Coarse-to-Fine Segmentation and Tracking Using Sobolev Active Contours *

Ganesh Sundaramoorthi*, Anthony Yezzi, Senior Member, IEEE, Andrea C. Mennucci †

May 12, 2007

Abstract

Recently proposed Sobolev active contours introduced a new paradigm for minimizing energies defined on curves by changing the traditional cost of perturbing a curve and thereby redefining their gradients. Sobolev active contours evolve more globally and are less attracted to certain intermediate local minima than traditional active contours, and it is based on a well-structured Riemannian metric, which is important for shape analysis and shape priors. In this paper, we analyze Sobolev active contours using scale-space analysis in order to understand their evolution across different scales. This analysis shows an extremely important and useful behavior of Sobolev contours, namely, that they move successively from coarse to increasingly finer scale motions in a continuous manner. This property illustrates that one justification for using the Sobolev technique is for applications where coarse-scale deformations are preferred over fine scale deformations. Along with other properties to be discussed, the coarse-to-fine observation reveals that Sobolev active contours are, in particular, ideally suited for tracking algorithms that use active contours. We will also justify our assertion that the Sobolev metric should be used over the traditional metric for active contours in tracking problems by experimentally showing how a variety of active contour based tracking methods can be significantly improved merely by evolving the active contour according to the Sobolev method.

1 Introduction

Tracking objects in video sequences with active contours has been an active research area ever since the introduction of *snakes* in [?] (see [?] for a survey). This is often a two step procedure. The first step is *detection*. Here an initial estimate of the object boundary being tracked in a particular image (video frame) is given, and the goal is to evolve this initial contour towards the object of interest in that particular frame. A wide variety of different energy-based schemes have been proposed to do this, including both edge-based [?, ?, ?, ?] and region-based [?, ?, ?, ?, ?, ?] active contours. The second step is *prediction*, where the objective is to predict the object's boundary in the upcoming image based on the presently detected contour as well as contours detected in previous images. Measured (or assumed) dynamics are then extrapolated forward to predict the upcoming contour. Many times, the result from prediction is then averaged in an appropriate manner with the result

*Note that a preliminary version of this manuscript has appeared as a conference proceeding [?].

†G. Sundaramoorthi and A. Yezzi are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, 777 Atlantic Drive, Atlanta, GA 30332 USA (e-mail: {ganeshs, ayezzi}@ece.gatech.edu, phone: 404-894-7515, fax: 404-894-4641). A. C. Mennucci is with Scuola Normale Superiore, Pisa, Italy (a.mennucci@sns.it).

of detection on the prediction to form an *estimate* of the contour (for example see [?, ?]). A trivial approach, which we call the *naive tracker*, assumes no change in dynamics and therefore uses the contour detected in the current frame as the prediction (initial contour) for the next frame. More sophisticated prediction steps may be found in [?, ?, ?, ?] for parametric snakes and more recently [?, ?, ?] for geometric active contours. In [?], a dynamical model of the object being tracked is used, and [?] uses particle filtering for geometric active contours building on the approach in [?].

The prediction step in many contour tracking algorithms is needed because the detection step is too sensitive to initial contour placement, thereby rendering the *naive tracker* inadequate. Indeed, if we had a robust detection scheme that could operate in real-time, then the prediction step could be eliminated and the *naive tracker* would suffice. This sensitivity of active contour models comes in part due to a lack of inherent smoothness in the way the active contours evolve or deform.

Typically an object being tracked deforms rather smoothly from frame to frame, otherwise a prediction would make no sense. Note that we are referring to smoothness of the contour *deformation* (along the contour), not the contour itself. Active contour energies, through the use of regularizers, may easily be adapted to favor smoothness in the final detected contour. However, in tracking it makes sense to ensure smoothness of the deformation of the contour from one frame to the next, regardless of how smooth we want the contour to be. Most current and previous active contour algorithms allow an initial contour to deform in very complex ways, as it flows toward an energy minimum. Even if the final contour has the exact same shape as the initial contour up to translation, the intermediate contours attained during the evolution may vary immensely from the initial and final shapes. This non-preferential *freedom* of the contour to undergo arbitrarily complicated deformations as it flows can attract the contour to undesirable, intermediate local minima before it reaches the desired object boundary.

It would thus be beneficial, when tracking with active contours, to evolve the initial contour, whether or not it was obtained by the naive tracker or by a prediction step, toward its final configuration in a manner that mimics the evolution behavior of objects we wish to track. In particular, it would be ideal if the evolution first favored rigid motions that did not change the actual shape of the evolving contour and then gave preferential treatment to coarser or more global deformations, resorting only at the end to finer or more local deformations when necessary.

Recently, *Sobolev Active Contours* [?, ?] introduced a new paradigm for minimizing energies defined on curves (see also [?, ?]). This yields a completely new way to evolve active contours by exploiting the fact that the gradient flow used to evolve a contour is not only influenced by the energy it minimizes but also by how we measure the cost of perturbing the curve. The works [?, ?] revealed many undesirable properties associated with the usual cost (H^0) inherent in all previous geometric active contour models. Accordingly, [?] and [?] considered using other norms for perturbing active contours based on Sobolev spaces. Sobolev active contours evolve more globally and are less attracted to certain intermediate local minima than traditional active contours. In contrast to the usual strategy of substituting simple energies with more complex (and costly) energies exhibiting fewer local minima, Sobolev active contours minimize the same energy, but follow an entirely different deformation to reach their steady state configuration, thereby avoiding many local minima that would otherwise have been encountered along the way.

