Z$ is a smooth immersed submanifold of codimension 3 in L^2.*

Proof. By the implicit function theorem. Indeed, suppose by contradiction that $\nabla\Phi_1, \nabla\Phi_2, \nabla\Phi_3$ are linearly dependant at $\alpha \in S$, that is, there exists $a \in \mathbb{R}^3, a \neq 0$ s.t.

$$a_1 \cos(\alpha(s)) + a_2 \sin(\alpha(s)) + a_3 = 0$$

for almost all s; then, by integrating, $a_3 = 0$, therefore $a_1 \cos(\alpha(s)) + a_2 \sin(\alpha(s)) = 0$ that means that $\alpha \in Z$. □

The manifold $S \setminus Z$ inherits a Riemannian structure, induced by the scalar product of L^2; (critical) geodesics may be prolonged smoothly as long as they do not meet Z.

Even if S may not be a manifold at Z, we may define the **geodesic distance** $d^g(x, y)$ in S as the infimum of the length of Lipschitz paths $\gamma : [0, 1] \to L^2$ going from x to y and whose image is contained in S;[11] since $d^g(x, y) \geq \|x - y\|_{L^2}$, and S is closed in L^2, then the metric space (S, d^g) is metrically complete.

We don't know if (S, d^g) admits minimal geodesics, or if it falls in the same category as the Atkin example 3.34.

8.2.1 Multiple measurable representations

We may represent any Lipschitz closed arc parameterized curve ξ using a measurable $\alpha \in S$.

Definition 8.6 *A* ***measurable lifting*** *is a measurable function $\alpha : \mathbb{R} \to \mathbb{R}$ satisfying (8.1). Consequently, a* ***measurable representation*** *is a measurable lifting α satisfying conditions $\Phi(\alpha) = (2\pi^2, 0, 0)$ (from Definition 8.2).*

We remark that

[11] It seems that S is Lipschitz-arc-connected, so $d^g(x, y) < \infty$; but we did not carry on a detailed proof

- a measurable representation always exists; for example, let

$$\mathrm{arc} : S^1 \to [0, 2\pi)$$

be the inverse of

$$\alpha \mapsto (\cos(\alpha), \sin(\alpha))$$

when $\alpha \in [0, 2\pi)$. arc() is a Borel function; then $((\mathrm{arc} \circ \xi')(s) + a) \in S$, for a choice of $a \in \mathbb{R}$.

- The measurable representation is never unique: for example, given any measurable $A, B \subset [0, 2\pi]$ with $|A| = |B|$,

$$\alpha(s) + 2\pi \mathbf{1}_A(s) - 2\pi \mathbf{1}_B(s)$$

will as well represent ξ.

This implies that the family A_ξ of measurable $\alpha \in S$ that represent the same curve ξ is infinite. It may be then advisable to define a **quotient distance** $\hat{d}$ as follows:

$$\hat{d}(\xi_1, \xi_2) := \inf_{\alpha_1 \in A_{\xi_1}, \alpha_2 \in A_{\xi_2}} d(\alpha_1, \alpha_2) \tag{8.3}$$

where $d(\alpha_1, \alpha_2) = \|\alpha_1 - \alpha_2\|_{L^2}$, or alternatively $d = d^g$ is the geodesic distance on S.

8.2.2 Continuous vs measurable lifting — no rotation index

If $\xi \in C^1$, we have an unique[12] continuous representation $\alpha \in S$; but note that, even if $\xi_1, \xi_2 \in C^1$, the infimum (8.3) may not be given by the continuous representations α_1, α_2 of ξ_1, ξ_2. Moreover there are rectifiable curves ξ that do not admit a continuous representation α, as for example the polygons.

A problem similar to the above is expressed by this proposition.

Proposition 8.7 *For any $h \in \mathbb{Z}$, the set of closed smooth curves ξ with* rotation index *h, when represented in S using the continuous lifting, is dense in $S \setminus Z$.*

It implies that we cannot properly extend the concept of *rotation index* to S.

The proof is based on this lemma.

Lemma 8.8 *Suppose that ξ is not flat, let τ be one of the measurable liftings of ξ. There exists a smooth projection $\pi : V \to S$ defined in a neighborhood $V \subset L^2$ of τ such that, if $f \in L^2 \cap C^\infty$ then $\pi(f)$ is in $L^2 \cap C^\infty$.*

This is the proof of both the lemma and the proposition.

[12]Indeed, the continuous lifting is unique up to addition of a constant to $\alpha(s)$, which is equivalent to a rotation of ξ; and the constant is decided by $\Phi_1(\alpha) = 2\pi^2$

Proof. Fix $\alpha_0 \in S \setminus Z$. Let $T = T_{\alpha_0} S$ be the tangent at α_0. T is the vector space orthogonal to $\nabla\phi_i(\alpha_0)$ for $i = 1, 2, 3$. Let $e_i = e_i(s) \in L^2 \cap C_c^\infty$ be near $\nabla\phi_i(\alpha_0)$ in L^2, so that the map $(x, y) : T \times \mathbb{R}^3 \to L^2$

$$(x, y) \mapsto \alpha = \alpha_0 + x + \sum_{i=1}^{3} e_i y_i \tag{8.4}$$

is an isomorphism. Let S' be S in these coordinates; by the Implicit Function Theorem (5.9 in Lang [1999]), there exists an open set $U' \subset T$, $0 \in U'$, an open $V' \subset \mathbb{R}^3$, $0 \in V'$, and a smooth function $F : U \to \mathbb{R}^3$ such that the local part $S' \cap (U' \times V')$ of the manifold S' is the graph of $y = F(x)$.

We immediately define a smooth projection $\pi : U' \times V' \to S'$ by setting $\pi'(x, y) = (x, F(x))$; this may be expressed in the original L^2 space; let $(x(\alpha), y(\alpha))$ be the inverse of (8.4) and $U = x^{-1}(U')$; we define the projection $\pi : U \to S$ by setting

$$\pi(\alpha) = \alpha_0 + x + \sum_{i=1}^{3} e_i F_i(x(\alpha))$$

Then

$$\pi(\alpha)(s) - \alpha(s) = \sum_{i=1}^{3} e_i(s) a_i \quad , \quad a_i := (F_i(x(\alpha)) - y_i(\alpha)) \in \mathbb{R} \tag{8.5}$$

so if $\alpha(s)$ is smooth, then $\pi(\alpha)(s)$ is smooth.

Let α_n be smooth functions such that $\alpha_n \to \alpha$ in L^2, then $\pi(\alpha_n) \to \alpha_0$; if we choose them to satisfy $\alpha_n(2\pi) - \alpha_n(0) = 2\pi h$, then, by the formula (8.5), $\pi(\alpha)(2\pi) - \pi(\alpha)(0) = 2\pi h$ so that $\pi(\alpha_n) \in S$ and it represents a smooth curve with the assigned rotation index h. □ □

8.3 A metric with explicit geodesics

A similar method has been proposed recently in Younes et al. [2008], based on an idea originally in Younes [1998]. We consider immersed planar curves and again we identify $\mathbb{R}^2 = \mathbb{C}$.

Proposition 8.9 (Continuous lifting of square root) *If $\xi : [0, 2\pi] \to \mathbb{C}$ is an immersed planar curve, then ξ' is continuous and $\xi' \neq 0$, so there exists a continuous function $\alpha :\to \mathbb{C}$ satisfying*

$$\xi'(\theta) = \alpha(\theta)^2 \tag{8.6}$$

and α is uniquely identified up to multiplying by ± 1.

If the rotation index of ξ is even then $\alpha(0) = \alpha(2\pi)$, whereas if the rotation index of ξ is odd then $\alpha(0) = -\alpha(2\pi)$.

We will use this lifting to obtain, as a first step, a representation of *curves up to rotation, translation, scaling.* To this end, let ξ be an immersed planar closed curve of $\mathrm{len}(\xi) = 2$ (not necessarily parameterized by arc parameter); let α be the **square root lifting** of the derivative ξ'. Let e, f be the real and imaginary part of α, that is, $\alpha = e + if$. The condition that ξ is a closed curve translates into $\int_0^{2\pi}(e + if)^2 \,\mathrm{d}\theta = 0$ (where equality is in $\mathbb{C}$), hence we have two equalities (for real and imaginary part)

$$\int_0^{2\pi} e^2 - f^2 \,\mathrm{d}\theta = 0, \quad \int_0^{2\pi} ef \,\mathrm{d}\theta = 0$$

while the condition that $\mathrm{len}(\xi) = 2$ translates into

$$\int_0^{2\pi} |\xi'| \,\mathrm{d}\theta = \int_0^{2\pi} |e + if|^2 \,\mathrm{d}\theta = \int_0^{2\pi} (e^2 + f^2) \,\mathrm{d}\theta = 2 \ .$$

With some algebra we obtain that the above conditions are equivalent to

$$\int_0^{2\pi} e^2 \,\mathrm{d}\theta = 1, \int_0^{2\pi} f^2 \,\mathrm{d}\theta = 1, \int_0^{2\pi} ef \,\mathrm{d}\theta = 0.$$

Let $L^2 = L^2([0, 2\pi] \to \mathbb{R})$; let $S \subset L^2 \times L^2$ be defined by

$$S := \left\{ (e, f) \mid \int_0^{2\pi} e^2 \,\mathrm{d}\theta = 1 = \int_0^{2\pi} f^2 \,\mathrm{d}\theta, \int_0^{2\pi} ef \,\mathrm{d}\theta = 0. \right\}$$

S is the **Stiefel manifold** of orthonormal pairs in L^2. S is a smooth manifold, and inherits the (flat) metric of $L^2 \times L^2$. What is most surprising is that

Theorem 8.10 (2.2 in Younes et al. [2008]) *Let $\delta\xi$ be a small deformation of ξ, and $\delta e, \delta f$ be the corresponding small deformation of the representation e, f. Then*

$$\int_\xi |D_s \delta\xi|^2 \,\mathrm{d}s = \int_0^{2\pi} (\delta e)^2 + (\delta f)^2 \,\mathrm{d}\theta$$

that is, the (geometric) Riemannian metric in M is mapped into the (flat and parametric) metric in S.

It is then natural to "embed" closed planar curves (up to translation and scaling) into S.

8.3.1 Curves up to rotation represented as a Grassmanian

We note that rotation of ξ by an angle τ is equivalent to rotation of the frame (e, f) by an angle $\tau/2$. So the orbit of all rotations of ξ is associated to the plane generated by (e, f) in L^2: the space of *curves up to rotation/translation/scaling* is represented by the **Grassmanian manifold** of 2-planes in L^2. A theorem by Neretin then applies, that provides a closed form formula for critical and minimal geodesics. See section 4 in Younes et al. [2008].

8.3.2 Representation

The *Stiefel manifold* is a complete smooth Riemannian manifold; it contains the (representation of) all closed rectifiable parametric planar curves, up to scaling and translation. So with this choice of metric and representation, we obtain that the completion of the Fréchet manifold of smooth curves is a Riemannian smooth manifold. There are two problems left.

- The quotient w.r.to reparameterization. This is studied in Younes et al. [2008], where it is proven that unfortunately geodesics in the quotient space may develop singularities in the reparameterizations at the end times of the geodesic.
- But there is also the quotient w.r.to representation.

8.3.3 Quotient w.r.to representation

As in the previous section, we may define a *measurable square root lifting*; this lifting will not be unique. Indeed, let (e, f) be the square root representation of ξ, that is $\xi' = (e + if)^2$; choose any function $a : S^1 \to \{-1, 1\}$ arbitrarily (but measurable); then (ae, af) represents the same curve. So again we may define a quotient metric $\hat{d}(\xi_1, \xi_2)$ as we did in eqn. (8.3); similar comments to those at the end of Section 8.2.2 hold.

9 Riemannian metrics of immersed curves

Metrics of "geometric" curves have been studied by Michor and Mumford [2006, 2007, 2005] and Yezzi and Mennucci [2005, 2004a,b]; more recently, Yezzi-M.-Sundaramoorthi Sundaramoorthi et al. [2005, 2006, 2007a,b, 2008a,b]; Mennucci et al. [2008] have studied Sobolev-like metrics of curves and shown many good properties for applications to shape optimization; similar results have also been shown independently by Charpiat et al. [2004, 2005, 2007].

We now discuss some Riemannian metrics on immersed curves.

- The H^0 metric
$$\langle h_1, h_2 \rangle_{H^0} = \int_c \langle h_1(s), h_2(s) \rangle \, \mathrm{d}s \ .$$
- Michor and Mumford [2006]'s metric
$$\langle h_1, h_2 \rangle_{H^A} = \int_c (1 + A|\kappa_c|^2) \langle h_1, h_2 \rangle \, \mathrm{d}s \ .$$
- Yezzi and Mennucci [2004b] *conformal metric*
$$\langle h_1, h_2 \rangle_{H^0_\phi} = \phi(c) \int_c \langle h_1, h_2 \rangle \, \mathrm{d}s \ .$$

- Charpiat et al. [2005] *rigidified norms*
- Sundaramoorthi et al. [2005] Sobolev type metrics

$$\langle h_1, h_2\rangle_{H^n} = \fint_c \langle h_1, h_2\rangle \,\mathrm{d}s + \mathrm{len}(c)^{2n} \fint_c \langle \partial_s^n h_1, \partial_s^n h_2\rangle \,\mathrm{d}s$$

$$\langle h_1, h_2\rangle_{\tilde H^n} = \left\langle \fint_c h_1 \,\mathrm{d}s, \fint_c h_2 \,\mathrm{d}s \right\rangle + \mathrm{len}(c)^{2n} \fint_c \langle \partial_s^n h_1, \partial_s^n h_2\rangle \,\mathrm{d}s \ .$$

We will now present a quick overview of all metrics (but for the latter, that is discussed in the next section).

9.1 H^0

The H^0 inner product was defined in eqn. (2.4) as

$$\left\langle h_1, h_2 \right\rangle_{H^0} := \int_c h_1(s) \cdot h_2(s) \,\mathrm{d}s \ ; \tag{9.1}$$

it is possibly the simplest geometric inner product that we may imagine to apply to curves. We already noted that the minimizing flows implemented in traditional active contour methods are "gradient flows" only if we use this H^0 inner product.

We will show in Sec. 11.3 that the H^0-induced distance is identically zero. In Yezzi and Mennucci [2004b] there is a result that shows that the distance is non degenerate, and minimal geodesics exists, when the shape space is restricted to curves with uniformly bounded curvature.

9.2 H^A

Michor and Mumford [2006] propose the metric H^A

$$\langle h_1, h_2\rangle_{H^A_{|c}} = \int_0^L (1 + A|\kappa_c|^2)\langle h_1, h_2\rangle \,\mathrm{d}s$$

where κ_c is the curvature of c, and $A > 0$ is fixed.

Properties 9.1

- *The induced distance is non degenerate.*
- *The completion $\overline{M}$ (intended in the metric sense) is*

  $$BV^2 \subset \overline{M} \subset Lip$$

  where BV^2 are the curves that admit curvature as a measure and Lip are the rectifiable curves.
- *There are compactness bounds.*

9.3 Conformal metrics

Yezzi and Mennucci [2004b] proposed to change the metric, from H^0 to a **conformal metric**

$$\langle h_1, h_2 \rangle_{H^0_\psi} = \psi(c) \int_c \langle h_1, h_2 \rangle \, \mathrm{d}s$$

where $\psi(c)$ associates to each curve c a positive number. Then the *gradient descent flow* of an energy E defined on curves $C(t, \cdot)$ is

$$\frac{\partial C}{\partial t} = -\nabla^\psi E(C) = -\frac{1}{\psi(C)} \nabla E(C)$$

where $\nabla E(C)$ is the gradient for the H^0 metric. This is equivalent to a change of time variable t in the gradient descent flow. So all properties of the flows are unaffected if we switch from a H^0 to a conformal-H^0 metric.

Properties 9.2 *Consider a conformal metric where $\psi(c)$ depends (monotonically) on the length* $\mathrm{len}(c)$ *of the curve.*

- *If $\psi(c) \geq \mathrm{len}(c)$, the induced metric is non degenerate;*
- *unfortunately, according to a result by Shah [2008], when $\psi(c) \equiv L(c)$ very few (minimal) geodesics do exist (only "grassfire" geodesics, moving by constant normal speed).*

9.4 "Rigidified" norms

Charpiat et al.Charpiat et al. [2005] consider norms that favor pre-specified rigid motions. They decompose a motion h using the H^0 projection as

$$h = h_{\text{rigid}} + h_{\text{rest}}$$

where h_{rigid} contains the rigid part of the motion; then they choose λ large, and construct the norm

$$\|h\|^2_{\text{rigid}} = \lambda \|h_{\text{rigid}}\|^2_{H^0} + \|h_{\text{rest}}\|^2_{H^0} .$$

Note that these norms are equivalent to the H^0-type norm; as a result the induced distance is (again) degenerate.

10 Sobolev type Riemannian metrics

In this part we discuss the Sobolev norms, with applications and experiments. What follows summarizes Sundaramoorthi et al. [2007a, 2008a,b]; Mennucci et al. [2008].