Applying Sobolev norms to variational problems has been done in areas other than active contours to gain many of the same advantages that are gained in active contour problems. For example, the book [?] (see also references within) presents the theory of Sobolev gradients and applies it to various physical problems (an interesting physical application is [?]). The key difference between those metrics and the ones we are considering is that `we are defining` *`geometrized`* `Sobolev norms; these are to be though of as the metrics for a (yet to be completely studied) Riemannian manifold of curves, equivalent up to reparametrization -- whereas the usual Sobolev norms are associated to Hilbert (or Banack) vector spaces of` *`parametrized`* `curves`. Recent work that uses the Sobolev gradient for boundary reconstruction is [?]. Geometrized Sobolev metrics have been used for defining shape spaces for *shape analysis*, where the objective is to be able to perform statistical operations on *shapes*, which may be contours. For example, [?, ?, ?] consider Sobolev metrics on the space of plane contours for shape analysis, and other variants are considered by [?].

Although Sobolev active contours was introduced earlier and tracking is mentioned in both [?, ?], our contribution is to expand on the ideas and give a detailed new analysis, which among other things explains the compelling reasons for using the Sobolev metric in tracking situations. Indeed, we examine Sobolev active contours using a scale-space type analysis which shows, along with other properties to be discussed, that these active contours are quite naturally suited for tracking problems, performing (given the exact same energy functional) significantly better than the corresponding traditional active contour. This makes the generic tracking algorithm less dependent on its prediction step as the initial contour does not need to be placed within as narrow an attraction basin in order to reach the desired minimum. In fact, we will see Sobolev active contours often allows even the *naive tracker* to perform well with simple energies that are otherwise plagued by undesirable local minima problems. For a more detailed discussion and analysis of the benefits of Sobolev active contours for tracking, see Section 4.

1.1 Related Works

We now discuss previous techniques that have been explicitly designed to obtain multiscale and global motion properties that are naturally inherent in Sobolev active contours, discuss the advantages of Sobolev active contours in relation to these techniques.

Many active contour works have explicitly incorporated information from successive *scales* of an image to perform a systematic segmentation that matches image data at both coarse and fine scales. For example, in [?] the image is down-sampled to a coarse scale, and an active contour [?] is evolved until convergence. The resulting active contour is up-sampled to a finer scale of the image, and the process is continued on successively finer scale representations of the image until the active contour is evolved in the image itself. The method of [?] makes it less likely that the active contour becomes stuck in irrelevant local minimum of the underlying energy caused by fine scale features of the image. There is also a computational advantage of these methods since the algorithm works with down-sampled images.

- One problem with this method is knowing which and how many scales of the image to use. Ideally, one would like to use a `continuum` of scales that have gradually more information added at each successive scale. However, this is practically not possible in the framework of [?].
- When an arbitrary discrete sampling of scales is chosen, there is a greater possibility of the evolving contour becoming trapped in artificial local minima.
- Choosing a large number of scales reduces the chance of being trapped in a local minimum; however, the computational cost increases.

- One limitation of multiscale techniques like [?] is that there is a limit to how much the domain of the image may be coarsened whenever complicated geometrical objects are present in the image, and it is not trivial to obtain the limit. To illustrate this point, consider a simple example of a binary image consisting of two large circles connected by a long thin strip (an elongated version of the object in Fig. 5). If coarsened or down-sampled enough, the object will become two circles separated by a space, a completely different topology than the actual object. Imagine that an active contour is initialized to enclose the two circles, and the geodesic active contour energy [?, ?] (see also (12)) is used. The contour will split and become two circles. When the contour is evolved in a finer scale of the image, which shows the strip, the contour will not capture the strip because there is a more favorable local minimum of the energy, which is the curve consisting of two circles that slightly protrude into the strip, but do not capture it. In general, this problem is exhibited whenever thin structures are present in the image. This example serves to illustrate that pathologies arising from topological inconsistency of the object to be segmented *across scale* can result from multiscale methods like [?] if careful attention is not paid to the amount of coarsification performed. We note that Sobolev active contours has no such limitation in that the object to be segmented stays a consistent topology across each instant of time during the contour evolution (even when complicated geometrical objects and even thin structures are present), and still performs a coarse-to-fine segmentation.

A solution to the problem of selecting scales is proposed in [?], which considers segmenting all scales of a scale space of an image simultaneously. Although this method has some advantages, the main drawback is a significant increase in computational time because a surface rather than a curve evolution is needed.

As shown in [?] (see also Section 2), Sobolev active contours are *global* flows in that they incorporate image and curve information from the entire curve in order to evolve a single point on the curve. Typical multiscale methods that are used for curve evolutions also use global information in order to deform the contour, but the notion of "global" is different than for Sobolev active contours. Indeed in the coarse scale evolution of a multiscale algorithm, a point $c(s)$ of the curve uses image information from the neighborhood $\{x \in \mathbb{R}^2 : \|x - c(s)\| < R\}$ where $R > 0$ depends on the amount of smoothing/down-sampling performed. While this gives some advantages over traditional active contours including that the evolution incorporates coarse scale information before finer details, the evolution of the curve is not coarse-to-fine and there is no preference to coarse scale deformations before finer deformations because the traditional metric is used. Our scale-space type analysis shows that Sobolev active contours has inherent multiscale behavior. Indeed, the Sobolev technique incorporates information from *all* scales of the image in a systematic fashion first incorporating coarse scale information and favoring coarse deformations of the curve, before gradually moving to finer deformations when coarse deformations can no longer minimize the energy of interest.

Numerous approaches have been applied to tracking that deal with the problem of traditional active contours being attracted to irrelevant local features of an image resulting from the flexibility of the contour to undergo arbitrary deformations. The approach taken by some to avoid arbitrary (and unlikely) deformations in tracking applications has been to restrict the degrees of freedom of the active contour so that more global deformations of the contour are only possible. For example, this approach has been taken by [?, ?, ?]. In [?], a deforming contour is represented by b-splines, and [?, ?] uses polygons. Since there are fewer control points for a b-spline and polygon than a typical parametric snake, this results in more global deformations of the contour. An advantage of [?, ?] over traditional active contours is that information is *integrated* over adjacent edges of the polygon in order to move the corresponding vertex, and hence this adds robustness of the polygon to noise and other local features. Other related methods use of a *finite* number of Fourier or wavelet coefficients to represent the evolving contour (for example, see [?]).

- An advantage of these methods that restrict the contour to more global deformations, besides being less attracted to local features, is that they are computationally fast compared to both parametric and level set implementations [?]. This is because they have inherent regularization and typically do not require an additional regularization term (such as a curvature term), which implies that these methods may take much larger time steps than traditional methods.
- One disadvantage is that topological changes become hard to handle (see [?]).
- Another disadvantage of these approaches is that the motion of the contour is restricted, and therefore, it becomes impossible to detect fine features of the image when they are needed.
- Moreover, these methods do not generally evolve in a coarse-to-fine fashion, which is an advantage of the Sobolev active contours.

Techniques have been designed to force a coarse-to-fine evolution of active contours in order to avoid undesirable local minima of energies in tracking applications. For example, in [?, ?], the authors propose to optimize energies that are defined on both the set of curves and on a set of *global* group motions. In the simplest case, these energies are defined as $E_{\text{new}}(c, g) := E(g \circ c)$ where E is an active contour energy, c is a curve and g is a global group action (e.g. affine motion). For tracking, it is beneficial to optimize with respect to g first while keeping c fixed since the global motion is the most important, and then to optimize (using the traditional metric) with respect to c to obtain fine scale changes of the object being tracked. More recently, [?, ?] (see a related idea in [?]) have proposed to optimize energies using "spatially coherent" flows to achieve the effect of [?, ?] by constructing inner products on the space of perturbations of a curve that favor various group motions such as affine motions. This approach is useful for tracking if one has a prior assumption that the object of interest is moving according to an affine motion. The advantage of using Sobolev active contours for tracking is the inherent coarse-to-fine behavior, which the inner products based on group motions do not have, and the fact that explicit groups do not need to be chosen. In many tracking situations (for example see the real sequences in Section 5), it is not necessarily the case that the object of interest is moving according to an affine motion nor close to it. However, one can generally say that the object is moving according to "coarse" deformations that cannot be captured with simple groups motions (such as affine). The Sobolev active contour is ideal for such a situation where one does not a explicit representation of the motion because it favors general coarse-scale deformations before finer deformations.

2 Review of Sobolev Active Contours

Sobolev active contours were introduced in [?, ?] (they were concurrently introduced in [?, ?]). We give a brief review of the theory. Let M denote the set of immersed curves in $\mathbb{R}^2$, which is a differentiable manifold [?]. For a curve $c \in M$, we denote by T_cM the tangent space of M at c, which is isomorphic to the set of smooth perturbations of the form $h : S^1 \to \mathbb{R}^2$ where S^1 denotes the circle. We also denote by $E : M \to \mathbb{R}$ an energy functional on M, which is known.

Definition 2.1 *Let $E : M \to \mathbb{R}$.*

If $c \in M$ and $h \in T_cM$, then the **variation of** *E is $\mathrm{d}E(c) \cdot h = \frac{\mathrm{d}}{\mathrm{d}t}E(c + th)\big|_{t=0}$, where $(c + th)(\theta) := c(\theta) + th(\theta)$ and $\theta \in S^1$.*

Assume $\langle\cdot,\cdot\rangle_c$ is an inner product on T_cM. The **gradient of** *E is a vector field $\nabla E(c) \in T_cM$ that satisfies $\mathrm{d}E(c) \cdot h = \langle h, \nabla E(c)\rangle_c$ for all $h \in T_cM$.*

One can interpret the gradient as the most efficient perturbation; that is, the gradient maximizes the change in energy per cost of perturbing the curve. The following proposition justifies the previous statement, and was stated in [**?**].

Proposition 2.1 *Let $\|\cdot\|_c$ be the norm induced from the inner product $\langle\cdot,\cdot\rangle_c$ on T_cM. Suppose $\mathrm{d}E(c) \neq 0$, and $\nabla E(c)$ exists; then the problem*

$$\sup_{\{h \in T_cM, \|h\|_c = 1\}} \mathrm{d}E(c) \cdot h = \sup_{\{k \in T_cM, k \neq 0\}} \frac{\mathrm{d}E(c) \cdot k}{\|k\|_c}$$

has a unique solution `up to a multiplicative constant`, *$k = \nabla E(c) \in T_cM, h = k/\|k\|$.*

All previous geometric active contour models that have been formulated as gradient flows of various energies use the same L^2-type inner product (aka H^0) to define the notion of gradient. We review the new inner products on T_cM proposed in [**?**, **?**], which are based on inner products in Sobolev spaces.