10.1 Sobolev-type metrics

Recently in Sundaramoorthi et al. [2005, 2006, 2007a,b, 2008a,b]; Mennucci et al. [2008] Yezzi–M.–Sundaramoorthi studied a family of Sobolev-type metrics. Let $D_s := \frac{1}{|c'|}\partial_\theta$ be the derivative with respect to arc parameter. Let $j \geq 1$ be an integer[13] and

$$\langle h_1, h_2 \rangle_{H_0^j} := \fint_c \langle D_s^j h_1, D_s^j h_2 \rangle \, \mathrm{d}s \tag{10.1}$$

where $\fint_c \cdots \mathrm{d}s$ was defined in 1.9. Let $\lambda > 0$ a fixed constant; we define the **Sobolev-type metrics**

$$\langle h_1, h_2 \rangle_{H^j} \quad := \quad \fint_c \langle h_1, h_2 \rangle \, \mathrm{d}s + \lambda L^{2j} \, \langle h_1, h_2 \rangle_{H_0^j} \tag{10.2}$$

$$\langle h_1, h_2 \rangle_{\tilde{H}^j} \quad := \quad \Big\langle \fint_c h_1 \, \mathrm{d}s, \fint_c h_2 \, \mathrm{d}s \Big\rangle + \lambda L^{2j} \, \langle h_1, h_2 \rangle_{H_0^j} \tag{10.3}$$

where $L = \mathrm{len}(c)$. Notice that these metrics are *geometric*:

- they are easily seen to be invariant w.r.to rotations and translations, (in a stronger sense than in page 39, indeed in this case

  $$\langle h_1, h_2 \rangle_{Ac} = \langle h_1, h_2 \rangle_c \quad , \quad \langle Rh_1, Rh_2 \rangle_c = \langle h_1, h_2 \rangle_c$$

  for any Euclidean transformation A and rotation R);
- they are *reparameterization invariant* due to the usage of the arc parameter in derivatives and integrals;
- they are *scale invariant*, since the normalizing factors L^{2j} make them 0-homogeneous w.r.to rescaling of the curve.

For this last reason, we **redefine the H^0 metric** to be

$$\Big\langle h_1, h_2 \Big\rangle_{H^0} := \fint_c h_1 \cdot h_2 \, \mathrm{d}s \tag{10.4}$$

so that it is again 0-homogeneous.This is a conformal version of the H^0 metric defined in eqn. (2.4), so a gradient descent flow is a time reparameterization of the flow for the original H^0 metric; and the induced distance is again degenerate.

When we will present a shape optimization energy E and we will minimize E using a gradient descent flow driven by the H^j or $\tilde{H}^j$ gradient of E, we will call the resulting algorithm a **Sobolev active contour** method (abbreviated as **SAC** in the following).

[13] It is though possible to define Sobolev metrics for any $j \in \mathbb{R}, j > 0$; see Prop. 3.1 in Sundaramoorthi et al. [2008a].

10.1.1 Related works

A family of metrics similar to H^j above (but for the length dependent scale factors[14]) was studied (independently) in Michor and Mumford [2007]: the Sobolev-type weak Riemannian metric on $\mathrm{Imm}(S^1, \mathbb{R}^2)$

$$\langle h, k \rangle_{G_c^j} = \int_c \sum_{i=0}^{j} \langle D_s^i h, D_s^i k \rangle \, \mathrm{d}s \quad ;$$

in that paper the geodesic equation, horizontality, conserved momenta, lower and upper bounds on the induced distance, and scalar curvatures are computed. Note that this metric is locally equivalent to the above metrics defined in equation (10.2), (10.3).

Charpiat *et al* in Charpiat et al. [2005, 2007] studied (again independently) some generalized metrics and relative gradient flows; in particular they defined the Sobolev-type metric

$$\fint_c \langle h_1, h_2 \rangle + \langle D_s h_1, D_s h_2 \rangle \, \mathrm{d}s \ .$$

10.1.2 Properties of H^j metrics

This is a list of important properties that will be discussed in the following sections.

- Flow regularization: Sobolev gradient flows are smoother than H^0 flows.
- PDE order reducing property: Sobolev gradient flows are lower order than H^0 flows.
- SAC does not require derivatives of the curve to be defined for many commonly used region-based and edge-based energies.
- Coarse-to-fine deformation property: a SAC automatically favors coarse-scale deformations before moving to fine-scale deformations; this is ideal for visual tracking.
- Sobolev-type norms induce a well-defined distance on space of curves;
- moreover the structure of the completion of the space of immersed curves w.r.to H^1 and H^2 norm is fairly well understood. So they offer a *consistent* theory of *shape optimization* and *shape analysis*.

[14] Though, a scale-invariant Sobolev metric is proposed in Michor and Mumford [2007] in §4.8 as a sensible generalization.

10.2 Mathematical properties

We start summarizing the main mathematical properties, that were presented in Mennucci et al. [2008] mostly. We first of all cite this lemma.

Lemma 10.1 (Poincaré inequality) *Pick $h : [0, l] \to \mathbb{R}^n$, weakly differentiable, with $h(0) = h(l)$ (so h is periodically extensible); let $\hat{h} = \frac{1}{l}\int_0^l h(x)\,\mathrm{d}x$; then*

$$\sup_u |h(u) - \hat{h}| \le \frac{1}{2}\int_0^l |h'(x)|\,\mathrm{d}x \ . \tag{10.5}$$

This is proved as Lemma 18 in Mennucci et al. [2008]; it is one of the main ingredients for the following propositions, whose full proofs are in Mennucci et al. [2008].

Proposition 10.2 *The H^j and $\tilde{H}^j$ distances are equivalent:*

$$d_{\tilde{H}^j} \le d_{H^j} \le \sqrt{\frac{1 + (2\pi)^{2j}\lambda}{(2\pi)^{2j}\lambda}}\, d_{\tilde{H}^j}$$

whereas $d_{H^j} \le d_{H^k}$ for $j < k$.

Proposition 10.3 *The H^j and $\tilde{H}^j$ distances are lower bounded (with appropriate constants depending on λ) by the* Fréchet distance *(defined in 7.2).*

Proposition 10.4 *$c \mapsto len(c)$ is Lipschitz in M with H^j metric, that is,*

$$|\operatorname{len}(c_0) - \operatorname{len}(c_1)| \le d_{H^j}(c_0, c_1)$$

Theorem 10.5 (Completion of B_i w.r.to H^1) *let d_{H^1} be the distance induced by H^1; the metric completion of the space of curves is equal to the space of all rectifiable curves.*

Theorem 10.6 (Completion of B_i w.r.to H^2) *Let $E(c) := \int |D_s^2 c|^2\,\mathrm{d}s$ be defined on non-constant smooth curves; then E is locally Lipschitz in w.r.to d_{H^2}. Moreover the completion of $C^\infty(S^1)$ according to the metric H^2 is the space of curves that admit curvature $D_s^2 c \in L^2(S^1)$.*

The hopeful consequence of the above theorem would be a possible solution to the problem 4.12; it implies that the space of geometric curves B with the H^2 Riemannian metric completes onto the usual Hilbert space H^2. [15] This result in turn would ease a (possible) proof of existence geodesics.

Concluding, it seems that, to have a complete Riemannian manifold of geometric (freely) immersed curves, a metric should penalize derivatives of 2nd order (at least).

[15] The detailed proof has not yet been written...

10.3 Sobolev metrics in shape optimization

We first present a definition.

Definition 10.7 (Convolution) *A* arc-parameterized convolutional kernel K *along the curve c of length L is a L-periodic function $K : \mathbb{R} \to \mathbb{R}$. Given a vector field $f : S^1 \to \mathbb{R}^n$ and a kernel K, we define the* ***convolution by arc parameter***[16] *formally as*

$$(f \star K)(s) := \int_c K(s - \hat{s}) f(\hat{s}) \, \mathrm{d}\hat{s} \ . \tag{10.6}$$

By defining the ***run-length function*** $l : \mathbb{R} \to \mathbb{R}$

$$l(\tau) := \int_0^\tau |c'(x)| \, \mathrm{d}x$$

we can rewrite the above eqn. (10.6) explicitly in θ parameter as

$$(f \star K)(\theta) := \int_0^{2\pi} K\big(l(\theta) - l(\tau)\big) f(\tau) |c'(\tau)| \, \mathrm{d}\tau \ . \tag{10.7}$$

Recall the definition 3.37 of gradient ∇E by means of the identity

$$\langle \nabla E, h \rangle_c = DE(c; h) \quad \forall h \in T_c M \ .$$

Let $f = \nabla_{H^0} E$, $g = \nabla_{H^1} E$ be the gradients w.r.to the inner products H^0 and H^1; by the definition of gradient, we obtain that

$$\langle f, h \rangle_{H^0, c} = \langle g, h \rangle_{H^1, c} \quad \forall h \in T_c M$$

that is[17]

$$⨍_c h \cdot f \, \mathrm{d}s = ⨍_c h \cdot g + \lambda L^2 (D_s h \cdot D_s g) \, \mathrm{d}s \quad \forall h$$

by integrating by parts this becomes

$$\int_c h \cdot (f - g + \lambda L^2 D_s^2 g) \, \mathrm{d}s = 0 \quad \forall h$$

then we conclude that

$$\nabla_{H^0} E = \nabla_{H^1} E - \lambda L^2 D_s^2 (\nabla_{H^1} E) \ . \tag{10.8}$$

With similar computation, for $\tilde{H}^1$ we obtain that

$$\nabla_{H^0} E = ⨍_c \nabla_{\tilde{H}^1} E \, \mathrm{d}s - \lambda L^2 D_s^2 (\nabla_{\tilde{H}^1} E) \ . \tag{10.9}$$

[16] Note this definition is different from the eqn. (13) in Sundaramoorthi et al. [2007a].

[17] We use the definition (10.4) of H^0.

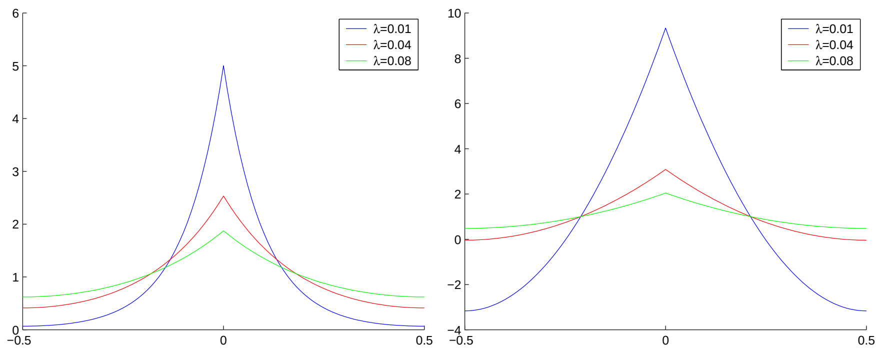


Figure 11: Plots of K_λ (left) and $\tilde{K}_\lambda$ (right) for various λ with $L = 1$. The plots show the kernels over one period.

Given $\nabla_{H^0} E$, both equations can be solved for $\nabla_{H^1} E$ and $\nabla_{\tilde{H}^1} E$, by means of suitable convolution kernels $\tilde{K}_\lambda, K_\lambda$, that is, we have the formulas

$$\nabla_{H^1} E = \nabla_{H^0} E \star K_\lambda \quad , \quad \nabla_{\tilde{H}^1} E = \nabla_{H^0} E \star \tilde{K}_\lambda \ ;$$

The kernels $\tilde{K}_\lambda, K_\lambda$ are known in closed form:

$$K_\lambda(s) = \frac{\cosh\left(\frac{s - \frac{L}{2}}{\sqrt{\lambda} L}\right)}{2L\sqrt{\lambda}\sinh\left(\frac{1}{2\sqrt{\lambda}}\right)}, \quad \text{for } s \in [0, L], \tag{10.10}$$

$$\tilde{K}_\lambda(s) = \frac{1}{L}\left(1 + \frac{(s/L)^2 - (s/L) + 1/6}{2\lambda}\right), \quad s \in [0, L]. \tag{10.11}$$

and $K_\lambda, \tilde{K}_\lambda$ are periodically extended to all of $\mathbb{R}$. See Fig. 11.

The above idea extends to any j: it is possible to obtain $\nabla_{H^j} E$ and $\nabla_{\tilde{H}^j} E$ from $\nabla_{H^0} E$, by convolution. But there is also a way to compute $\nabla_{\tilde{H}^j} E$ without resolving to convolutions, see Prop. 10.10.

The bad news is that, when shape optimization is implemented using a *level set method* (as is usually the case), the curve must be traced out to compute gradients.[18] This in particular leads to some problems when a curve undergoes a topological change, since the Sobolev gradient depends on the *global* shape of the curve; but those problems are easily bypassed when the optimization energy is continuous across topological changes of the curves: see Sec. 5.1 in Sundaramoorthi et al. [2007a].

[18] Though, an alternate method that does not need the tracing of the curve is described in Sec. 5.3 in Sundaramoorthi et al. [2007a] – but it was not successively used, since it depends on some difficult-to-tune parameters.

10.3.1 Smoothing of gradients, coarse-to-fine flowing

In Sundaramoorthi et al. [2008a] it was noted that the regularizing properties may be explained in the Fourier domain: indeed, if we calculate Sobolev gradients $\nabla_{H^j} E$ of an arbitrary energy E in the frequency domain, then

$$\widehat{\nabla_{H^j} E}(l) = \frac{\widehat{\nabla_{H^0} E}(l)}{1 + \lambda(2\pi l)^{2j}} \quad \text{for } l \in \mathbb{Z} \tag{10.12}$$

and

$$\widehat{\nabla_{\tilde{H}^j} E}(0) = \widehat{\nabla_{H^0} E}(0), \quad \widehat{\nabla_{\tilde{H}^j} E}(l) = \frac{\widehat{\nabla_{H^0} E}(l)}{\lambda(2\pi l)^{2j}} \quad \text{for } l \in \mathbb{Z}\backslash\{0\}, \tag{10.13}$$

It is clear from the previous expressions that high frequency components of $\nabla_{H^0} E(c)$ are increasingly less pronounced in the various forms of the H^j gradients. So H^j and $\tilde{H}^j$ gradients are much smoother than H^0 gradients. Note that using a SAC method smooths the gradients, so it induces smoother minimization flows, but it does **not** smooth the curves themselves.

The above phenomenon may be also explained by the following argument. The gradient satisfies the following property.

Proposition 10.8 *If $DE(c) \neq 0$, the* gradient $\nabla E(c)$ *is the vector v in $T_c M$ that provides the maximum in*

$$\frac{DE(c)(v)}{\|v\|_c} = \sup_{h \in T_c M \backslash \{0\}} \frac{|DE(c)(h)|}{\|h\|_c} \quad . \tag{10.14}$$

Thus, the gradient is the most *efficient* perturbation, *i.e.,* it maximizes

$$\frac{\text{change in energy by moving in direction } h}{\text{cost of moving in direction } h}$$

By constructing $\|\cdot\|_c$ to penalize "unfavorable" directions, we have control over the path taken to minimize E *without changing the energy E.* [19]

Since higher frequencies are more penalized in H^1 than in H^0, this explains why a SAC prefers to move the curves in a coarser scale w.r.to a traditional active contour method.

10.3.2 Flow regularization

It is also possible to show mathematically that the Sobolev–type gradients regularize the flows of well known energies, by reducing the degree of the P.D.E., as shown in this example.

[19] In a sense, the focus of research in active contours has been mostly on the numerator in eqn. (10.14) — whereas we now focus on the denominator.

Example 10.9 (Elastica) *In the case of the elastic energy*

$$E(c) = \int_c \kappa^2 \, \mathrm{d}s = \int_c |D_s^2 c|^2 \, \mathrm{d}s \ ,$$

the H^0 *gradient is*[20]

$$\nabla_{H^0} E = L D_s (2 D_s^3 c + 3 |D_s^2 c|^2 D_s c)$$

that includes fourth order derivatives; whereas the $\tilde{H}^1$*-gradient is*

$$-\frac{2}{\lambda L} D_s^2 c + 3L (|D_s^2 c|^2 D_s c) \star (D_s \tilde{K}_\lambda) \tag{10.15}$$

that is an integro-differential second order P.D.E.

A practical positive outcome of this phenomenon will be shown in Section 10.9.1.

10.4 $\tilde{H}^j$ is faster than H^j

We will now show that the metric $\tilde{H}^1$ (that is metrically equivalent to H^1, by Prop. 10.2) leads to a simpler calculus for gradients; with benefits in numerical applications as well.

We recall that $\mathrm{avg}_c(f) := \frac{1}{\mathrm{len}(c)} \int_c f(s) \, \mathrm{d}s$.

Let $E(c)$ be an energy of curves. Gradients are implicitly defined by the following relations

$$DE(c; h) = \langle h, \nabla_{H^0} E \rangle_{H^0} = \langle h, \nabla_{H^1} E \rangle_{H^1} = \langle h, \nabla_{\tilde{H}^1} E \rangle_{\tilde{H}^1} \quad \forall h \ .$$

As we saw in equations (10.8) and (10.9), the above leads to the ODEs

Norm	ODE
H^1	$\nabla_{H^0} E \;=\; \nabla_{H^1} E - \lambda L^2 D_s^2 (\nabla_{H^1} E)$
$\tilde{H}^1$	$\nabla_{H^0} E \;=\; \mathrm{avg}_c(\nabla_{\tilde{H}^1} E) - \lambda L^2 D_s^2 (\nabla_{\tilde{H}^1} E)$

The second ODE is much easier to solve.