Definition 2.2 *Let $c \in M$, L be the length of c, and $h, k \in T_cM$. Let $\lambda > 0$. We assume h and k are parameterized by the arc-length parameter of c.*

1. $\langle h, k\rangle_{H^0} := \frac{1}{L}\int_c h(s) \cdot k(s)\,\mathrm{d}s$
2. $\langle h, k\rangle_{H^n} := \langle h, k\rangle_{H^0} + \lambda L^{2n}\left\langle h^{(n)}, k^{(n)}\right\rangle_{H^0}$
3. $\langle h, k\rangle_{\tilde{H}^n} := avg(h) \cdot avg(k) + \lambda L^{2n}\left\langle h^{(n)}, k^{(n)}\right\rangle_{H^0}$

where $avg(h) := \frac{1}{L}\int_c h(s)\,\mathrm{d}s$, and $h^{(n)}$ is the n^{th} derivative of h with respect to arc-length.

Note that the length dependent scale factors give the above inner products and corresponding norms invariance under rescaling of the curve (*e.g.*, when the domain of the curve, for example the image domain, is scaled). Also, it should be noted that the general definition of the Sobolev inner product of order n contains the H^0 inner product of all lower than order n derivatives; however, as shown in [**?**, **?**] all these definitions are topologically equivalent to the definitions we present, and moreover, the qualitative behavior (and in particular for experiments) depends on the leading derivative. As noted in [**?**, **?**] (see also Section 5), for implementation purposes there is an algorithm for the gradient flow that is independent of the parameter λ.

Sobolev gradients are related to H^0 gradients by linear ODEs, and the solution to these ODEs results in the following expressions:

$$\nabla_{H^n} E = \nabla_{H^0} E * K_{\lambda,n}, \text{ and } \nabla_{\tilde{H}^n} E = \nabla_{H^0} E * \tilde{K}_{\lambda,n} \tag{1}$$

where $*$ denotes convolution on S^1, and for $n = 1$ we have

$$K_{\lambda,1}(s) = \frac{\cosh\left(\frac{s-\frac{L}{2}}{\sqrt{\lambda}L}\right)}{2L\sqrt{\lambda}\sinh\left(\frac{1}{2\sqrt{\lambda}}\right)}, \text{ and } \tilde{K}_{\lambda,1}(s) = \frac{1}{L}\left(1 + \frac{(s/L)^2 - (s/L) + 1/6}{2\lambda}\right), \quad \text{for } s \in [0, L]. \quad (2)$$

It was noted that gradient flows from H^n and $\tilde{H}^n$ have the same qualitative properties, and that they have similar geometric properties. The advantage of using the $\tilde{H}^n$ gradient is that the convolution formula need not be used; $\nabla_{\tilde{H}^n} E$ at all points of the curve can be solved from $\nabla_{H^0} E$ by computing a couple of integrals around the contour. This means that the computational costs of computing the H^0 and $\tilde{H}^n$ gradients are nearly the same; indeed, computing both gradients have the same order of computational complexity. If m is the number of sample points of the curve and we assume that computing $\nabla_{H^0} E$ is $\mathcal{O}(m)$ (which is typical), then to compute $\nabla_{\tilde{H}^n} E$ is $\mathcal{O}(m + nm) = \mathcal{O}(m)$, where $\mathcal{O}$ denotes Big-Oh notation. On the other hand, to compute $\nabla_{\tilde{H}^n} E$ is $\mathcal{O}(m^2)$, which is much more expensive.

2.1 Motivation for Sobolev Metrics

One of the main motivations of our work on Sobolev metrics in the space of curves is to obtain a *consistent* way of doing both shape analysis (*e.g.*, computing statistics of shapes) and shape optimization (*e.g.*, optimizing active contour energies) in a *consistent* manner, and to date this is the only work that offers such a consistent theory. Indeed, it has recently been proven that the Riemannian metric arising from the Sobolev metrics we consider in the space of curves yield well-defined, that is non-degenerate, distances, and that the metrics are complete with respect to Lipschitz curves [**?**]. Also recently, [**?**] computed the geodesic equation for the H^n metric, and proved the existence of geodesics `for small times and smooth initial data`. In the past, there have been many metrics proposed for doing shape analysis (e.g. [**?**, **?**, **?**, **?**]), but the optimization procedure for energies defined on the space of shapes is completely inconsistent with this metric and the geometry of the space, and therefore completely artificial.

There are many advantages of the Riemannian metric approach for shape analysis, for example being able to define a principal component analysis of a set of shapes: we may perform a PCA in the tangent space to the mean shape on the vectors that point in the geodesic direction of each individual shape. A consistent approach for optimization on the space of curves is desired for many reasons. For example, consider the simple application of incorporating prior information into segmentation using active contours (e.g. see [**?**, **?**, **?**, **?**, **?**]). In the simplest case, we may have prior information that the object to be detected from an image is close in the sense of our metric to the shape c_0. Our active contour energy may be the following:

$$E(c) = E_{\text{image}}(c) + d(c, c_0)$$

where c is a shape (e.g. curve), d is the Riemannian metric (shape metric) on the space of curves, and E_{image} is an image-based term. To minimize the energy, one can consider calculating the gradient descent flow, which depends on the metric used to define gradient. If we choose consistent Riemannian metrics for d and the gradient, then the gradient of $d(\cdot, c_0)$ becomes simply the vector pointing in the direction of the geodesic, which is quite natural. On the other hand, if we make an inconsistent choice, then the gradient of $d(\cdot, c_0)$ is the direction which maximizes the change in $d(c, c_0)$ while also minimizing the (inconsistent) cost of perturbing c, which is quite artificial. Moreover, when minimizing E with respect to a different metric than d, we end up far away (in the sense of d) from the initial curve as we step in the gradient direction, which may have detrimental effects for tracking applications where the initial curve is usually quite close to the target curve. Even when computing the average of two shapes, which is a typical computation for shape analysis, a typical procedure is to use a gradient descent to find the average. While using the same metric to define the gradient leads to a gradient descent that is intimately tied to geodesics and therefore to the geometry of the shape space, an inconsistent metric has no such natural connection to the geometry of the space.

In [**?**, **?**], we have demonstrated many important and useful properties of using Sobolev metrics for defining gradient descents of active contour energies. For example, it was shown that Sobolev flows are smooth in the space of curves, are not as dependent on local image information as H^0 flows, are less sensitive to certain local minima, are global flows, and reduce the order of the evolution PDE in comparison to H^0. This article gives more justification to use Sobolev active contours by showing important scale-space properties that are automatically inherent in Sobolev gradient flows. This property makes the method quite useful for tracking. Although ad-hoc methods (i.e., Gaussian smoothing of the H^0 gradient as considered by [**?**]) may be applied to gain smoothness properties like Sobolev active contours, it does not come from a Riemannian metric, which is very important for the reasons discussed above.

3 Scale-Space Type Analysis of Sobolev Active Contours

In this section, through Fourier analysis, we show that Sobolev active contours favor more rigid motions than regular active contours, and that they first undergo coarse scale motions before resorting to fine scale deformations in optimizing the chosen energy.

3.1 Sobolev Norms in Frequency Domain

Notice that since any $h \in T_c M$ is smooth on S^1, it follows that $h \in L^2(S^1)$. Thus, we may write h as a Fourier series, i.e.,

$$h(s) = \sum_{l \in \mathbb{Z}} \hat{h}(l) \exp\left(\frac{2\pi i}{L} l s\right) \tag{3}$$

with convergence in $L^2(S^1)$ (and in fact point wise since h is smooth) where $\hat{h} \in \ell^2(\mathbb{Z})$ is defined by

$$\hat{h}(l) = \frac{1}{L} \int_c h(s) \exp\left(-\frac{2\pi i}{L} l s\right) \mathrm{d}s. \tag{4}$$

It should be noted that (3) decomposes the perturbation into the orthonormal basis of exponentials. This allows us to write Definition 2.2 in the frequency domain. By Parseval's theorem,

$$\int_0^L h(s) \cdot k(s) \, \mathrm{d}s = L \sum_{l \in \mathbb{Z}} \hat{h}(l) \cdot \overline{\hat{k}(l)}.$$

where $\bar{\cdot}$ denotes complex conjugation. We also have that

$$\int_0^L h^{(n)}(s) \cdot k^{(n)}(s) \, \mathrm{d}s = L \sum_{l \in \mathbb{Z}} \left(\frac{2\pi}{L} l\right)^{2n} \hat{h}(l) \cdot \overline{\hat{k}(l)};$$

therefore,

Proposition 3.1 *If $h, k \in T_cM$, L is the length of c, and $\widehat{h}, \widehat{k} : \mathbb{Z} \to \mathbb{C}$ are defined by (4). Then,*

$$\langle h, k \rangle_{H^n} = \sum_{l \in \mathbb{Z}} (1 + \lambda(2\pi l)^{2n}) \widehat{h}(l) \cdot \overline{\widehat{k}(l)} \tag{5}$$

$$\langle h, k \rangle_{\tilde{H}^n} = \widehat{h}(0) \cdot \overline{\widehat{k}(0)} + \sum_{l \in \mathbb{Z}} \lambda(2\pi l)^{2n} \widehat{h}(l) \cdot \overline{\widehat{k}(l)}. \tag{6}$$

and the corresponding norms are

$$\|h\|^2_{H^n} = \sum_{l \in \mathbb{Z}} (1 + \lambda(2\pi l)^{2n}) |\widehat{h}(l)|^2 \tag{7}$$

$$\|h\|^2_{\tilde{H}^n} = |\widehat{h}(0)|^2 + \sum_{l \in \mathbb{Z}} \lambda(2\pi l)^{2n} |\widehat{h}(l)|^2. \tag{8}$$

Notice that Proposition 3.1 allows us to define the H^n and $\tilde{H}^n$ inner products for n that is any real number greater than 0. These inner products are defined the same way as in (5) and (6). It is easy to verify in this case too that the definitions are indeed inner products. Unfortunately for n that is not an integer, the inner products (therefore, norms) are not local, that is, they cannot be written as integrals of derivatives of the curves; but, given $r \in \mathbb{R}^+$ we can represent them, for n integer $n > r + 1/4$ as $\langle h, k \rangle_{\tilde{H}^r} = \operatorname{avg}(h) \cdot \operatorname{avg}(k) + L^{2n} \left\langle h^{(n)}, K * k^{(n)} \right\rangle_{H^0}$, for a kernel K with Fourier coefficients

$$\widehat{K}(l) = \begin{cases} 1 & \text{if } l = 0 \\ \lambda(2\pi l)^{2r-2n} & \text{if } l \neq 0. \end{cases}$$

The norms shown in (7) and (8) measure the perturbation magnitude in terms of its Fourier coefficients, which are the weights of its corresponding frequency components. We see that for both H^n and $\tilde{H}^n$ norms, high frequency components of the perturbation contribute increasingly to the norm of the perturbation.