Proposition 10.10 *Let* $f = \nabla_{H^0} E$*,* $g = \nabla_{\tilde{H}^1} E$*, then* g *is derived from* f *by the following closed form formulas*

$$\begin{aligned}
g(s) &= g(0) + s g'(0) - \frac{1}{\lambda L^2} \int_0^s (s - \hat{s})(f(\hat{s}) - avg_c(f)) \, \mathrm{d}\hat{s} \\
g'(0) &= -\frac{1}{\lambda L^3} \int_0^L s (f(s) - avg_c(f)) \, \mathrm{d}s \\
g(0) &= \int_0^L f(s) \tilde{K}_\lambda(s) \, \mathrm{d}s \ .
\end{aligned}$$

where $\tilde{K}_\lambda$ *was defined in (10.11).*

[20] We use the definition (10.4) of H^0; the directional derivative was computed in 4.6

Proof. The ODE
$$f = \mathrm{avg}_c(g) - \lambda L^2 D_s^2 g$$
immediately tells that $\mathrm{avg}_c(f) = \mathrm{avg}_c(g)$; so we rewrite it as
$$\lambda L^2 D_s^2 g = -f + \mathrm{avg}_c(f)$$
and we simply integrate twice! Moreover, exploiting the identity
$$\int_0^s \left(\int_0^t h(r)\,\mathrm{d}r \right) \mathrm{d}t = \int_0^s (s-r)\,h(r)\,\mathrm{d}r \ ,$$
and with some extra computations, we obtain the result. □

In the end we obtain that the $\tilde{H}^1$ gradient need not be computed as a convolution; so the $\tilde{H}^1$ gradient enjoys nearly the same computational speed as the H^0 gradient; moreover the resulting gradient flow is more stable, so it works fine in numerical implementations for a larger choice of time step discretization. For these reasons, $\tilde{H}^1$ Sobolev active contours are very fast.

10.5 Analysis and calculus of $\tilde{H}^1$ gradients

We write $h \in T_cM$ as $h = \mathrm{avg}_c(h) + \tilde{h}$; this decomposes
$$T_cM = \mathbb{R}^n \oplus D_cM \tag{10.16}$$
with
$$D_cM := \Big\{ h : S^1 \to \mathbb{R}^n \mid \mathrm{avg}_c(h) = 0 \Big\} \ .$$

If we assign to $\mathbb{R}^n$ its usual Euclidean norm, and to D_cM the H_0^j norm[21], then
$$\|h\|^2_{\tilde{H}^j} = |\mathrm{avg}_c(h)|^2_{\mathbb{R}^n} + \lambda L^{2j} \|\tilde{h}\|^2_{H_0^j} \ .$$

This means that the two spaces $\mathbb{R}^n$ and D_cM are **orthogonal** w.r.to $\tilde{H}^j$.

A remark on this decomposition is due.

Remark 10.12 *In the above, $\mathbb{R}^n$ is <u>akin</u> to be the* ***space of translations*** *and D_cM the* ***space of non-translating deformations****.*

That labeling is not rigorous, though! since the subspace of T_cM of deformations that do not move the center of mass $avg_c(c)$ is not D_cM, but rather
$$\Big\{ h : \int_c h + \big(c - avg_c(c)\big) \langle D_s h \cdot T \rangle \,\mathrm{d}s = 0 \Big\} \ ,$$
as is easily deduced from equation (2.8).

[21] H_0^j was defined in (10.1). Note that H_0^j is a seminorm on T_cM, since its value is zero on constant fields; but H_0^j is a norm on D_cM, by Poincaré inequality (10.5).

The decomposition (10.16) is quite useful in the calculus when analytically computing $\tilde{H}^1$ gradients and proving existence of flows. Indeed, let $f = \nabla_{H^0} E$, $g = \nabla_{\tilde{H}^1} E$; decompose

$$f = \mathrm{avg}_c(f) + \tilde{f} \quad , \quad g = \mathrm{avg}_c(g) + \tilde{g} \quad .$$

To solve

$$f = \mathrm{avg}_c(g) - \lambda L^2 D_s^2 g$$

we have to solve two equations:

$$\begin{array}{ccc} \mathrm{avg}_c(g) = \mathrm{avg}_c(f) & & \lambda L^2\, D_s^2 \tilde{g} = -\tilde{f} \\ \mathbb{R}^n & \oplus & D_c M \ . \end{array}$$

We will now show how to properly solve these coupled equations.

We will need to define some useful objects.

Definition 10.13 *We define the* ***projection operator***

$$\begin{array}{rccc} \Pi_c : & T_c M & \to & D_c M \\ & h & \mapsto & h - \fint_c h \,\mathrm{d}s \end{array} \tag{10.17}$$

Definition 10.14 *When we consider the* derivation with respect to the arc parameter *as a linear operator*

$$\begin{array}{rccc} \mathcal{D}_c : & D_c M & \to & D_c M \\ & h & \mapsto & D_s h \end{array} \tag{10.18}$$

then it admits the inverse, that is the ***primitive operator***

$$\mathcal{P}_c : D_c M \to D_c M \tag{10.19}$$

Proof. The proof is just based on noting that, for any smooth $h \in T_c M$, $h \in D_c M$ iff there is a smooth $k \in T_c M$ with $h = D_s k$. Vice versa, two primitives differ by a constant, hence when $h \in D_c M$ there is exactly one primitive k in $D_c M$ such that $h = D_s k$. □

Example 10.15 *The tangent vector field $D_s c$ is in $D_c M$, and its primitive is*

$$\mathcal{P}_c D_s c = c - avg_c(c) \ .$$

Proposition 10.16 *Fix a curve c, let $L = \mathrm{len}(c)$. We can extend the primitive operator by composing it with the projection; the resulting composite operator can be expressed in convolutional form as*

$$\mathcal{P}_c \Pi_c h = K_c^{\mathcal{P}} \star h \tag{10.20}$$

for any continuous $h \in T_c M$; where the kernel $K_c^{\mathcal{P}}$ is

$$K_c^{\mathcal{P}}(s) := -\frac{s}{L} + \frac{1}{2} \quad \textit{for } s \in [0, L) \tag{10.21}$$

and $K_c^{\mathcal{P}}(s)$ is extended periodically to $s \in \mathbb{R}$ (note that $K_c^{\mathcal{P}}(s)$ jumps at all points of the form $s = nL, n \in \mathbb{Z}$).

The above properties lead to this theorem, that can be used to shed a new light to what was expressed in 10.10 in the previous section.

Theorem 10.17 *Let $f = \nabla_{H^0} E$, $g = \nabla_{\tilde{H}^1} E$; the solution of*

$$f = avg_c(g) - \lambda L^2 D_s^2 g$$

can be expressed as

$$g = avg_c(f) - \frac{1}{\lambda L^2} \mathcal{P}_c \mathcal{P}_c \Pi_c f = avg_c(f) - \frac{1}{\lambda L^2} K_c^{\mathcal{P}} \star K_c^{\mathcal{P}} \star f \ .$$

Corollary 10.18 *In particular, the kernels defined in eqn. (10.10), (10.21) are related by*

$$\tilde{K}_\lambda = \frac{1}{L} - \frac{1}{\lambda L^2} K_c^{\mathcal{P}} \star K_c^{\mathcal{P}}$$

By deriving,

$$D_s K_c^{\mathcal{P}} = \delta_0 - \frac{1}{L} \tag{10.23}$$

where δ_0 is the Dirac's delta*; so we obtain the relations*

$$D_s \tilde{K}_\lambda = -\frac{1}{\lambda L^2} K_c^{\mathcal{P}}$$

$$D_{ss} \tilde{K}_\lambda = -\frac{1}{\lambda L^2} \delta_0 + \frac{1}{\lambda L^3} \ .$$

We also recall this Lemma.

Lemma 10.19 (De la Vallée-Poussin) *Suppose that $f : S^1 \to \mathbb{R}^n$ is integrable and satisfies*

$$\int_c k \cdot f \, \mathrm{d}s = 0$$

for all $k \in D_c M$ smooth; then f is almost everywhere equal to a constant.

For the proof, see Chapter 13 in Ambrosio et al. [2007], or Lemma VIII.1 in Brezis [1986].

We now compute the gradient of the centroid energy (2.9).

Proposition 10.20 (Gradient of the centroid-based energy) *Let in the following again, for simplicity of notation, $\overline{c} = avg_c(c)$ be the center of mass of the curve. Let*

$$E(c) = \frac{1}{2} |\overline{c} - v|^2 \ .$$

The $\tilde{H}^1$ gradient of E is

$$\nabla_{\tilde{H}^1} E(c) = \overline{c} - v + \frac{1}{\lambda L^2} \mathcal{P}_c \Pi_c \big(D_s c \langle (\overline{c} - v), (c - \overline{c}) \rangle \big) \ . \tag{10.24}$$

Proof. We already computed the *Gâteaux differential* of E; but this time we prefer to use the first form of eqn. (2.10), so as to write

$$DE(c;h) = (\overline{c} - v) \cdot \overline{h} + \fint_c \langle \overline{c} - v, c - \overline{c} \rangle (D_s h, D_s c) \, \mathrm{d}s \ .$$

Let $f = \nabla_{\tilde{H}^1} E(c)$ be the $\tilde{H}^1$ gradient of E. The equality

$$DE(c;h) = \langle h, f \rangle_{\tilde{H}^1} \quad \forall h$$

becomes

$$\langle \overline{c} - v - \overline{f}, \overline{h} \rangle \ + \ \fint_c D_s h \cdot \big(\langle \overline{c} - v, c - \overline{c} \rangle D_s c - \lambda L^2 D_s f \big) \ \mathrm{d}s = 0, \quad \forall h \ .$$

Since h is arbitrary, using *de la Vallée-Poussin* Lemma, we obtain that

$$\overline{f} = \overline{c} - v \quad , \quad \lambda L^2 \, D_s f = D_s c \langle (\overline{c} - v), (c - \overline{c}) \rangle + \alpha \ , \tag{10.25}$$

where the constant α is the unique $\alpha \in \mathbb{R}^n$ such that the rightmost term is in $D_c M$. In conclusion we obtain (10.24). $\square$

We similarly compute the gradient of the active contour energy.

Proposition 10.21 (Gradient of geodesic active contour model) *We consider once again the* geodesic active contour *model Caselles et al. [1995], Kichenassamy et al. [1995] (that was presented in Section 2.2) where the energy is*

$$E(c) = \int_c \phi(c(s)) \, \mathrm{d}s$$

with $\phi : \mathbb{R}^2 \to \mathbb{R}^+$ appropriately designed. The gradient with respect to $\tilde{H}^1$ is

$$\nabla_{\tilde{H}^1} E(c) = L avg_c(\nabla \phi(c)) - \frac{1}{\lambda L} \mathcal{P}_c \mathcal{P}_c \Pi_c \big(\nabla \phi(c) \big) + \frac{1}{\lambda L} \mathcal{P}_c \Pi_c \big(\phi(c) D_s c \big) \ . \tag{10.26}$$

Proof. Let us note [22] (recalling eqn. (2.3)) that

$$\nabla_{H^0} E = L \nabla \phi(c) - L D_s(\phi(c) D_s c) \ , \tag{10.27}$$

so the above equation (10.26) can be obtained by the relation in Theorem 10.17.

Alternatively, we know that

$$DE(c;h) = L \fint_c \nabla \phi(c) \cdot h + \phi(c)(D_s h \cdot D_s c) \, \mathrm{d}s \ ;$$

let $f = \nabla_{\tilde{H}^1} E(c)$ be the $\tilde{H}^1$ gradient of E; the equality

$$DE(c;h) = \langle h, f \rangle_{\tilde{H}^1} \quad \forall h$$

[22] We use the definition (10.4) of H^0.

becomes

$$\int_c (L\nabla\phi(c) - \mathrm{avg}_c(f)) \cdot h + D_s h \cdot (L\phi(c) D_s c - \lambda L^2 D_s f)\, \mathrm{d}s = 0$$

imitating the proof of the previous proposition, we obtain (10.26). □ □

We remark a few things.

- Note that the first term in the formula (10.26) is in $\mathbb{R}^n$ while the other two are in $D_c M$.
- Using the kernel $\tilde{K}_\lambda$ that was defined in (10.10), we can rewrite

$$\nabla_{\tilde{H}^1} E(c) = L\tilde{K}_\lambda \star (\nabla\phi(c)) + \frac{1}{\lambda L}\mathcal{P}_c \Pi_c\big(\phi(c) D_s c\big) \tag{10.28}$$

- The formula for the gradient does not require that the curve be twice differentiable: we will use this fact to prove an existence result for the gradient flow, in Theorem 10.31.

10.6 Existence of gradient flows

We recall this definition (that was already presented informally in the introduction).

Definition 10.22 (Gradient descent flow) *Given a differentiable energy $E : M \to \mathbb{R}$, and a metric $\langle,\rangle_c$, let $\nabla E(c)$ be the gradient. Let us fix moreover $c_0 : S^1 \to \mathbb{R}^n$, $c_0 \in M$. The **gradient descent flow** of E is the solution $C = C(t, \theta)$ of the initial value P.D.E.*

$$\begin{cases} \partial_t C = -\nabla E(C) \\ C(0,\theta) = c_0(\theta) \end{cases}$$

We present an example computation for the *geodesic active contour* model on a radially symmetric "image".

Example 10.23 *Let*

$$C(t,\theta) = r(t)\,(\cos\theta, \sin\theta) \quad , \quad \phi(x) = d(|x|)$$

with $r, d : \mathbb{R} \to \mathbb{R}^+$; the energy takes the form

$$E(C) = \int_C \phi(C(t,s))\,\mathrm{d}s = 2\pi\, r(t)\, d(r(t))$$

Then C is the $\tilde{H}^1$ gradient descent iff

$$r'(t) = -\frac{1}{\lambda 2\pi}\big(d(r(t)) + r(t)\, d'(r(t))\big) \ .$$

whereas C is the H^0 gradient descent iff

$$r'(t) = -2\pi\big(d(r(t)) + r(t)d'(r(t))\big) \ ,$$

where we use the definition (10.4) of H^0.

Note also that

$$\partial_t E(C) = r'(t) \left(d(r(t)) + r(t)d'(r(t))\right) \ .$$

Proof. Indeed,

$$\begin{aligned}
\mathrm{d}s &= r(t)\,\mathrm{d}\theta \\
D_s &= \frac{1}{r(t)} D_\theta \\
\mathrm{len}(C) &= 2\pi r(t) \\
\phi(C) &= d(r(t)) \\
\nabla\phi(x) &= \frac{x}{|x|} d'(|x|) \quad \text{for } x \neq 0 \\
\nabla\phi(C) &= d'(r(t))(\cos\theta, \sin\theta) \\
\mathrm{avg}_c(\nabla\phi(C)) &= 0
\end{aligned}$$

$$\begin{aligned}
\mathcal{P}_c \Pi_C \nabla\phi(C) &= \mathcal{P}_c \nabla\phi(C) = r(t)\, d'(r(t))\, (\sin\theta, -\cos\theta) \\
\mathcal{P}_c \mathcal{P}_c \Pi_C \nabla\phi(C) &= -r(t)^2\, d'(r(t))\, (\cos\theta, \sin\theta) \\
\phi(C) D_s C &= d(r(t))(-\sin\theta, \cos\theta) \\
\mathcal{P}_c(\phi(C) D_s C) &= r(t) d(r(t))\, (\cos\theta, \sin\theta) \\
\nabla_{\tilde{H}^1} E(C) &= L\mathrm{avg}_c(\nabla\phi(C)) - \frac{1}{\lambda L} \mathcal{P}_C \mathcal{P}_C \Pi_C \big(\nabla\phi(C)\big) + \frac{1}{\lambda L} \mathcal{P}_C \Pi_C \big(\phi(C) D_s C\big) = \\
&= \frac{1}{\lambda 2\pi} \big(d(r(t)) + r(t)\, d'(r(t))\big)\, (\cos\theta, \sin\theta) \ ;
\end{aligned}$$

whereas for the $\tilde{H}^0$ gradient descent we use the formula (10.27)

$$\begin{aligned}
\nabla_{H^0} E &= L\nabla\phi(C) - L D_s(\phi(C) D_s C) = \\
&= 2\pi\big(r(t) d'(r(t)) + d(r(t))\big)\, (\cos\theta, \sin\theta)
\end{aligned}$$

□

We will show in the following theorems how to prove that $\tilde{H}^1$ gradient flows of common energies are well defined; to this end, we will provide a detailed proof for the *centroid energy* E (2.9) discussed in the previous section, and for the *geodesic active contour* model Caselles et al. [1995]; Kichenassamy et al. [1995] (that was presented in Section 2.2); those proofs illustrates methods that may be used for many other energies.

10.6.1 Lemmas and inequalities

Let us also prepare the proof by presenting some useful inequalities in three Lemma.

Definition 10.24 *We recall from 3.11 that $C^0 = C^0(S^1 \to \mathbb{R}^n)$ is the space of continuous functions, that is a Banach space with norm*

$$\|c\|_0 := \sup_{\theta\in[0,2\pi]} |c(\theta)| \ .$$

Similarly, $C^1 = C^1(S^1 \to \mathbb{R}^n)$ is the space of continuously differentiable functions, that is a Banach space with norm

$$\|c\|_1 := \|c\|_0 + \|c'\|_0$$

where $c'(\theta) = \frac{\partial c}{\partial \theta}(\theta)$ is the usual parametric derivative of c.