3.2 Sobolev Gradients in Frequency Domain

We now calculate Sobolev gradients of an arbitrary energy in the frequency domain. By Definition 2.1, if the H^0 and H^n gradients of an energy $E : M \to \mathbb{R}$ exist, then it follows that

$$\mathrm{d}E(c) \cdot h = \langle \nabla_{H^0} E(c), h \rangle_{H^0} = \langle \nabla_{H^n} E(c), h \rangle_{H^n}$$

for all $h \in T_cM$. Using Parseval's Theorem, the last expression becomes

$$\sum_{l \in \mathbb{Z}} (1 + \lambda(2\pi l)^{2n}) \widehat{\nabla_{H^n} E}(l) \cdot \widehat{h}(l) = \sum_{l \in \mathbb{Z}} \widehat{\nabla_{H^0} E}(l) \cdot \widehat{h}(l).$$

Since the last expression holds for all $h \in T_cM$, we have

$$\widehat{\nabla_{H^n} E}(l) = (1 + \lambda(2\pi l)^{2n})^{-1} \widehat{\nabla_{H^0} E}(l) \quad \text{for } l \in \mathbb{Z} \tag{9}$$

Using a similar argument, we see that

$$\widehat{\nabla_{\tilde{H}^n} E}(l) = \begin{cases} \widehat{\nabla_{H^0} E}(0) & l = 0 \\ (\lambda(2\pi l)^{2n})^{-1} \widehat{\nabla_{H^0} E}(l) & l \in \mathbb{Z} \backslash \{0\}. \end{cases}$$

Figure 1: Increasingly higher frequency perturbations applied to a circle (left to right, $l = 0, 2, 5, 10$).

It is clear from the previous expressions that high frequency components of $\nabla_{H^0} E$ are less pronounced in the various forms of Sobolev gradients when compared with the H^0 gradient, with higher order Sobolev gradients damping high frequency components with faster decay rates.

3.3 Coarse-To-Fine Motion of Sobolev Contours

We now discuss the implications of the results in the previous sections. Note that the Fourier basis of the perturbations of a curve decomposes $T_c M$ from global perturbations (low frequency perturbations) to increasingly more finer perturbations (high frequency perturbations). Indeed the zero frequency perturbation is a simple translation of the curve, which is completely global. See Fig. 1. Therefore, by (9), and comments in the previous section, it is apparent that Sobolev gradients yield perturbations with more pronounced global components than the standard H^0 gradient. While H^0 gradients give equal weighting across all scales, Sobolev gradients give less weight to finer scales. However, this does not mean that very fine scale deformations of the curve are restricted in Sobolev gradient flows. It just means that if there exists a low order perturbation (a more global motion) that increases the given energy just as would a higher order perturbation (a more finer deformation), then the low order perturbation will be preferred in the Sobolev gradient, as shown by Proposition 2.1. Also, if no perturbations in G_m, given by

$$G_m = \left\{ \sum_{|l| \leq m} a_l \exp\left(\frac{2\pi i}{L} l \cdot \right) : a_l \in \mathbb{C}, a_{-l} = \overline{a_l} \right\},$$

can increase the energy, E; that is $\mathrm{d}E(c) \cdot h \leq 0$ for all $h \in G_m$, then by Definition 2.1, we must have that $\widehat{\nabla_{H^0} E}(l) = 0$ for $l \leq m$, and therefore, we can write

$$\widehat{\nabla_{\tilde{H}^n} E}(l) = \frac{1}{\lambda (m+1)^{2n}} \begin{cases} 0, & |l| \leq m \\ \frac{1}{(2\pi l/(m+1))^{2n}} \widehat{\nabla_{H^0} E}(l), & |l| > m \end{cases}.$$

We see that since the gradient flow does not geometrically depend on a scale factor, the Sobolev gradient automatically has the weights on high order perturbations of the gradient readjusted (so that perturbations near $|l| = m + 1$ become more pronounced). This means the Sobolev gradient flow at this particular instant of the evolution changes the finer scale structure of the curve. Thus, with Sobolev active contours, this implies at least locally, a progression from coarse scale motion to finer scale motion, unlike the standard H^0 active contour. These comments are illustrated in Fig. 2, which shows the tracking of a noisy square image using both H^0 and H^1 active contours. Notice that with the H^0 active contour, the fine structure of the curve is changed immediately, while the H^1 active contour gradually changes finer scale features of the curve after changing coarse-scale features.

The effect of using higher order (n large) Sobolev gradients is higher favorability to lower order perturbations in the flow.

Figure 2: Standard (H^0) active contour (top) alters fine structure of the curve immediately; Sobolev (H^1) active contour (bottom) moves from coarse to finer scale motions. Both use same energy.

3.4 Analytic Examples of Coarse-To-Fine Motion of Sobolev Flows

In this section, we give two analytic examples of Sobolev gradient flows and show explicitly that as the artificial time parameter of the evolution increases, the gradient generally moves from coarse scale to finer scale perturbations. This will verify that the curve initially deforms in a coarse manner before resorting to finer deformations to decrease the energy of interest. Since geometric energies make use of the arc-length parameterization, which lead to non-linear equations that are difficult to analyze analytically, we perform an analysis on related parametric energies. Indeed, it took decades to analyze the H^0 gradient flow of a simple geometric energy, length, analytically in the mathematical community [**?**, **?**]. However, the parametric analysis we perform is relevant in inferring the qualitative behaviors, particularly the coarse-to-fine properties, of the geometric Sobolev gradient flows as we shall justify throughout this section by illustrating the similarities between the parametric and geometric evolutions in the cases where the geometric behavior is known through analysis from other methods.

The inner products that we use are the parametric equivalent of those given in Definition 2.2; that is, if $h, k : S^1 \to \mathbb{R}^2$ are perturbations of a curve $c : S^1 \to \mathbb{R}^2$ then we define

$$\langle h, k \rangle_{H^0} := \int_{S^1} h(u) \cdot k(u) \, \mathrm{d}u$$

$$\langle h, k \rangle_{\tilde{H}^n} := \mathrm{avg}(h) \cdot \mathrm{avg}(k) + \lambda \left\langle h^{(n)}, k^{(n)} \right\rangle_{H^0}$$

where $\mathrm{avg}(h) := \int_{S^1} h(u) \, \mathrm{d}u$, and $h^{(n)}$ is the n^{th} derivative with respect to the parameter u. Note that we use $\tilde{H}^n$ instead of H^n for simplicity, but similar conclusions hold for the H^n inner product

The first energy we consider is the following curve matching energy, $E_m : C^\infty(S^1, \mathbb{R}^2) \to \mathbb{R}^+$:

$$E_m(c) = \frac{1}{2} \int_{S^1} |c(u) - c_0(u)|^2 \, \mathrm{d}u \tag{10}$$

where $c_0 : S^1 \to \mathbb{R}^2$ is a pre-specified target curve, which is the known *data*. This energy is a representative of data-based energies, e.g., possibly energies that depend on image data such as the image-based term of the Chan and Vese energy [**?**].

Remark 3.1 *A geometric version of the energy* (10) *might be the following:*

$$E(c) = \inf_{b \in [0, L_0]} \frac{1}{2} \int_c |c(s) - c_0(L_0/L_c \cdot s + b)|^2 \, \mathrm{d}s$$

where s is the arc-length parameter of c, L_0 is the length of c_0, and L_c is the length of c. An optimization of this energy would require a joint evolution of the parameter b and the curve c. The H^0 gradient of this energy w.r.t c is

$$\nabla_{H^0} E(c)(s) = (c(s) - c_0(L_0/L_c \cdot s + b)) \cdot \mathcal{N}(s) - |c(s) - c_0(L_0/L_c \cdot s + b)|^2 c_{ss}(s).$$

As we will see, the first term is the normal component of the evolution of E_m, and we will show later when we consider the energy E_s below that the second term in the Sobolev domain has the *coarse-to-fine property we seek to show. Thus, our parametric analysis gives us a good idea of the behavior of the geometric energy also.*

Obviously the optimum curve for (10) is $c = c_0$, which is the global minimizer. However, we are interested in *how* the curve evolves in the frequency domain to attain the minimum. To see this, let us note our definition of Fourier transform:

$$\widehat{f}(k) = \int_{S^1} f(u) \exp(-2\pi i k u) \, \mathrm{d}u$$

where $f : S^1 \to \mathbb{R}$. Note that

$$\mathrm{d}E_m(c) \cdot h = \int_{S^1} (c(u) - c_0(u)) \cdot h(u) \, \mathrm{d}u,$$

and so, $\nabla_{H^0} E(c) = c - c_0$. Note that

$$\widehat{\nabla_{\tilde{H}^n} E_m}(l) = \begin{cases} \widehat{\nabla_{H^0} E_m}(0), & l = 0 \\ \frac{\widehat{\nabla_{H^0} E_m}(l)}{\lambda(2\pi l)^{2n}}, & l \in \mathbb{Z}\backslash\{0\}. \end{cases}$$

Let's now consider the $\tilde{H}^n$ gradient flow

$$\partial_t c(u,t) = -\nabla_{\tilde{H}^n} E_m(c);$$

we may write c as a Fourier series:

$$\begin{aligned} c(u,t) &= \sum_{l \in \mathbb{Z}} \widehat{c}(l,t) \exp(2\pi i l u) \\ \nabla_{\tilde{H}^n} E_m(u,t) &= \widehat{\nabla_{H^0} E_m}(0,t) + \sum_{l \in \mathbb{Z}\backslash\{0\}} \frac{\widehat{\nabla_{H^0} E_m}(l,t)}{\lambda(2\pi l)^{2n}} \exp(2\pi i l u) \\ &= \widehat{c}(0,t) - \widehat{c_0}(0) + \sum_{l \in \mathbb{Z}\backslash\{0\}} \frac{\widehat{c}(l,t) - \widehat{c_0}(l)}{\lambda(2\pi l)^{2n}} \exp(2\pi i l u). \end{aligned}$$

Assuming that $\widehat{c}(\cdot,t)$ is uniformly bounded by an $\ell^1(\mathbb{Z})$ function, the flow becomes

$$\sum_{l \in \mathbb{Z}} \partial_t \widehat{c}(l,t) \exp(2\pi i l u) = -(\widehat{c}(0,t) - \widehat{c_0}(0)) - \sum_{l \in \mathbb{Z}\backslash\{0\}} \frac{\widehat{c}(l,t) - \widehat{c_0}(l)}{\lambda(2\pi l)^{2n}} \exp(2\pi i l u).$$