Lemma 10.25

- *We will use repeatedly the two following inequalities. If $a_1, a_2, b_1, b_2 \in \mathbb{R}$ then*

$$\begin{aligned} |a_1 b_1 - a_2 b_2| &\leq \tfrac{|b_1|+|b_2|}{2}|a_1 - a_2| + \tfrac{|a_1|+|a_2|}{2}|b_1 - b_2| \\ |\tfrac{a_1}{b_1} - \tfrac{a_2}{b_2}| &\leq \tfrac{|b_1|+|b_2|}{2|b_1 b_2|}|a_1 - a_2| + \tfrac{|a_1|+|a_2|}{2|b_1 b_2|}|b_1 - b_2| \end{aligned} \qquad (i)$$

- *The length functional (from C^1 to $\mathbb{R}$) is Lipschitz, since*

$$\mathrm{len}(c_1) - \mathrm{len}(c_2) \leq 2\pi\|c_1' - c_2'\|_0 \ ; \qquad (ii)$$

- *if $h_1, h_2 : S^1 \to \mathbb{R}^n$ are continuous fields, by (i)*

$$\begin{aligned} |\int_{c_1} h_1(s)\,\mathrm{d}s - \int_{c_2} h_2(s)\,\mathrm{d}s| &\leq \pi(\|h_1\|_0 + \|h_2\|_0)\|c_1' - c_2'\|_0 + \\ &+ \pi(\|c_1'\|_0 + \|c_2'\|_0)\|h_1 - h_2\|_0 \ ; \end{aligned} \qquad \text{(iii)}$$

- *similarly*

$$|⨍_{c_1} h_1(s)\,\mathrm{d}s - ⨍_{c_2} h_2(s)\,\mathrm{d}s| \leq A\|c_1' - c_2'\|_0 + \|h_1 - h_2\|_0 \qquad (iv)$$

  where

$$A := \frac{2\pi^2(\|h_1\|_0 + \|h_2\|_0)(\|c_1'\|_0 + \|c_2'\|_0)}{\mathrm{len}(c_0)\,\mathrm{len}(c_1)} \ . \qquad (v)$$

- *Let then Π_c be the* projection operator *(as in definition (10.17)); note that $(\Pi_c h)' = h'$; from all above inequalities we obtain that*

$$\begin{aligned} \|\Pi_{c_1} h_1 - \Pi_{c_2} h_2\|_0 &\leq A\|c_1' - c_2'\|_0 + 2\|h_1 - h_2\|_0 \\ \|\Pi_{c_1} h_1 - \Pi_{c_2} h_2\|_1 &= \|\Pi_{c_1} h_1 - \Pi_{c_2} h_2\|_0 + \|(\Pi_{c_1} h_1)' - (\Pi_{c_2} h_2)'\|_0 \leq \\ &\leq A\|c_1' - c_2'\|_0 + 2\|h_1 - h_2\|_0 + \|h_1' - h_2'\|_0 \\ &\leq A\|c_1 - c_2\|_1 + 2\|h_1 - h_2\|_1 \ . \end{aligned} \qquad \text{(vi)}$$

We will now prove the local Lipschitz regularity of the operators $\mathcal{P}_c(h)$, $\mathcal{D}_c(h)$ (that were defined in 10.14) in both the two variables h and c. Unfortunately to prove the theorem 10.29 we will need to also compare the action of $\mathcal{P}_c$ for different values of c; since $\mathcal{P}_c h$ is properly defined only when $h \in D_c M$ (and in general $D_{c_1} M \neq D_{c_2} M$), we will actually need to study the composite operator $\mathcal{P}_c \Pi_c h$.

Lemma 10.26 *Let Π_c the projector from $T_c M$ to $D_c M$ (that we defined in eqn. (10.17)). Let c, c_1, c_2 be C^1 immersed curves, and h, h_1, h_2 be continuous fields; then*

$$\|\mathcal{P}_c \Pi_c h\|_1 \leq (2\pi + 2)\|c'\|_0 \|h\|_0 \tag{10.29}$$

and

$$\begin{aligned} \|\mathcal{P}_{c_1} \Pi_{c_1} h_1 - \mathcal{P}_{c_2} \Pi_{c_2} h_2\|_1 \quad &\leq \quad P\,\|c_1' - c_2'\|_0 + \\ &+ \quad Q\,\|h_2 - h_0\|_0 \end{aligned} \tag{10.30}$$

where

$$Q := (1/2 + \pi)(\|c_2'\|_0 + \|c_1'\|_0)\ .$$

and similarly P is the evaluation

$$P = p\big(\|h_1\|_0,\ \|h_2\|_0,\ \|c_1'\|_0,\ \|c_2'\|_0,\ 1/(\mathrm{len}(c_1)\,\mathrm{len}(c_2))\big)$$

of the polynomial

$$p(x_1, x_2, x_3, x_4, x_5) = (\pi + 1/2)(x_1 + x_2) + 2\pi^3 (x_1 x_3 + x_2 x_4)(x_3 + x_4) x_5 \tag{10.31}$$

(that has constant positive coefficients).

Proof. Fix c_i immersed, and $h_i \in T_{c_i} M$ for $i = 1, 2$; let $L_i = \mathrm{len}(c_i)$; let

$$k_i := \mathcal{P}_{c_i} \Pi_{c_i} h_i(s)$$

for simplicity. We rewrite this in the convolutional form

$$k_i(s) := \int_{c_i} K_i(s - \hat{s}) h_i(\hat{s})\,\mathrm{d}\hat{s} \tag{10.32}$$

when integrals are performed in arc parameter, and the kernel (following (10.21)) is

$$K_i(s) := -\frac{s}{L_i} + \frac{1}{2} \qquad \text{for } s \in [0, L_i]$$

and K_i is extended periodically. By substituting and integrating on only one period of K_i,

$$k_i(s) = \int_{s - L_i}^{s} \left(\frac{1}{2} - \frac{s - \hat{s}}{L_i} \right) h_i(\hat{s})\,\mathrm{d}\hat{s}\ .$$

We can then prove easily (10.29): indeed by the convolutional representation (10.32)

$$|(\mathcal{P}_c\Pi_c h)(t)| \leq \|h\|_0 \int_0^{\mathrm{len}(c)} \left| -\frac{s}{\mathrm{len}(c)} + \frac{1}{2} \right| \mathrm{d}s \leq \mathrm{len}(c)\|h\|_0$$

and instead, deriving and applying (iv) from lemma 10.25,

$$|(\mathcal{P}_c\Pi_c h)'(\theta)| = |\Pi_c h(\theta)|\,|c'(\theta)| \leq 2\|h\|_0\|c'\|_0 \ .$$

To prove (10.30), we write (10.32) in θ parameter, as was done in the Definition 10.7; we need the *run-length functions* $l_i : \mathbb{R} \to \mathbb{R}$

$$l_i(\tau) := \int_0^\tau |c_i'(x)|\,\mathrm{d}x$$

so that, setting $s = l_i(\tau)$ and $\hat{s} = l_i(\theta)$ we can write

$$k_i(\tau) = \int_{\tau-2\pi}^{\tau} \Big(\frac{1}{2} - \frac{l_i(\tau) - l_i(\theta)}{L_i}\Big) h_i(\theta)|c_i'(\theta)|\,\mathrm{d}\theta$$

We can eventually estimate the difference $|k_1(\tau) - k_2(\tau)|$ by using *e.g.* the inequality

$$|A_2B_2C_2 - A_1B_1C_1| \leq |A_1B_1|\,|C_2 - C_1| + |A_1C_2|\,|B_2 - B_1| + |B_2C_2|\,|A_2 - A_1|$$

on the difference of the integrands

$$\underbrace{\Big(\frac{1}{2} - \frac{l_2(\tau) - l_2(\theta)}{L_2}\Big)}_{A_2} \underbrace{h_2(\theta)}_{B_2} \underbrace{|c_2'(\theta)|}_{C_2} - \underbrace{\Big(\frac{1}{2} - \frac{l_1(\tau) - l_1(\theta)}{L_1}\Big)}_{A_1} \underbrace{h_1(\theta)}_{B_1} \underbrace{|c_1'(\theta)|}_{C_1}$$

to obtain that the above term is less or equal than

$$\|h_1\|_0\,\|c_2' - c_1'\|_0 + \|c_2'\|_0\,\|h_2 - h_0\|_0 + \|h_2\|_0\|c_2'\|_0 \left| \frac{l_2(\theta) - l_2(\tau)}{L_2} - \frac{l_1(\theta) - l_1(\tau)}{L_1} \right|$$

since $|A_i| \leq 1/2$. In turn (since the formulas defining $k_1(\tau), k_2(\tau)$ the parameters are bound by $\tau - 2\pi \leq \theta \leq \tau$) then

$$\begin{aligned}
|A_2 - A_1| &= \left| \frac{l_2(\tau) - l_2(\theta)}{L_2} - \frac{l_1(\tau) - l_1(\theta)}{L_1} \right| = \left| \frac{\int_\theta^\tau |c_2'(x)|\,\mathrm{d}x}{L_2} - \frac{\int_\theta^\tau |c_1'(x)|\,\mathrm{d}x}{L_1} \right| \leq \\
&\leq \pi\frac{L_1 + L_2}{L_1L_2}\|c_1' - c_2'\|_0 + \pi\frac{\|c_1'\|_0 + \|c_2'\|_0}{L_1L_2}|L_1 - L_2| \leq \\
&\leq 4\pi^2\frac{\|c_1'\|_0 + \|c_2'\|_0}{L_1L_2}\|c_1' - c_2'\|_0
\end{aligned}$$

(by equations (ii) and (i) in lemma 10.25). Summarizing

$$
\begin{aligned}
|k_1(\tau) - k_2(\tau)| \quad &\le \quad 2\pi \left(\|h_1\|_0 + \|h_2\|_0 \|c_2'\|_0 \, 4\pi^2 \frac{\|c_1'\|_0 + \|c_2'\|_0}{L_1 L_2} \right) \|c_1' - c_2'\|_0 + \\
&+ \quad 2\pi \|c_2'\|_0 \, \|h_2 - h_0\|_0 \ .
\end{aligned}
$$

If we derive, $k_i'(\theta) = h_i(\theta)|c_i'(\theta)|$, so

$$
|k_2'(\theta) - k_2'(\theta)| \le \frac{|h_2(\theta)| + |h_1(\theta)|}{2} |c_2'(\theta) - c_1'(\theta)| + \frac{|c_2'(\theta)| + |c_1'(\theta)|}{2} |h_2'(\theta) - h_1'(\theta)| \ .
$$

Symmetrizing we prove (10.30).

□

Lemma 10.27 *Conversely, let c_1, c_2 be two C^1 immersed curves, and h_1, h_2 be two differentiable fields; then (by using once again eqn. (i) from the lemma 10.25)*

$$
\|\mathcal{D}_{c_1} h_1 - \mathcal{D}_{c_2} h_2\|_0 \le \frac{\|c_1\|_1 + \|c_2\|_1}{2\varepsilon^2} \|h_1 - h_2\|_1 + \frac{\|h_1\|_1 + \|h_2\|_1}{2\varepsilon^2} \|c_1 - c_2\|_1 \quad (10.35)
$$

where $\varepsilon = \min(\inf_{S^1} |c_1'|, \inf_{S^1} |c_2'|)$.

10.6.2 Existence of flow for the centroid energy (2.9)

Theorem 10.29 *Let us fix $v \in \mathbb{R}^n$, let $E(c) = \frac{1}{2}|avg_c(c) - v|^2$; let the initial curve $c_0 \in C^1$, then the $\tilde{H}^1$ gradient descent flow of E has an unique solution $C = C(t, \theta)$, for all $t \in \mathbb{R}$, and $C(t, \cdot) \in C^1$.*

Results of a numerical simulation of the above gradient descent flow are shown in figure 12 on the following page.

Before proving the theorem, let us comment on an interesting property of the above gradient flow.

Remark 10.30 *The length of the curves* $\operatorname{len}(C(t, \cdot))$ *is constant during the evolution in t.*

Proof. Indeed it is easy to prove that $\partial_t \operatorname{len}(C(t, \cdot)) = 0$: the Gâteaux differential of the length of a curve c is

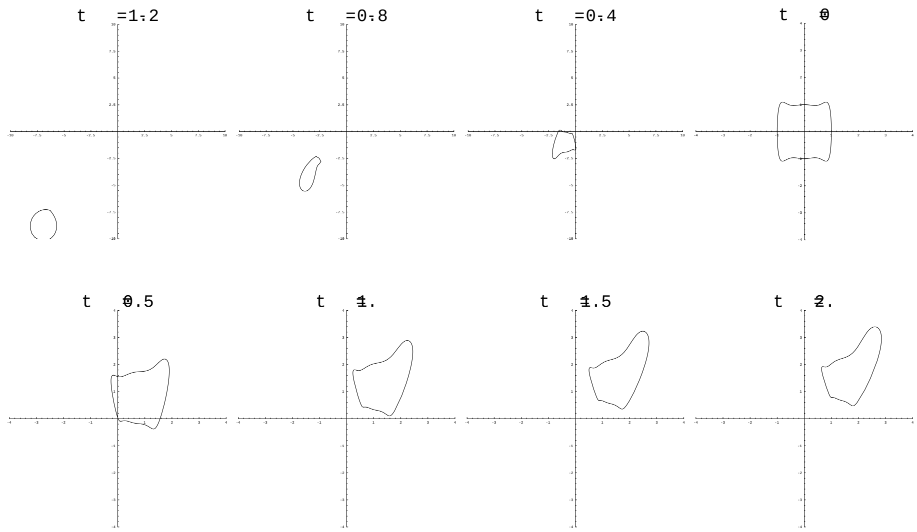
$$
D \operatorname{len}(c; h) = \int_c (D_s c \cdot D_s h) \, \mathrm{d}s \ ;
$$

substituting the value of $D_s C(t, \cdot))$ from eqn. (10.25),

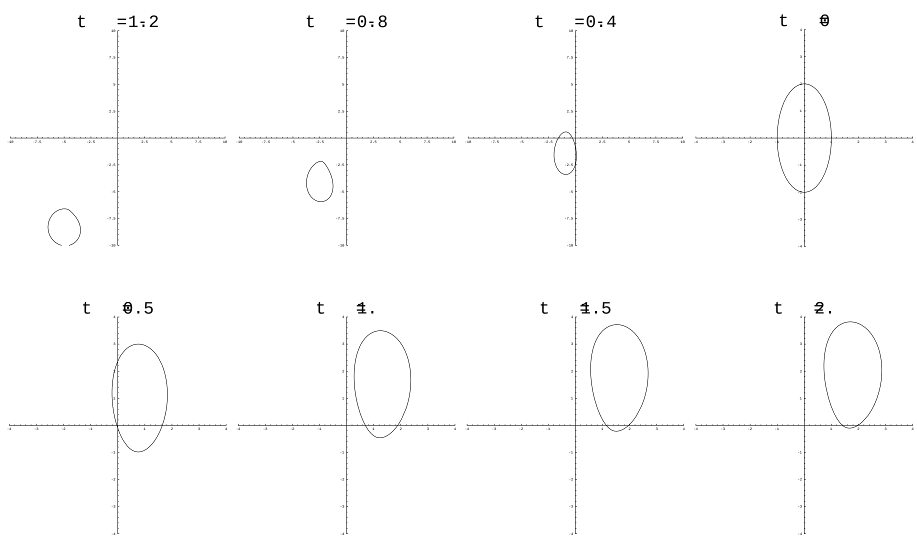
$$
\begin{aligned}
\partial_t \operatorname{len}(C(t, \cdot)) \quad &= \quad D(\operatorname{len}(C))(\partial_t C) \\
&= \quad \frac{1}{\lambda L^2} \int \left\langle D_s C, \big(D_s C \langle (\operatorname{avg}_c(C) - v), (C - \operatorname{avg}_c(C)) \rangle + \alpha \big) \right\rangle \mathrm{d}s = \\
&= \quad \frac{1}{\lambda L^2} \left\langle (\operatorname{avg}_c(C) - v), \int (C - \operatorname{avg}_c(C)) \, \mathrm{d}s \right\rangle + \int (D_s C \cdot \alpha) \, \mathrm{d}s = 0 \ .
\end{aligned}
$$

□

$$\boxed{C(0,\theta) = c_0(\theta) = \big(\cos(\theta), \sin(\theta)(1 + 2\cos(\theta)^4)\big)}$$



$$\boxed{C(0,\theta) = c_0(\theta) = \big(\cos(\theta), 2\sin(\theta)\big)}$$



Two numerical simulation of the $\tilde{H}^1$ gradient descent flow of the centroid energy $E(c) = \frac{1}{2}|avg_c(c) - v|^2$, *for two different initial curves c_0. The constants were set to $\lambda = 1/(4\pi^2)$, $v = (2,2)$. Steps for numerical discretization were set to $\Delta t = 1/30$, $\Delta\theta = \pi/64$ (on $[0, 2\pi]$). Note that plots for $t < 0$ are shown in a smaller scale.*

Figure 12: $\tilde{H}^1$ gradient descent flow of the centroid energy. See 10.20 and 10.29.

We now present the proof of the theorem 10.29.

Proof. We will first of all prove existence and uniqueness of the gradient flow for small time, using the Cauchy–Lipschitz theorem[24] *in the space M of immersed curves, seen as an open subset of the C^1* Banach space *(see Definition 10.24).*

To this end we will prove that $c \mapsto \nabla_{\tilde{H}^1} E(c)$ is a locally Lipschitz functional from C^1 into itself. Afterward, we will directly prove that the solution exists for all times.

So, let us fix $c_1, c_2 \in C^1 \cap M$ that are two curves near c_0. Let us fix $\varepsilon > 0$ by

$$\varepsilon := \inf_\theta |c_0'(\theta)|/2 \ .$$

By "near" we mean that, in all of the following, we will require that $\|c_i - c_0\|_1 < \varepsilon$, for $i = 1, 2$. We will need the following (easy to prove) facts. (In the following, the index i will represent both $1, 2$).

- $\inf_\theta |c_i'(\theta)| \geq \varepsilon$, *so all curves in the neighborhood are immersed.*
- *Setting $a_0 := \|c_0\|_1 + \varepsilon$, we have $\|c_i\|_1 \leq a_0$.*
- *By equation (ii) in lemma 10.25,*

  $$\mathrm{len}(c_0) - 2\pi\|c_0 - c_i\|_1 \leq \mathrm{len}(c_i) \leq \mathrm{len}(c_0) + 2\pi\|c_0 - c_i\|_1$$

  and in particular

  $$2\pi\varepsilon \leq \mathrm{len}(c_i) \leq \mathrm{len}(c_0) + 2\pi\varepsilon \ .$$

Let $f_i = \nabla_{\tilde{H}^1} E(c_i)$ be the gradient, whose formula was expressed in (10.24). We decompose f_i using the relation $T_cM = \mathbb{R}^n \oplus D_cM$, as done previously:

$$\overline{f}_i = \mathrm{avg}_{c_i}(f_i) \in \mathbb{R}^n \quad , \quad \tilde{f}_i = f - \mathrm{avg}_{c_i}(f_i) \in D_{c_i}M$$

and obtain

$$\begin{aligned} \overline{f}_i &= \mathrm{avg}_{c_i}(c_i) - v \\ \tilde{f}_i &= \frac{1}{\lambda L_i^2} \mathcal{P}_{c_i} \Pi_{c_i} h_i, \ \textit{where} \\ & \quad h_i := D_s c_i \langle (\mathrm{avg}_{c_i}(c_i) - v), (c_i - \mathrm{avg}_{c_i}(c_i)) \rangle \ , \end{aligned} \tag{10.36}$$

by rewriting (10.25) in the notation of this proof.

We will then exploit all the inequalities in the lemmas to prove that

$$|\overline{f}_1 - \overline{f}_2| \leq a_1 \|c_1 - c_2\|_1 \quad , \quad \|\tilde{f}_1 - \tilde{f}_2\|_1 \leq a_2 \|c_1 - c_2\|_1$$

for two constants $a_1, a_2 > 0$ and all c_1, c_2 near c_0; this effectively proves that

$$|f_1 - f_2| \leq (a_1 + a_2)\|c_1 - c_2\|_1$$

[24]A.k.a. as the *Picard–Lindelöf theorem*

that is, $\nabla_{\tilde{H}^1} E(c)$ is a locally Lipschitz functional.

The first term is dealt with equation (iv) in lemma 10.25, whence we obtain

$$|\overline{f}_1 - \overline{f}_2| = |\fint_{c_1} c_1(s)\,\mathrm{d}s - \fint_{c_2} c_2(s)\,\mathrm{d}s| \leq a_1 \|c_1 - c_2\|_1 \text{ with } a_1 := \frac{2a_0^2}{\varepsilon^2} \ .$$

By repeated applications of the inequalities listed in lemma 10.25, we prove that, in the designed neighborhood, the following inequalities hold

$$\begin{aligned} \|h_2 - h_1\|_0 &\leq a_3 \|c_2 - c_1\|_1 \\ \|h_i\| &\leq a_4 \end{aligned}$$

for two constants $a_3, a_4 > 0$. So we can choose constants $P, Q > 0$ such that the inequality (10.30) holds uniformly in the neighborhood, and then

$$\begin{aligned} \|\mathcal{P}_{c_1} \Pi_{c_1} h_1 - \mathcal{P}_{c_2} \Pi_{c_2} h_2\|_1 &\leq P\,\|c_1' - c_2'\|_0 + Q\,\|h_2 - h_1\|_0 \leq \\ &\leq (P + Qa_3)\|c_1 - c_2\|_1 \end{aligned}$$

By using (i) and (ii) from lemma 10.25 once again, we can conclude that

$$\|\tilde{f}_1 - \tilde{f}_2\|_1 \leq a_2 \|c_1 - c_2\|_1$$

for a constant $a_2 > 0$. The Cauchy–Lipschitz theorem is now invoked to guarantee that the gradient descent flow does exist and is unique for small times. Let then $C(t, \theta)$ be the solution, that will exist for $t \in (t^-, t^+)$, the maximal interval.

In the following, given any $g = g(t, \theta)$ we will simply write $\|g\|_0$ instead of

$$\|g(t, \cdot)\|_0 = \sup_\theta |g(t, \theta)|$$

and similarly

$$\|g\|_1 = \|g\|_0 + \|\partial_\theta g\|_0 = \sup_\theta (|g(t, \theta)| + |\partial_\theta g(t, \theta)|) \ ;$$

we will also write $\mathrm{len}(C) = \mathrm{len}(C(t, \cdot))$ *for the length of the curve at time t.*

We want to prove that the maximal interval is actually $\mathbb{R}$, that is, $t^+ = -t^- = \infty$. The base and rough idea of the proof is assuming that t^+ or t^- are finite, and derive a contradiction by showing that $\nabla E(C)$ does not blow up *when $t \searrow t^-$ or $t \nearrow t^+$.*

More precisely, we will show that, if $t^- > -\infty$ then

$$\limsup_{t \searrow t^-} \|\nabla E(C)\|_1 < \infty \qquad (**)$$

this implies that the flow $C(t, \cdot)$ admits a limit (inside the Banach space C^1) as $t \to t^-$, and then it may continued (contradicting the fact that t^- is the lowest time limit of existence of the flow). A similar result may be derived when $t \nearrow t^+$ (but we will omit the proof, that is actually simpler).

One key step in showing that $(**)$ *holds is to consider the two fundamental quantities*

$$I(t) := \inf_\theta |\partial_\theta C(t,\theta)| \quad , \quad N(t) := \|C\|_1$$

and prove that

$$\liminf I(t) > 0 \quad , \quad \limsup N(t) < \infty$$

when $t^- > -\infty$ *and* $t \searrow t^-$. [25]. *We proceed in steps.*

- *By remark 10.30, we know that* $\operatorname{len}(C)$ *is constant (for as long as the flow is defined), and then equal to* $\operatorname{len}(c_0)$.
- *At any fixed time, by (10.5),*

$$|C - \overline{C}| \le \operatorname{len}(C)/2 = \operatorname{len}(c_0)/2 \tag{10.37}$$

  where $\overline{C}$ *is the center of mass of* $C(t,\cdot)$*; and then*

$$|C| \le |C - \overline{C}| + |v - \overline{C}| + |v| \le \operatorname{len}(c_0)/2 + \sqrt{2E(C)} + |v| \tag{10.38}$$

- *During the flow, the value of the energy changes with rate*

$$\begin{aligned} \partial_t E(C) &= -\|\nabla_{\tilde{H}^1} E(C)\|^2_{\tilde{H}^1} = \\ &= -|\overline{C} - v|^2 - \frac{1}{\lambda \operatorname{len}(C)^2} \fint_C \langle \overline{C} - v, C - \overline{C}\rangle^2 \, \mathrm{d}s \end{aligned}$$

  using (10.37), we can bound

$$|\partial_t E(C)| \le E(C)(2 + 1/\lambda)$$

  so from Gronwall inequality we obtain that $E(C)$ *does not blow up; consequently by (10.38),* $\|C\|_0$ *does not blow up as well.*

- *Let*

$$H := D_s C \langle (\overline{C} - v), (C - \overline{C})\rangle \ ,$$

  from the above we obtain that H *does not blow up as well, since*

$$\|H(t,\cdot)\|_0 \le \operatorname{len}(c_0)\sqrt{2E(C)} \tag{10.39}$$

- *The parameterization of the curves changes according to the law*

$$\partial_t \log(|\partial_\theta C|^2) = 2\langle D_s \partial_t C, D_s C\rangle = \frac{2H \cdot D_s C}{\lambda \operatorname{len}(c_0)^2}$$

  so

$$|\partial_t \log(|\partial_\theta C|^2)| \le \frac{2\sqrt{2E(C)}}{\lambda \operatorname{len}(c_0)}$$

[25] Actually, by tracking the first part of the proof in detail, it possible to prove that all other constants a_1, a_2, a_3, a_4, P, Q may be bounded in terms of these two quantities $I(t), N(t)$

this proves that

$$\liminf_{t \searrow t^-} I(t) > 0$$

and

$$\limsup_{t \searrow t^-} \sup_{\theta} |\partial_\theta C(t,\theta)| < \infty \ .$$

(As a consequence, $\partial_\theta C(t,\theta) \neq 0$ at all times: curves will always be immersed)

- *So by (10.29)*

  $$\|\mathcal{P}_C \Pi_C H\|_1 \leq (2\pi + 2)\|C'\|_0 \|H\|_0$$

  does not blow up as well.

- *Since both $\|C\|_0$ and $\|\partial_\theta C\|_0$ do not blow up, then $\|C\|_1$ does not blow up.*

- *Decomposing $F = -\partial_t C = \nabla_{\tilde{H}^1} E(C)$ (as was done in (10.36)) we write*

  $$\|\tilde{F}\|_1 = \frac{1}{\lambda \operatorname{len}(c_0)^2} \|\mathcal{P}_C \Pi_C H\|_1$$

  and from all above $\|\tilde{F}\|_1$ does not blow up; but also

  $$|\overline{F}| = |\operatorname{avg}_C(C) - v| \leq \sqrt{2E(C)}$$

  as well; but then $\|\nabla_{\tilde{H}^1} E(C)\|_1$ itself does not blow up, as we wanted to prove.

$\square$

10.6.3 Existence of flow for geodesic active contour

Theorem 10.31 *Let once again*

$$E(c) = \int_c \phi(c(s))\, \mathrm{d}s$$

be the geodesic active contour *energy. Suppose that $\phi \in C^{1,1}_{loc}$ (that is, $\phi \in C^1$ and its derivative is Lipschitz on compact subsets of $\mathbb{R}^n$), let the initial curve be $c_0 \in C^1$; then the gradient flow*

$$\frac{dc}{dt} = -\nabla_{\tilde{H}^1} E(c)$$

exists and is unique in C^1 for small times.

Proof. We rapidly sketch the proof, that is quite similar to the proof of the previous theorem. We show that $\nabla_{\tilde{H}^1} E(c)$ is locally Lipschitz in the C^1 Banach space (see Definition 10.24). Indeed, since $\phi \in C^{1,1}_{\mathrm{loc}}$, the maps

$$\begin{aligned} c &\mapsto \nabla\phi(c) \\ c &\mapsto \phi(c) D_s c \end{aligned}$$

are locally Lipschitz as maps from C^1 to C^0, so (using Lemmas 10.25 and 10.26) we obtain that

$$c \mapsto \operatorname{avg}_c(\nabla\phi(c))$$

is loc.Lip. from C^1 to $\mathbb{R}^n$, and

$$\begin{array}{rcl} c & \mapsto & \mathcal{P}_c\mathcal{P}_c\Pi_c\big(\nabla\phi(c)\big) \\ c & \mapsto & \mathcal{P}_c\Pi_c\big(\phi(c)D_s c\big) \end{array}$$

are locally Lipschitz as maps from C^1 to C^1; combining all together (and using the Lemma 10.25 again)

$$\nabla_{\tilde{H}^1} E(c) = L\operatorname{avg}_c(\nabla\phi(c)) - \frac{1}{\lambda L}\mathcal{P}_c\mathcal{P}_c\Pi_c\big(\nabla\phi(c)\big) + \frac{1}{\lambda L}\mathcal{P}_c\Pi_c\big(\phi(c)D_s c\big)$$

is locally Lipschitz as maps from C^1 to C^1; so by the *Cauchy–Lipschitz theorem* we know that the gradient flow exists for small times. □

Remark 10.32 (Gradient descent of curve length) *Of particular interest is the case when $\phi = 1$, that is $E = L$, the length of the curve. In this case, we already know that the H^0 flow is the* geometric heat flow*. By 10.15, the $\tilde{H}^1$ gradient instead reduces to*

$$\nabla_{\tilde{H}^1} L = \frac{c - avg_c(c)}{\lambda L} \ . \tag{10.40}$$

So the $\tilde{H}^1$ gradient flow constitutes a simple rescaling of the curve about its centroid (!).

It is interesting to notice that the H^1 and $\tilde{H}^1$ gradient flows are well-posed for both ascent and descent while the H^0 gradient flow is only well-posed for descent. This is related to the fact that the H^0 gradient descent flow smooths the curve, whereas the $\tilde{H}^1$ gradient descent (or ascent) has neither a beneficial nor a detrimental effect on the regularity of the curve.

10.7 Regularization of energy *vs* regularization of flow/metric

Imagine an energy E minimized on curves (numerically sampled to p points). Suppose it is not satisfactory in applications (it may be ill posed, or not robust to noise). We have two solutions available.

- Add a regularization term $R(c)$ and minimize $E(c)+\varepsilon R(c)$ by H^0 gradient descent. The numerical complexity of computing $\nabla_{H^0} R$ is of order $O(p)$.
- Minimize $E(c)$ by $\tilde{H}^1$ gradient descent. The numerical complexity of computing $\nabla_{\tilde{H}^1} E$ using the equations in Prop. 10.10 is of order $O(p)$.

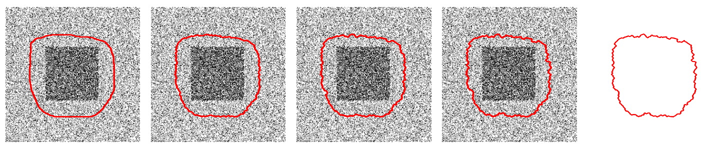
The numerical complexity of the two approaches is similar; but in the second case we are minimizing the original energy. This can bring evident benefits, as shown in the following example (originally presented in the *2006 IMA conference on shape spaces*).

Example 10.33 *We consider a* region-based segmentation *of a synthetic noisy image, where the energy is the* Chan-Vese *energy*

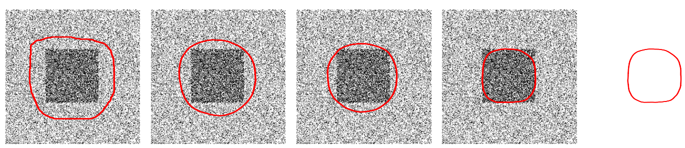
$$E(c) = \int_{c_{in}} (I-u)^2 \, \mathrm{d}A + \int_{c_{out}} (I-v)^2 \, \mathrm{d}A$$

plus the regularization terms $\alpha \, \mathrm{len}(c)$ *for* H^0 *flows.*

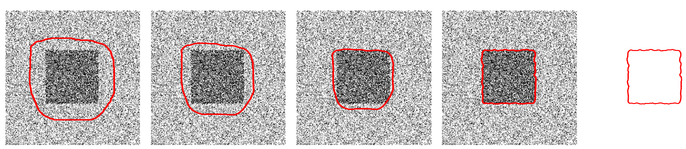
When flowing according to $-\nabla_{H^0} E + \alpha \kappa N$ *and a small value of* $\alpha > 0$*, we obtain the following evolution of the curve, where the curve gets trapped in a local minima due to noise.*



For a larger value of α*, the regularization term forces the curve away from the square shape.*



When evolving using $-\nabla_{\tilde{H}^1} E$*, the curve captures the correct shape on coarse scale, and then segments the noisy boundary of the square further.*



10.7.1 Robustness w.r.to local minima due to noise

The concept of *local* depends on the norm.

Definition 10.34 *A curve* c_0 *is a* local minimum *of an energy* E *iff* $\exists \epsilon > 0$ *such that if* $\|h\|_{c_0} < \epsilon$ *then* $E(c_0) \leq E(c_0 + h)$*.*

Note that Sobolev-type norms dominate H^0-type norm:

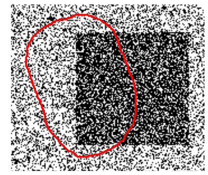
$$\|h\|_{H^0} \leq \|h\|_{\mathrm{Sobolev}},$$

and the norms are <u>not</u> equivalent. As a result the *neighborhood* of critical points in Sobolev-type space is different than in H^0 space.

We present an experimental demonstration of what "local" means (originally presented in Sundaramoorthi et al. [2008a]).

Example 10.35 *1. We initialize a contour in a noisy image.*

2. *We run the H^0 gradient flow on energy*



$$E(c) = E_{cv}(c) + \alpha\,\mathrm{len}(c),$$

   where E_{cv} is the Chan-Vese *energy. Let's call the converged contour c_0; it is shown in the picture on the right.*

3. *We adjust c_0 at* **one sample point** *by* **one pixel**, *and call the modification $\hat{c}_0$. Note: $\hat{c}_0$ is an H^0 (but also "rigidified") local perturbation of c_0.*

4. *We run and compare H^0, "rigidified," and Sobolev gradient flows initialized with $\hat{c}_0$.*

The results are in Figure 13. As we can see, the SAC evolution can escape from the local minimum that is induced by noise. Surprisingly, in the numerical experiments, the **same** *result holds for SAC even without perturbing the local minimum: this is due to the effect of numerical noise.*

Many other examples, on synthetic and on real images, are present in Sundaramoorthi et al. [2007a, 2008a,b].

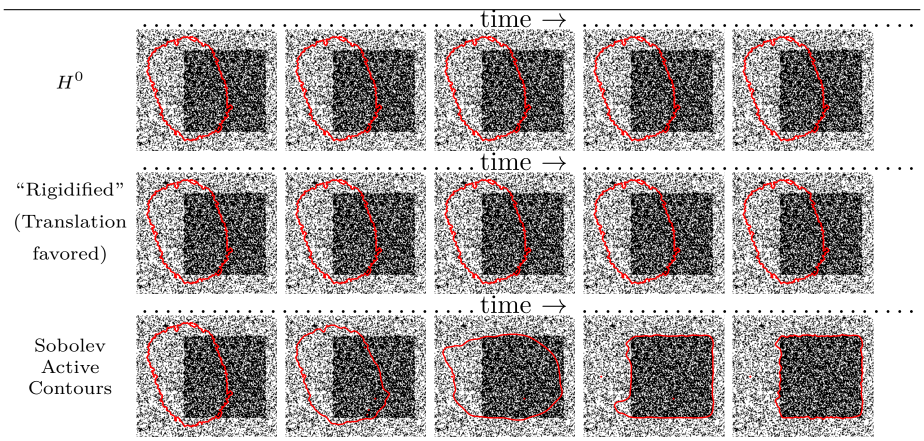


Figure 13: Comparison of minimization flows when initialized near a local minimum; see Example 10.35. *(From Sundaramoorthi et al. [2008a] © 2008 IEEE. Reproduced with permission).*

10.8 New shape optimization energies

Most of what follows was originally presented in Sundaramoorthi et al. [2007b, 2008b].

There are many useful active contour energies that cannot be optimized using H^0 gradient flow. There are mainly two reasons why this happens.

- Some energies result in *ill-posed* H^0 flows.
- Some energies result in high-order PDE and are difficult to implement numerically.

In many cases, we can optimize these energies using Sobolev active contours and avoid ill-posed problems and reduce order of PDE.

We remark that other gradients (e.g. the "rigidified" active contours of Charpiat et al. [2005]) cannot be used in the examples we will present: the minimization flows still are ill posed or high order. Similarly global methods (e.g. graph cuts) cannot deal with these types of energies.

10.8.1 Average weighted length

A simple example that falls in the first category is the **average weighted length** energy that we presented in equation (2.6) in the introduction:

$$E_1(c) = \fint_c \phi(c(s))\,\mathrm{d}s = \mathrm{avg}_c(\phi)\ ;$$

this energy was introduced since it does not present the *short length bias* (that was discussed in §2.4.2).

Let $L_\phi = \int_c \phi\,\mathrm{d}s$ and $L = \mathrm{len}(c)$ so that $E_1(c) = L_\phi/L$, then the gradient is

$$\nabla E_1 = \frac{1}{L}\nabla L_\phi - \frac{L_\phi}{L^2}\nabla L\ .$$

We already presented all the calculus needed to compute the H^0 and $\tilde{H}^1$ gradients of E_1 (see Section 2.4.4 and Example 10.21); let us summarize all the results.

- In the case of H^0 gradient descent flow, the second term is ill posed , since $L_\phi/L^2 > 0$ and $\nabla_{H^0} L$ is the driving term in the *geometric heat flow*.

  This is also clear from the explicit formula

  $$\nabla_{H^0} E_1(c) = [\nabla\phi \cdot N - \kappa(\phi - \mathrm{avg}_c(\phi))]N \tag{10.41}$$

  since $\mathrm{avg}_c(\phi)$ is the *average* value of ϕ, therefore $\phi - \mathrm{avg}_c(\phi)$ is negative on roughly half of the curve; so the second term tries to increase length on half of curve using a *geometric heat flow* (and this is ill posed); whereas the first term $\nabla\phi$ does not stabilize the flow.

- In the case of the Sobolev gradient, we saw in the proof of Theorem 10.31 that the gradients $\nabla_{\tilde{H}^1} L_\phi$ and $\nabla_{\tilde{H}^1} L$ are locally Lipschitz in C^1, so, by using Lemma 10.25 we conclude that the gradient $\nabla_{\tilde{H}^1} E_1$ is locally Lipschitz as well, hence the gradient flow is well defined.

10.8.2 New edge-based active contour models

The *average weighted length* idea can be used to build new energies with interesting applications in tracking.

Let us call

$$E_{\mathrm{old}}(c) = \int_c \phi(c(s))\, \mathrm{d}s$$

the traditional edge-based active contour energy.

When using a Sobolev–type norm, we can consider *non-shrinking* edge-based models, as in the following examples.

Example 10.36 *Let us consider the energy*

$$E_{new}(c) = \int_c \phi(c(s)) \left(L^{-1} + \alpha L \kappa^2(s)\right) \, \mathrm{d}s$$

where

- *the first term is edge-detection without shrinking bias (nor regularity),*
- *and the second term is a kind of* ***elastic regularization*** *(that will be discussed also in the next section) that is moreover relaxed near edges;*

the length terms render the whole energy scale invariant*.*

The frames in Figure 14 on the following page show initialization and final results obtained with different norms and energies.

Example 10.37 *Let us consider a* ***length increasing*** *edge-based model. In this case, we* ***maximize*** *the energy*

$$E_{inc}(c) = \int_c \phi(c(s))\, \mathrm{d}s - \alpha \int_c \kappa^2(s)\, \mathrm{d}s$$

where $\phi > 0$ is high near edges. We compare numerical results with a typical ***edge-based balloon model***

$$E_{bal}(c) = \int_c \phi(c(s))\, \mathrm{d}s - \alpha \int_R \phi(x)\, \mathrm{d}x$$

where R is the area enclosed by c. See figure 15 on the next page.

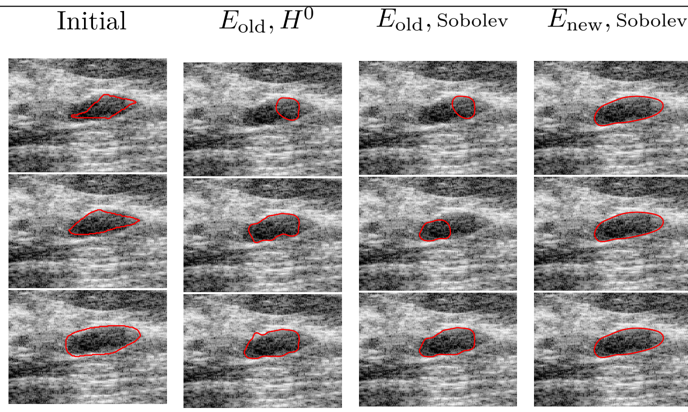


Figure 14: Comparison of segmentations obtained with different energies, as explained in Example 10.36. *(From Sundaramoorthi et al. [2008b] © 2008 IEEE. Reproduced with permission).*

Figure 15: Left to right we see the initial contour, the minima of E_{bal} for $\alpha = 0.2, 0.25, 0.4$ using H^0; the maximum for E_{inc} with $\alpha = 0.1$ using $\tilde{H}^1$ flowing. See Example 10.37. *(From Sundaramoorthi et al. [2008b] © 2008 IEEE. Reproduced with permission).*

10.9 New regularization methods

Typically, in active contour energies a *length penalty* is added to obtain regularity of the evolving contour:

$$E(c) = E_{\text{data}}(c) + \alpha \operatorname{len}(c) \ .$$

It is important to note that this *length penalty* will regularize the curve only if the curve evolution is derived as a H^0 gradient descent flow. If another curve metric is used to derive the gradient descent flow, then *a priori* the *length penalty* may have no regularizing effect on the curve regularity (cf. 10.32).

10.9.1 Elastic regularization

We can now consider an alternative approach for regularity of curve:

$$E(c) = E_{\text{data}}(c) + \alpha \operatorname{len}(c) \int_c \kappa^2(s)\, \mathrm{d}s$$

where $\alpha > 0$ is a fixed constant. This energy favors regularity of curve, but this regularization does not rely on properties of the metric; and the regularization is *scale invariant*. Note that the H^0 gradient flow of this energy is ill posed.

The numerical results (originally presented in Sundaramoorthi et al. [2007b]) are in Fig. 16. In all frames, the final limit of the gradient descent flow is shown; between different frames, the value of α is increased.

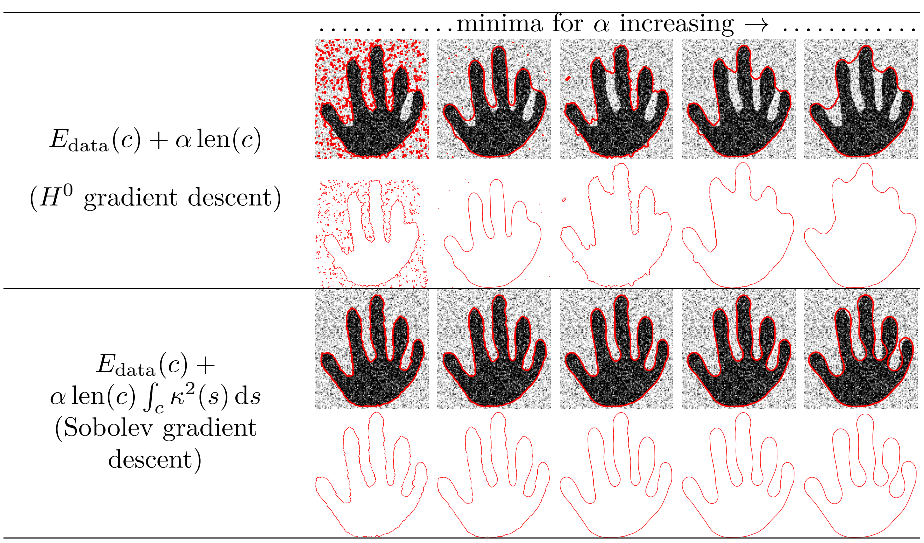


Figure 16: Elastic regularization. *(From Sundaramoorthi et al. [2008b] © 2008 IEEE. Reproduced with permission).*

11 Mathematical properties of the Riemannian space of curves

In this section we study some mathematical properties, and add some final remarks, regarding the Riemannian manifold of geometric curves, when this is endowed with the metrics that were presented previously.

11.1 Charts

Let again

$$M = M_{i,f} = \mathrm{Imm}_f(S^1, \mathbb{R}^n)$$

be the space of all smooth freely-immersed curves.

We already remarked in Remark 4.4 that, if the topology on M is strong enough then the immersed curves are an open subset of all functions. We moreover represent both curves $c \in M$ and deformations $h \in T_c M$ as functions $S^1 \to \mathbb{R}^n$; this is a special structure that is not usually present in abstract manifolds: so we can easily define "charts" for M.

Proposition 11.1 (Charts in $M_{i,f}$) *Given a curve c, there is a neighborhood U_c of $0 \in T_c M$ such that for $h \in U_c$, the curve $c + h$ is still immersed; then this map $h \mapsto c + h$ is the simplest natural candidate to be a chart of $\Phi_c : U_c \to M$. Indeed, if we pick another curve $\tilde{c} \in M$ and the corresponding $U_{\tilde{c}}$ such that $U_{\tilde{c}} \cap U_c \neq \emptyset$, then the equality $\Phi_c(h) = c + h = \tilde{c} + \tilde{h} = \Phi_{\tilde{c}}(\tilde{h})$ can be solved for h to obtain $h = (\tilde{c} - c) + \tilde{h}$.*

The main goal of this section is to study the manifold

$$B = B_{i,f} := M/\mathrm{Diff}(S^1)$$

of geometric curves. Since B is an abstract object, we will actually work with M in everyday calculus. To recombine the two needs, we will identify an unique family of "small deformations" inside M, that has a specific meaning in B. A common choice is to restrict the family of infinitesimal motions h to those such that $h(\theta)$ is orthogonal to the curve, that is, to $c'(\theta)$.

Proposition 11.2 (Charts in $B_{i,f}$) *Let Π be the projection from M to the quotient B. Let $[c] \in B$: we pick a curve c such that $\Pi(c) = [c]$. We represent the tangent space $T_{[c]}B$ as the space of all $k : S^1 \to \mathbb{R}^n$ such that $k(s)$ is orthogonal to $c'(s)$.*

We choose $U_{[c]} \subset T_{[c]}B$ a neighborhood of 0; then the chart is defined by

$$\Psi_{[c]} : U_{[c]} \to B \quad , \quad \Psi_{[c]}(k) := \Pi(c(\cdot) + k(\cdot))$$

that is, it moves $c(u)$ in direction $k(u)$.

If $U_{[c]}$ is small enough, then this chart is a smooth diffeomorphism; and, if we pick another curve $\tilde{c} \in M$ and the corresponding $U_{[\tilde{c}]}$ such that $U_{[\tilde{c}]} \cap U_{[c]} \neq \emptyset$, then the equality

$$\Psi_{[c]}(k) = \Psi_{[\tilde{c}]}(\tilde{k})$$

can be solved for k (this is not so easy to prove: see Michor and Mumford [2006], or 4.4.7 and 4.6.6 in Hamilton [1982]).

11.2 Reparameterization to normal motion

In the preceding proposition we decided to use "orthogonal motion" as distinguished chart for the manifold B. This choice leads also to a "lifting".

Lemma 11.3 *Given any smooth homotopy C of immersed curves, there exists a reparameterization given by a parameterized family of diffeomorphisms $\Phi : [0,1] \times S^1 \to S^1$, so that setting*

$$\tilde{C}(t,\theta) := C(t, \Phi(t,\theta)) \ ;$$

we have that $\partial_t \tilde{C}$ is orthogonal to $\partial_\theta \tilde{C}$ at all points; more precisely,

$$\pi_{\tilde{T}} \partial_t \tilde{C} = 0 \quad , \quad \pi_{\tilde{N}} \partial_t \tilde{C} = \pi_N \partial_t C$$

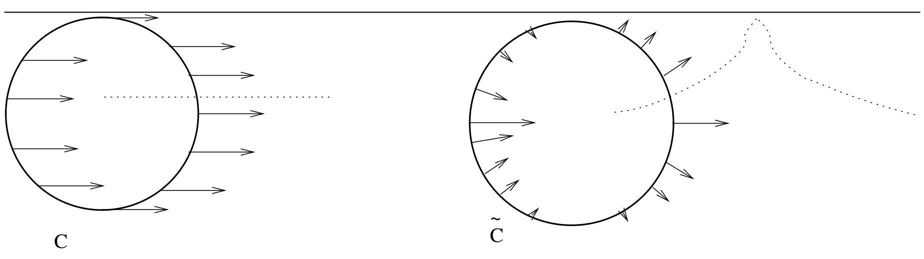
(where the last equality is up to the reparameterization Φ).

For the proof, see Thm. 3.10 in Yezzi and Mennucci [2004b] or §2.5 in Michor and Mumford [2006].

This explains what is commonly done in the *level set method*, where the tangent part of the flow is discarded.

It is unclear that this choice is actually the best possible choice for *computer vision* application. Consider the following example.

Example 11.4 *Suppose that $c(\theta) = (\cos(\theta), \sin(\theta))$ is a planar circle. Let $C(t,\theta) = c(\theta) + vt$ be an uniform translation of c. Let $\tilde{C}$ be as in the previous proposition; the Figure 17 shows the two motions. The two motions C and $\tilde{C}$ coincide in B but are represented differently in M. Which one is best? Both "translation" and "orthogonal motions" seem a natural idea at first glance.*



The dotted line represents the trajectory of a point on the curve.

Figure 17: Translation of a circle as a normal-only motion, by Prop. 11.3.

11.3 The H^0 distance is degenerate

In §C in Yezzi and Mennucci [2004b], or §3.10 in Michor and Mumford [2006], the following theorem is proved.

Theorem 11.5 *The H^0-induced distance is degenerate: the distance between any two curves is 0.*

This results is generalized in Michor and Mumford [2005] to L^2-type metrics of submanifolds of any codimension.

Here we will sketch the main idea of the proof for a very simple case: it is possible to connect two segments

$$c_0(u) = (u, 0) \quad , \quad c_1(u) = (u, 1)$$

with a family of "zigzag" homotopies $C^k(t, \theta)$ so that the H^0 *action* is infinitesimal when $k \to \infty$. A snapshot of the curves along the homotopy, for $t = 0, 1/8 \dots 1$ and $k = 5$, is in Fig. 18.

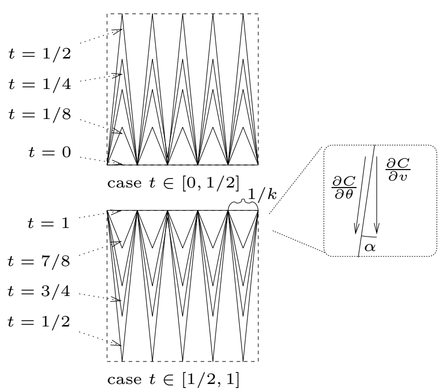


Figure 18: Zig-zag homotopy

In the following, let $C = C^k$ for k large. We recall that the H^0 *action* is

$$\mathbb{E}(C) = \int_0^1 \int_C |\partial_t C|^2 \, \mathrm{d}s \, \mathrm{d}t = \int_0^1 \int_0^1 |\partial_t C|^2 |\partial_\theta C| \, \mathrm{d}\theta \, \mathrm{d}t \ ;$$

and we also define

$$\mathbb{E}_\perp(C) = \int_0^1 \int_C |\pi_N \partial_t C|^2 \, \mathrm{d}s \, \mathrm{d}t = \int_0^1 \int_0^1 |\pi_N \partial_t C|^2 |\partial_\theta C| \, \mathrm{d}\theta \, \mathrm{d}t \qquad (11.1)$$

(this "action" will be explained in Sec. 11.10). The argument goes as follows.

1. The angle α is (the absolute value of) the angle between $\partial_\theta C$ and the vertical direction (that is also the direction of $\partial_t C$). Note that

   - $\alpha = \alpha(k, t)$, and

- at $t = 1/2$ it achieves the maximum $\alpha(k, 1/2) = \arctan(1/(2k))$, and also
- $|\partial_\theta C| = 1/\sin(\alpha)$.

2. Now
$$|\pi_N \partial_t C|^2 |\partial_\theta C| = |\partial_t C|^2 |\partial_\theta C| \sin^2(\alpha) \sim \sin(\alpha)$$
so if k is large, then the angle α is small and then $\mathbb{E}_\perp(C) < \varepsilon \sim 1/k$.

3. We now smooth C just a bit to obtain $\hat{C}$, so that $\mathbb{E}_\perp(\hat{C}) < 2\varepsilon$.

4. We then apply Prop. 11.3 to $\hat{C}$ to finally obtain $\tilde{C}$, that moves only in the normal direction, so
$$\mathbb{E}_\perp(\hat{C}) = \mathbb{E}_\perp(\tilde{C}) = \mathbb{E}(\tilde{C}) < 2\varepsilon$$
as we wanted to show.

11.4 Existence of critical geodesics for H^j

In contrast, it is possible to show that a Sobolev-type metric does admit critical geodesics.

Theorem 11.6 (4.3 in Michor and Mumford [2007]) *Consider the Sobolev type metric*
$$\langle h, k \rangle_{G_c^n} = \int_c \langle h, k \rangle + \langle D_s^n h, D_s^n k \rangle \, \mathrm{d}s$$
with $n \geq 1$. Let $k \geq 2n + 1$; suppose c, h are in the (usual and parametric) H^k Sobolev space *(that is, c, h admit k-th (weak)derivative, that is square integrable). Then it is possible to shoot the geodesic, starting from c and in direction h, for short time.*

In §4.8 in Michor and Mumford [2007] it is suggested that the theorem may similarly hold for Sobolev metrics with length-dependent scale factors.

11.5 Parameterization invariance

Let M be the manifold of (freely)immersed curves. Let $\langle h_1, h_2 \rangle_c$ be a Riemannian metric, for $h_1, h_2 \in T_c M$; let $\|h\|_c = \sqrt{\langle h, h \rangle_c}$ be its associated norm. Let $\mathbb{E}(\gamma) := \int_0^1 \|\dot{\gamma}(v)\|^2_{\gamma(v)} \, \mathrm{d}v$ be the *action*.

Let $C : [0, 1] \times S^1 \to S^1$ be a smooth homotopy. We recall the definition that we saw in Section 4.6, and add a second stronger version.

Definition 11.7 *A Riemannian metric is*

- ***curve-wise parameterization invariant** when the metric does not depend on the parameterization of the curve, that is $\|\tilde{h}\|_{\tilde{c}} = \|h\|_c$ when $\tilde{c}(t) = c(\varphi(t))$ and $\tilde{h}(t) = h(\varphi(t))$;*

- ***homotopy-wise parameterization invariant*** *if, for any* $\varphi : [0,1] \times S^1 \to S^1$ *smooth,* $\varphi(t,\cdot)$ *a diffeomorphism of* S^1 *for all fixed* t*, let* $\tilde{C}(t,\theta) = C(t,\varphi(t,\theta))$*, then* $\mathbb{E}(\tilde{C}) = \mathbb{E}(C)$*.*

It is not difficult to prove that the second condition implies the first. There is an important remark to note: a homotopy-wise-parameterization-invariant Riemannian metric cannot be a proper metric.

Proposition 11.8 *Suppose that a Riemannian metric is* homotopy-wise parameterization invariant*; if* $h_1, h_2 \in T_cM$ *and* h_1 *is tangent to* c *at all points, then* $\langle h_1, h_2\rangle_c = 0$*. So the Riemannian metric in this case is actually a* ***semimetric*** *(and* $\|\cdot\|_c$ *is not a norm, but rather a* seminorm *in* T_cM*).*

Proof. Let $C(t,u) = c(\varphi(t,u))$ with $\varphi(t,\cdot)$ be a time-varying family of diffeomorphisms of S^1. So $\mathbb{E}(C) = \mathbb{E}(c) = 0$, that is

$$\int_0^1 \|c'\partial_t\phi\|_c^2 \,\mathrm{d}t = 0$$

so $\|c'\partial_t\phi\|_c = 0$; but note that, by choosing appropriately ϕ, we may represent any $h \in T_cM$ that is tangent to c as $h = c'\partial_t\phi$. By polarization, we obtain the thesis. □

11.6 Standard and geometric distance

The **standard distance** $d(c_0, c_1)$ in M is the infimum of $\mathrm{Len}(\gamma)$ where γ is any homotopy that connects $c_0, c_1 \in M$ (cf. Definition 3.25).

We are though interested in studying metrics and distances in the quotient space $B := M/\mathrm{Diff}(S^1)$.

We suppose that the metric G is "*curve-wise parameterization invariant*"; then G may be projected to $B := M/\mathrm{Diff}(S^1)$. So in the following we will use a different definition of "geometric distance".

Definition 11.9 (geometric distance) $d_G(c_0, c_1)$ *is the infimum of the length* $\mathrm{Len}(C)$ *in the class of all homotopies* C *connecting the curve* c_0 *to any reparameterization* $c_1 \circ \phi$ *of* c_1*.*

This implements the quotienting formula that we saw in Definition 3.6 (in this case, the group is $\mathcal{G} = \mathrm{Diff}(S^1)$): so we will use d_G as a distance on B.

At the same time, note that in writing $d_G(c_0, c_1)$ we are abusing notation: d_G is not a distance in the space M, but rather it is a *semidistance*, since the distance between c and a reparameterization $c \circ \phi$ is zero.

11.7 Horizontal and vertical space

Consider a metric G (curve-wise parameterization invariant) on M. Let Π once again be projection from $M := \mathrm{Imm}_f(S^1)$ to the quotient $B = B_{i,f} = M/\mathrm{Diff}(S^1)$.

We present a list of definitions (see also the figure 19 on the next page).

Definition 11.10 • *The* ***orbit*** *is* $O_c = [c] = \{c \circ \phi \mid \phi\} = \Pi^{-1}(\{c\})$.

- *The* ***vertical space*** V_c *is the tangent to* O_c*:*

$$V_c := T_c O_c \subset T_c M$$

that can be explicitly written as

$$V_c := \{h = b(s)c'(s) \mid b : S^1 \to \mathbb{R}\}$$

i.e. all the vector fields h *where* $h(s)$ *is tangent to* c*.*

- *The* ***horizontal space*** W_c *is the orthogonal complement inside* $T_c M$

$$W_c := V_c^{\perp} \ .$$

Note that W_c *depends on* G*, but* V_c *does not.*

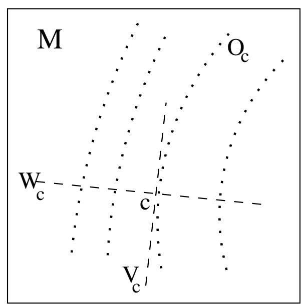


Figure 19: The action of $\mathrm{Diff}(S^1)$ on M. The orbits O_c are dotted, the spaces W_c and V_c are dashed.

The figure 20 may also help in understanding. The whole orbit O_c is projected to $[c]$. The vertical space V_c is the kernel of $D\Pi_c$, so it is projected to 0.

11.8 From curve-wise parameterization invariant to homotopy-wise parameterization invariant

In the following sections, we just aim to give an overview of the theory; we will not dwelve in depicting the most general hypotheses such that all presented results do hold for a generic metric G; we will though, case by case, present references to how the results can be proven for the metrics discussed in this paper, and in particular the Sobolev-type metrics.

We suppose that this theorem holds.

Theorem 11.11 *Given any* $h \in T_c M$*, there is an unique minimum point for*

$$\min_{x \in W_c} \|x - h\|_G$$

and the minimum is called the ***horizontal projection*** *of* h*.*

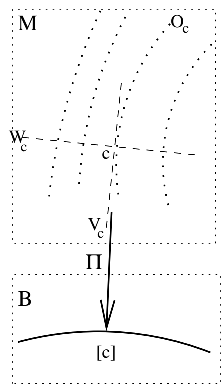


Figure 20: The vertical and horizontal spaces, and the projection Π from M to B.

This theorem is verified when G is a Sobolev-type metrics, see §4.5 in Michor and Mumford [2007]; and it is trivially verified by H^0.

Definition 11.12 *The* ***horizontally projected metric*** $G^\perp$ *is defined by*

$$\langle h,k\rangle_{G^\perp,c} := \langle \tilde h,\tilde k\rangle_{G,c} \tag{11.2}$$

where $\tilde h,\tilde k$ *are the projections of* h,k *to* W_c.

Proposition 11.13 *Equivalently the norm* $\|h\|_{G^\perp}$ *can be defined by*

$$\|h\|_{G^\perp} = \inf_b \|h+bc'\|_G \tag{11.3}$$

where the infimum is in the class of smooth $b : S^1 \to \mathbb{R}$.

Proof. Indeed by the projection theorem,

$$\inf_b \|h+bc'\|_G = \|\tilde h\|_G$$

where $\tilde h$ is the projections of h to W_c. By polarization we obtain (11.2). □

It is easy to prove that $G^\perp$ is *curve-wise parameterization invariant*, but moreover

Proposition 11.14 $G^\perp$ *is* homotopy-wise parameterization invariant.

Proof. Let $\tilde C(t,\theta) = C(t,\varphi(t,\theta))$, then

$$\partial_t \tilde C = \partial_t C + C'\varphi'$$

so at any given time t

$$\|\partial_t \tilde C(t,\cdot)\|_{G^\perp} = \|\partial_t C + C'\varphi'\|_{G^\perp} = \|\partial_t C\|_{G^\perp}$$

by eqn. (11.3), where the terms RHS are evaluated at $(t, \varphi(t, \cdot))$; since $G^\perp$ is *curve-wise parameterization invariant*, then

$$\|\partial_t \tilde{C}(t, \cdot)\|_{G^\perp} = \|\partial_t C(t, \cdot)\|_{G^\perp} \ .$$

□

Consequently,

Corollary 11.15 *G is homotopy-wise parameterization invariant if and only if*

$$\|h\|_G = \|h + bc'\|_G$$

for all b.

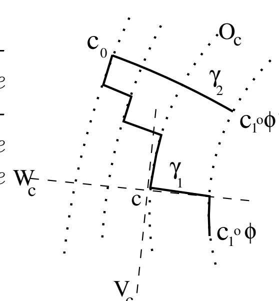
Summarizing all above theory, we obtain a method to understand/study/design the metric G on B, following these steps:

1. choose a metric G on M that is curve-wise parameterization invariant
2. G generates the horizontal space $W = V^\perp$
3. project G on W to define $G^\perp$ that is homotopy-wise parameterization invariant.

11.8.1 Horizontal $G^\perp$ as length minimizer

As we defined in 11.9, to define the distance we minimize the length of the paths $\gamma : [0, 1] \to M$ that connect a curve c_0 to a reparameterization $c_1 \circ \phi$ of the curve c_1. Consider the following sketchy example.

Example 11.16 *Consider two paths γ_1 and γ_2 connecting c_0 to reparameterizations of c_1. Recall that the space W_c is orthogonal to the orbits, the space V_c is tangent to the orbits. The path γ_1 (that moves in some tracts along the orbits, with $\dot{\gamma}_1 \in V_\gamma$) is longer than the path γ_2.*



The above example explains the following lemma.

Lemma 11.17 *Let us choose G to be a* curve-wise parameterization invariant *metric (so, by Prop. 11.8, it cannot be homotopy-wise parameterization invariant).*

1. *Let $\tilde{C}(t, \theta) = C(t, \varphi(t, \theta))$ where $\varphi(t, \cdot)$ is a diffeomorphism for all fixed t. Minimize*

$$\min_{\varphi} \mathbb{E}_G(\tilde{C})$$

Then at the minimum $\tilde{C}^*$, $\partial_t \tilde{C}^*$ is horizontal at every point, and

$$\mathbb{E}_G(\tilde{C}^*) = \mathbb{E}_{G^\perp}(\tilde{C}^*) \ .$$

2. *So the distance can be equivalently computed as the infimum of* $\mathrm{Len}_{G^\perp}$ *or of* Len_G*; and distances are equal,* $d_{G^\perp} = d_G$*.*

In practice, when computing minimal geodesic (possibly by numerical methods), it is usually best to minimize $\mathbb{E}_G$, since it also penalizes the vertical part of the motion, and hence the minimization will produce a "smoother" geodesic.

The proof of the above lemma is based on the validity of *"the lifting Lemma"*:

Lemma 11.18 (lifting Lemma) *Given any smooth homotopy C of immersed curves, there exists a reparameterization given by a parameterized family of diffeomorphisms $\Phi : [0,1] \times S^1 \to S^1$, so that setting*

$$\tilde{C}(t,\theta) := C(t, \Phi(t,\theta))$$

we have that $\partial_t \tilde{C}$ is in the horizontal space W_C at all times t.

For the metric H^0, this reduces to lemma 11.3 in this paper; for Sobolev-type metrics, the proof is in §4.6 in Michor and Mumford [2007].

11.9 A geometric gradient flow is horizontal

Proposition 11.19 *If E is a* geometric energy *of curves, in the sense that E is invariant w.r.to reparameterizations $\phi \in \mathrm{Diff}^+(S^1)$; then ∇E is horizontal.*

Proof. Let c be a smooth curve; suppose for simplicity (and with no loss of generality) that it has $\mathrm{len}(c) = 2\pi$ and that it is parameterized by arc parameter, so that $|c'| \equiv 1$ and $c' = D_s c$. Let $b : S^1 \to \mathbb{R}$ be smooth, and define $\phi : \mathbb{R} \times S^1 \to S^1$ by solving the b flow, that is the ODE

$$\partial_t \phi(t,\theta) = b(\phi(t,\theta))$$

with initial data $\phi(0,\theta) = \theta$. Note that, for all t, $\phi(t,\cdot) \in \mathrm{Diff}^+(S^1)$. Let $C(t,\theta) = c(\phi(t,\theta))$, note that

$$\partial_t C(t,\theta) = b(\phi(t,\theta)) c'(\phi(t,\theta))$$

then $E(C) = E(c)$, so that (deriving and setting $t = 0$)

$$\partial_t E(C) = 0 = \langle \nabla E(c), bc' \rangle$$

and we conclude by arbitrariness of b. $\square$

11.10 Horizontality according to H^0

Proposition 11.20 *The **horizontal space** W_c w.r.to H^0 is*

$$W_c := \{h : h(s) \perp c'(s) \forall s\}$$

that is the space of vector fields orthogonal to the curve.

So Prop. 11.19 guarantees that the H^0 gradients of geometric energies have only a normal component w.r.to the curve — and this couples well with *level set methods*.

The horizontal version of H^0 is

$$\Big\langle h_1, h_2 \Big\rangle_{H^{0,\perp}} := \int_c \pi_N h_1 \cdot \pi_N h_2 \,\mathrm{d}s \tag{11.4}$$

where $\pi_N h_1(s)$ is the component of $h_1(s)$ that is orthogonal to c' (as was defined in Section 7.1); note that the action of this (semi)metric was already presented in eqn. (11.1).

So a good paradigm is to think at the metric H^0 as

$$\begin{array}{rl} \int_c h_1 \cdot h_2 \,\mathrm{d}s & \quad \text{on } M, \\ \int_c \pi_N h_1 \cdot \pi_N h_2 \,\mathrm{d}s & \quad \text{on } B. \end{array}$$

Remark 11.21 *The horizontal space W_c is isometric (thru $D\Pi_c$) with the tangent space $T_{[c]}B$: this implies that we may decide to use W_c and Π_c as a "local chart" for B. If the metric is $G = H^0$, this brings us back the chart 11.2; with other metrics, this process instead defines a novel choice of chart.*

The same reasoning above holds for the H^A metric (from Sec. 9.2), and for the Finsler metrics F^1 and F^∞ (from Section 7); in this latter case, we already presented the horizontally projected version of the metrics.

For all above reasons, it is common thinking that *"a good movement of a curve is a movement that is orthogonal to the curve"*. But, when using a different metric (as for example, the Sobolev-type metrics) the above is not true anymore.

11.11 Horizontality according to H^j

Horizontality according to H^j is a complex matter. To study the horizontally projected metric, we need to express the solution of

$$\inf_b \|h + bc'\|_{H^j}$$

in closed form, and this needs (pseudo)differential operators: see Michor and Mumford [2007] for details. We just remark that in this case, *"a good movement of a curve is **not necessarily** a movement that is orthogonal to the curve."*

11.12 Horizontality is for any group action

In the above we studied the horizontality properties of the quotient $B = M/\mathrm{Diff}(S^1)$. But the theory works in general for any other quotient. So we may apply the same study and ideas to further quotients (such as $B/E(n)$).

11.13 Momenta

What follows comes from §2.5 in Michor and Mumford [2007].

Consider again the action of a group $\mathcal{G}$ on a Riemannian manifold M (as defined in Sect. 3.8). Most groups that act on curves are smooth differentiable manifolds themselves, and the group operations are smooth: so they are **Lie groups**. In our cases of interest, the actions are smooth as well.

We want to study some differentiable properties of the action $g \cdot m$; using the rule (3.5), we see that we can restrict to studying the case $e \cdot m$, where $e \in \mathcal{G}$ is the identity element.

11.13.1 Conservation of momenta

Given a curve $c \in M$, and a tangent vector $\xi \in T_e\mathcal{G}$ (where $T_e\mathcal{G}$ 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 $\zeta = \zeta_{\xi,c} \in T_cM$ (depending linearly on ξ). This direction ζ is intuitively *"the infinitesimal motion of c, when we infinitesimally act in direction ξ on c"*.

To exemplify the above process, we provide a simple toy example.

Example 11.22 (Rotation action on the plane) *The group of 2D rotations $\mathcal{G} = \mathbb{R}/(2\pi)$ (the real line modulus 2π, with the $+$ operation) acts on the plane, as follows*

$$\begin{array}{ccc} \mathcal{G} \times \mathbb{R}^2 & \to & \mathbb{R}^2 \\ g, m & \mapsto & g \cdot m = R_g m \end{array}$$

with

$$R_g := \begin{pmatrix} \cos g & -\sin g \\ \sin g & \cos g \end{pmatrix} .$$

Since $T_e\mathcal{G} = \mathbb{R}$, there is only one direction in the tangent to $\mathcal{G}$, hence the derivative in direction $\xi = 1$ is just the standard derivative d/dg; similarly $T\mathbb{R}^2 = \mathbb{R}^2$; by deriving the action in point $g = e = 0$ (the identity element), we obtain that

$$\zeta_{1,m} = Jm$$

with

$$J := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} .$$

Note that Jm is the vector (normal to m) obtained by rotating m of an angle $\pi/2$ counterclockwise. The physical interpretation is that, when rotating the plane in counterclockwise direction with speed $\xi = 1$, each point m will move with velocity $\zeta = Jm$.

In the above toy example, the result ζ is vector; but in the case of immersed curves M, since $\zeta \in T_cM$, then $\zeta = \zeta(\theta) : S^1 \to \mathbb{R}^n$ is a vector field.

We then recall a very interesting condition by Emmy Noether

Theorem 11.23 *Suppose that the Riemannian metric $\langle\cdot,\cdot\rangle_c$ on M is invariant w.r.to the action of $\mathcal{G}$: this is equivalent to saying that the action is an* isometry. *Let $\gamma(t)$ be a critical geodesic: then*

$$\left\langle \zeta_{\xi,\gamma(t)}, \dot\gamma(t) \right\rangle_{\gamma(t)} \tag{11.5}$$

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

The quantity (11.5) is called "the **momentum** of the action".

As an alternative interpretation, note that the vectors ζ that are obtained by deriving the action are exactly all the vectors in the vertical spaces V_c of the corresponding action $\mathcal{G}$ (since ζ are infinitesimal motions inside the orbit). Recall that $h \in T_cM$ is horizontal (that is, $h \in W_c$) iff it is orthogonal to V_c $\langle \zeta, h'\rangle = 0$ for all $\zeta \in V_c$. So, as corollary of Emmy Noether's theorem, we obtain that

Corollary 11.24 *if a geodesic is shot in a horizontal direction $\dot\gamma(0)$, then the geodesic will be horizontal for all subsequent times.*

The following are examples of momenta that are related to the actions on curves (that we saw in Example 4.8).

Example 11.25

- *The* rescaling *group is represented by $\mathbb{R}^+$, that is one dimensional, so there is only one tangent direction $\xi = 1$ in $\mathbb{R}^+$. The action is $l, c \mapsto lc$, so there is only one direction, that is $\zeta = c$.*
- *The* translation *group is represented by $\mathbb{R}^n$, that is a vector space, so the tangent directions are $\xi \in \mathbb{R}^n$; the action is $\xi, c \mapsto \xi + c$; then $\zeta = \xi$ (and note that this is constant in θ).*
- *The* rotation *group is represented by orthogonal matrixes, so we set $\mathcal{G} = O(n)$; the group identity e is the identity matrix; the action is matrix-vector multiplication $A, c(\theta) \mapsto Ac(\theta)$; the tangent $T_e\mathcal{G}$ is the set of the antisymmetric matrixes $B \in \mathbb{R}^{n\times n}$; then $\zeta = Bc$.*
- *The* reparameterization *group is $\mathcal{G} = \mathrm{Diff}(S^1)$; a tangent vector is a scalar field $\xi : S^1 \to \mathbb{R}$; the action is the composition $\phi, c \mapsto c \circ \phi$; we in the end have that*
$$\zeta(\theta) = \xi(\theta)c'(\theta)$$
(where $c' = D_\theta c$) that is, ζ is a generic vector field parallel to the curve.

For the $\tilde H^1$ metric, this implies the following properties of a geodesic path $\gamma(t)$.

Proposition 11.26

- *Scaling momentum:*
$$\langle \gamma, \dot\gamma\rangle_{\tilde H^1} = avg_c(\gamma)\cdot avg_c(\dot\gamma) + \lambda L^2 \fint D_s\gamma \cdot D_s\dot\gamma \,\mathrm{d}s = constant.$$

- *Linear momentum:* $avg_c(\dot\gamma)$ *is constant, since, for all* $\xi \in \mathbb{R}^n$,

$$\langle \xi, \dot\gamma \rangle_{\tilde H^1} = \xi \cdot avg_c(\dot\gamma) = constant;$$

since ξ *is arbitrary, this means that* $avg_c(\dot\gamma)$ *is constant in* t.

- *Angular momentum: for any antisymmetric matrix* $B \in \mathbb{R}^{n\times n}$,

$$\langle B\gamma, \dot\gamma \rangle_{\tilde H^1} = (B\,avg_c(\gamma)) \cdot avg_c(\dot\gamma) + \lambda L^2 ⨍ (BD_s\gamma) \cdot (D_s\dot\gamma)\,\mathrm{d}s = constant.$$

- *Reparameterization momentum: for any scalar field* $\xi : S^1 \to \mathbb{R}$, *setting*

$$\zeta(\theta, t) = \xi(\theta)\gamma'(\theta, t)$$

we get

$$\langle \zeta, \dot\gamma \rangle_{\tilde H^1} = avg_c(\xi\gamma') \cdot avg_c(\dot\gamma) + \lambda L^2 ⨍ D_s(\xi\gamma') \cdot D_s\dot\gamma = constant;$$

integrating by parts,

$$avg_c(\xi\gamma') \cdot avg_c(\dot\gamma) - \lambda L^2 ⨍ (\xi\gamma') \cdot D_{ss}\dot\gamma = constant;$$

since ξ *is arbitrary, this means that*

$$\gamma' \cdot avg_c(\dot\gamma) - \lambda L^2 \gamma' \cdot D_{ss}\dot\gamma$$

is constant in t, *for any* $\theta \in S^1$.

Contents

Index

References

Luigi Ambrosio, Giuseppe Da Prato, and Andrea Mennucci. An introduction to measure theory and probability, 2007. URL `http//dida.sns.it/dida2/cl/07-08/folde2/pdf0`.

T. M. Apostol. *Mathematical Analysis*. Addison - Wesley, 1974.

C. J. Atkin. The Hopf-Rinow theorem is false in infinite dimensions. *Bull. London Math. Soc.*, 7(3):261–266, 1975. doi: 10.1112/blms/7.3.261.

H. Brezis. *Analisi Funzionale*. Liguori Editore, Napoli, 1986. (italian translation of *Analyse fonctionelle*, Masson, 1983, Paris).

J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698, 1986. ISSN 0162-8828.

V. Caselles, F. Catte, T. Coll, and F. Dibos. A geometric model for edge detection. *Num. Mathematik*, 66:1–31, 1993.

V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In *Proceedings of the IEEE Int. Conf. on Computer Vision*, pages 694–699, Cambridge, MA, USA, June 1995.

T. Chan and L. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, February 2001.

G. Charpiat, O. Faugeras, and R. Keriven. Approximations of shape metrics and application to shape warping and empirical shape statistics. *Foundations of Computational Mathematics*, 2004. doi: 10.1007/s10208-003-0094-xgg819. INRIA report 4820 (2003).

G. Charpiat, R. Keriven, J.P. Pons, and O. Faugeras. Designing spatially coherent minimizing flows for variational problems based on active contours. In *ICCV*, 2005. doi: 10.1109/ICCV.2005.69.

G. Charpiat, P. Maurel, J.-P. Pons, R. Keriven, and O. Faugeras. Generalized gradients: Priors on minimization flows. *International Journal of Computer Vision*, 2007. doi: 10.1007/s11263-006-9966-2.

Y. Chen, H. Tagare, S. Thiruvenkadam, F. Huang, D. Wilson, K. Gopinath, R. Briggs, and E. Geiser. Using prior shapes in geometric active contours in a variational framework. *International Journal of Computer Vision*, 50(3): 315–328, Dec 2002.

D. Cremers and S. Soatto. A pseuso distance for shape priors in level set segmentation. In *IEEE Int. Workshop on Variational, Geometric and Level Set Methods*, pages 169–176, 2003.

Manfredo Perdigão do Carmo. *Riemannian Geometry*. Birkhäuser, 1992.

Alessandro Duci and Andrea Mennucci. Banach-like metrics and metrics of compact sets. 2007.

Alessandro Duci, Anthony J. Yezzi, Sanjoy K. Mitter, and Stefano Soatto. Shape representation via harmonic embedding. In *International Conference on Computer Vision (ICCV03)*, volume 1, pages 656 – 662, Washington, DC, USA, 2003. IEEE Computer Society. ISBN 0-7695-1950-4. doi: 10.1109/ICCV.2003.1238410.

Alessandro Duci, Anthony J. Yezzi, Stefano Soatto, and Kelvin Rocha. Harmonic embeddings for linear shape. *J. Math Imaging Vis*, 25:341–352, 2006. doi: 10.1007/s10851-006-7249-8.

J. Eells and K. D. Elworthy. Open embeddings of certain Banach manifolds. *Ann. of Math. (2)*, 91:465–485, 1970.

Ivar Ekeland. The Hopf-Rinow theorem in infinite dimension. *J. Differential Geom*, 13(2):287–301, 1978.

A. T. Fomenko. *The Plateau problem*. Studies in the development of modern mathematics. Gordon and Breach, 1990.

M. Gage and R. S. Hamilton. The heat equation shrinking convex plane curves. *Journal of Differential Geometry*, (23):69–96, 1986.

Joan Glaunès, Alain Trouvé, and Laurent Younes. Modeling planar shape variation via Hamiltonian flows of curves. In Anthony Yezzi and Hamid Krim, editors, *Analysis and Statistics of Shapes*, Modeling and Simulation in Science, Engineering and Technology, chapter 14. Birkhäuser - Springer - Verlag, 2005.

M. Grayson. The heat equation shrinks embedded planes curves to round points. *J. Differential Geometry*, (26):285–314, 1987.

Richard S. Hamilton. The inverse function theorem of Nash and Moser. *Bull. Amer. Math. Soc. (N.S.)*, 7(1):65–222, 1982. ISSN 0273-0979.

J. Itoh and M. Tanaka. The Lipschitz continuity of the distance function to the cut locus. *Trans. A.M.S.*, 353(1):21–40, 2000.

H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30:509–541, 1977.

M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1:321–331, 1987.

S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi. Gradient flows and geometric active contour models. In *Proceedings of the IEEE Int. Conf. on Computer Vision*, pages 810–815, 1995.

Eric Klassen, Anuj Srivastava, Washington Mio, and Shantanu H. Joshi. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:372–383, 2004. ISSN 0162-8828. doi: 10.1109/TPAMI.2004.1262333.

Wilhelm Klingenberg. *Riemannian Geometry.* W. de Gruyter, Berlin, 1982.

Serge Lang. *Fundamentals of differential geometry.* Springer–Verlag, 1999.

M. Leventon, E. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *IEEE Conf. on Comp. Vision and Patt. Recog.*, volume 1, pages 316–323, 2000.

YanYan Li and Louis Nirenberg. The distance function to the boundary, Finsler geometry and the singular set of viscosity solutions of some Hamilton-Jacobi equations. *Comm. Pure Appl. Math.*, LVIII, 2005. (first received as a personal communication in June 2003).

R. Malladi, J. Sethian, and B. Vemuri. Shape modeling with front propagation: a level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (17):158–175, 1995.

A. C. G. Mennucci. On asymmetric distances. 2nd version, preprint, 2004. URL `http://cvgmt.sns.it/papers/and04/`.

Andrea Mennucci, Anthony Yezzi, and Ganesh Sundaramoorthi. Properties of Sobolev Active Contours. *Interf. Free Bound.*, 10:423–445, 2008.

Peter W. Michor and David Mumford. Vanishing geodesic distance on spaces of submanifolds and diffeomorphisms. *Documenta Math.*, 10:217–245, 2005. URL `http://www.univie.ac.at/EMIS/journals/DMJDMV/vol-10/05.pdf`.

Peter W. Michor and David Mumford. Riemannian geometris of space of plane curves. *J. Eur. Math. Soc. (JEMS)*, 8:1–48, 2006.

Peter W. Michor and David Mumford. An overview of the Riemannian metrics on spaces of curves using the Hamiltonian approach. *Applied and Computational Harmonic Analysis*, 23:76–113, 2007. doi: 10.1016/j.acha.2006.07.004. URL `http://www.mat.univie.ac.at/~michor/curves-hamiltonian.pdf`.

Washington Mio and Anuj Srivastava. Elastic-string models for representation and analysis of planar shapes. In *CVPR (2)*, pages 10–15, 2004a.

Washington Mio and Anuj Srivastava. Elastic-string models for representation and analysis of planar shapes. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2004b. URL `http://stat.fsu.edu/~anuj/pdf/papers/CVPR_Paper_04.pdf`.

D. Mumford and J. Shah. Boundary detection by minimizing functionals. In *Proc. IEEE Conf. Computer Vision Pattern Recognition*, 1985.

D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42: 577–685, 1989.

S. Osher and J. Sethian. Fronts propagating with curvature-dependent speed: algorithms based on the Hamilton-Jacobi equations. *J. Comp. Physics*, 79: 12–49, 1988.

T. R. Raviv, N. Kiryati, and N. Sochen. Unlevel-set: Geometry and prior-based segmentation. In *Proc. European Conf. on Computer Vision*, 2004.

Remi Ronfard. Region based strategies for active contour models. *International Journal of Computer Vision*, 13(2):229–251, October 1994. URL `http://perception.inrialpes.fr/Publications/1994/Ron94`.

M. Rousson and N. Paragios. Shape priors for level set representations. In *Proc. European Conf. Computer Vision*, volume 2, pages 78–93, 2002.

Walter Rudin. *Functional analysis.* McGraw–Hill, 1973.

Walter Rudin. *Real and complex analysis.* McGraw–Hill, 1987.

J. Shah. H^0 type Riemannian metrics on the space of planar curves. *Quarterly of Applied Mathematics*, 66:123–137, 2008.

Stefano Soatto and Anthony J. Yezzi. DEFORMOTION: Deforming motion, shape average and the joint registration and segmentation of images. In *ECCV (3)*, pages 32–57, 2002.

Anuj Srivastava, Shantanu H. Joshi, Washington Mio, and Xiuwen Liu. Statistical shape analysis: Clustering, learning, and testing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:590–602, 2005. ISSN 0162-8828. doi: 10.1109/TPAMI.2005.86.

Ganesh Sundaramoorthi, Anthony Yezzi, and Andrea Mennucci. Sobolev active contours. In Nikos Paragios, Olivier D. Faugeras, Tony Chan, and Christoph Schnörr, editors, *VLSM*, volume 3752 of *Lecture Notes in Computer Science*, pages 109–120. Springer, 2005. ISBN 3-540-29348-5. doi: 10.1007/11567646_10.

Ganesh Sundaramoorthi, Jeremy D. Jackson, Anthony Yezzi, and Andrea Mennucci. Tracking with Sobolev active contours. In *Conference on Computer Vision and Pattern Recognition (CVPR06)*. IEEE Computer Society, 2006. ISBN 0-7695-2372-2. doi: 10.1109/CVPR.2006.314.

Ganesh Sundaramoorthi, Anthony Yezzi, and Andrea Mennucci. Sobolev active contours. *Intn. Journ. Computer Vision*, 2007a. doi: 10.1007/s11263-006-0635-2.

Ganesh Sundaramoorthi, Anthony Yezzi, Andrea Mennucci, and Guillermo Sapiro. New possibilities with Sobolev active contours. In *Scale Space Variational Methods 07*, 2007b. URL `http://ssvm07.ciram.unibo.it/ssvm07_public/index.html`. "Best Numerical Paper-Project Award" ; also Sundaramoorthi et al. [2008b].

Ganesh Sundaramoorthi, Anthony Yezzi, and Andrea Mennucci. Coarse-to-fine segmentation and tracking using Sobolev Active Contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2008a. doi: 10.1109/TPAMI.2007.70751.

Ganesh Sundaramoorthi, Anthony Yezzi, Andrea Mennucci, and Guillermo Sapiro. New possibilities with Sobolev active contours. *Intn. Journ. Computer Vision*, 2008b. doi: 10.1007/s11263-008-0133-9.

Alain Trouvé and Laurent Younes. Local geometry of deformable templates. *SIAM J. Math. Anal.*, 37(1):17–59 (electronic), 2005. ISSN 0036-1410.

A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, E. Grimson, and A. Willsky. Model-based curve evolution technique for image segmentation. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 463–468, Dec. 2001a. doi: 10.1109/CVPR.2001.990511.

A. Tsai, A. Yezzi, and A. S. Willsky. Curve evolution implementation of the mumford-shah functional for image segmentation, denoising, interpolation, and magnification. *IEEE Transactions on Image Processing*, 10(8):1169–1186, Aug 2001b.

L. A. Vese and T. F. Chan. A multiphase level set framework for image segmentation using the mumford and shah model. *Int. J. Computer Vision*, 50 (3):271–293, 2002.

A. Yezzi, A. Tsai, and A. Willsky. A statistical approach to snakes for bimodal and trimodal imagery. In *Int. Conf. on Computer Vision*, pages 898–903, October 1999.

Anthony Yezzi and Andrea Mennucci. Geodesic homotopies. In *EUSIPCO04*, 2004a. URL `http://www.eurasip.org/content/Eusipco/2004/defevent/papers/cr1925.pdf`.

Anthony Yezzi and Andrea Mennucci. Metrics in the space of curves. *arXiv*, 2004b.

Anthony Yezzi and Andrea Mennucci. Conformal metrics and true "gradient flows" for curves. In *International Conference on Computer Vision (ICCV05)*, pages 913–919, 2005. doi: 10.1109/ICCV.2005.60. URL `http://research.microsoft.com/iccv2005/`.

Laurent Younes. Computable elastic distances between shapes. *SIAM Journal of Applied Mathematics*, 58(2):565–586, 1998. doi: 10.1137/S0036139995287685.

Laurent Younes, Peter W. Michor, Jayant Shah, and David Mumford. A metric on shape space with explicit geodesics. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.*, 19(1):25–57, 2008. ISSN 1120-6330. doi: 10.4171/RLM/506.

Charles T. Zahn and Ralph Z. Roskies. Fourier descriptors for plane closed curves. *IEEE Trans. Comput.*, 21(3):269–281, 1972. ISSN 0018-9340. doi: 10.1109/TC.1972.5008949.

Song Chun Zhu, Tai Sing Lee, and Alan L. Yuille. Region competition: Unifying snakes, region growing, energy/bayes/MDL for multi-band image segmentation. In *ICCV*, pages 416–, 1995. URL `citeseer.ist.psu.edu/zhu95region.html`.