Since $\{\exp(2\pi i l \cdot)\}_{l \in \mathbb{Z}}$ is an orthogonal basis; we have that

$$\begin{aligned} \partial_t \widehat{c}(0,t) &= -(\widehat{c}(0,t) - \widehat{c_0}(0)) \quad \text{and} \\ \partial_t \widehat{c}(l,t) &= -\frac{\widehat{c}(l,t) - \widehat{c_0}(l)}{\lambda(2\pi l)^{2n}} \quad \text{for } l \in \mathbb{Z}\backslash\{0\}. \end{aligned}$$

or

$$\begin{aligned} \partial_t \widehat{c}(0,t) + \widehat{c}(0,t) &= \widehat{c_0}(0) \quad \text{and} \\ \partial_t \widehat{c}(l,t) + \frac{1}{\lambda(2\pi l)^{2n}} \widehat{c}(l,t) &= \frac{\widehat{c_0}(l)}{\lambda(2\pi l)^{2n}} \quad \text{for } l \in \mathbb{Z}\backslash\{0\} \end{aligned}$$

solving the previous equations yield

$$\begin{aligned}\widehat{c}(0,t) &= \exp(-t)\widehat{c}(0,0) + \widehat{c_0}(0)\left[1 - \exp(-t)\right] \\ \widehat{c}(l,t) &= \exp\left(-\frac{t}{\lambda(2\pi l)^{2n}}\right)\widehat{c}(l,0) + \widehat{c_0}(l)\left[1 - \exp\left(-\frac{t}{\lambda(2\pi l)^{2n}}\right)\right].\end{aligned}$$

Therefore,

$$\begin{aligned}\widehat{\nabla_{\tilde{H}^n} E_m}(0,t) &= \widehat{c}(0,t) - \widehat{c_0}(0) = \exp(-t)(\widehat{c}(0,0) - \widehat{c_0}(0)) \\ \widehat{\nabla_{\tilde{H}^n} E_m}(l,t) &= \frac{\widehat{c}(l,t) - \widehat{c_0}(l)}{\lambda(2\pi l)^{2n}} = \frac{1}{\lambda(2\pi l)^{2n}}\exp\left(-\frac{t}{\lambda(2\pi l)^{2n}}\right)(\widehat{c}(l,0) - \widehat{c_0}(l)).\end{aligned}$$

For convenience define,

$$g_n(l,t) := \frac{1}{\lambda(2\pi l)^{2n}}\exp\left(-\frac{t}{\lambda(2\pi l)^{2n}}\right).$$

Note that

- $\lim_{t\to+\infty} g_n(l,t) = 0$ for all $l \in \mathbb{Z}$, and also the rate of convergence to zero as $|l|$ becomes larger is slower (when $n > 0$). Thus, we find that the coarse scale frequency components ($|l|$ small) of c converge to the global minimum much faster than the fine scale frequencies.

- $\lim_{|l|\to+\infty} g_n(l,t) = 0$.

- We want to show that as time increases, the low frequency components of $\widehat{\nabla_{\tilde{H}^n} E_m}$ decrease to zero monotonically for a gradually increasing number of low frequency components. Note that $\|\widehat{\nabla_{\tilde{H}^n} E_m}\|_{\ell^\infty(\mathbb{Z})} \to 0$ as $t \to +\infty$. Because a common scale factor for $\widehat{\nabla_{\tilde{H}^n} E_m}$ does not have any effect on the geometry of the curve evolution, we scale $\widehat{\nabla_{\tilde{H}^n} E_m}$ by its maximum value in l to show the convergence to zero of low frequency components relative to the rest of the components of $\widehat{\nabla_{\tilde{H}^n} E_m}$. Consider

  $$\tilde{g}_n(l,t) = \frac{g_n(l,t)}{1/(et)} = etg_n(l,t).$$

  Note that
  $$\frac{\partial \tilde{g}_n}{\partial t} = \frac{e}{\lambda(2\pi l)^{2n}}\exp\left(-\frac{t}{\lambda(2\pi l)^{2n}}\right)\left(1 - \frac{t}{\lambda(2\pi l)^{2n}}\right) < 0$$

  when $l < \frac{1}{2\pi}(t/\lambda)^{1/(2n)}$. Note also that $\lim_{t\to+\infty} \tilde{g}_n(l,t) = 0$ for all $l \in \mathbb{Z}$. These facts verify our assertion.

- For a fixed time, we approximate the frequency component that is changing the most. To do this, we suppose l is a real number (even though it is an integer) and compute the derivative,

  $$\frac{\partial g_n}{\partial l} = \frac{2}{\lambda(2\pi)^{2n} l^{2n+1}}\exp\left(-\frac{t}{\lambda(2\pi l)^{2n}}\right)\left[-1 + \frac{t}{l^{2n}}\right],$$

  and so $l_* \approx \pm t^{\frac{1}{2n}}$ is the frequency component being changed the most at a particular time instant. In particular, this shows a successive transition through *all* possible frequencies in a coarse-to-fine manner (see the right plot in Fig. 3).

Figure 3: Left: Plot of $\tilde{g}_n(\cdot, t)$ with $n = 2$ for various t, which shows coarse-to-fine behavior. Right: Plot of frequency component changing most rapidly versus time, which shows that the evolution transitions through *all* possible frequencies.

Figure 4: Minimizing E_m using a "spiked circle" (which is shown on the right of each row) for c_0, and the initialized curve is a circle: $c(u, 0) = \epsilon(\cos 2\pi u, \sin 2\pi u), \epsilon = 0.001$ for $u \in S^1$ (left of each row, enlarged for visibility). Top: Snapshots of the H^0 evolution. Bottom: Snapshots of the $\tilde{H}^2$ evolution. The right of each row shows that both flows converge to the desired "spiked" circle, but taking quite different paths.

In the left of Fig. 3, we show a plot of $\tilde{g}_n$ to visually show the movement of $\widehat{\nabla_{\tilde{H}^n} E_m}$ from coarse to finer scale perturbations. The figure shows the case when $n = 2$; however, other $n > 0$ have the same qualitative behavior. Larger n shows a slower and more pronounced progression from coarse to fine. Note that when $n = 0$, $\tilde{g}_n(l, t) = \frac{et}{\lambda} \exp\left(-\frac{t}{\lambda}\right)$, which is constant in l for a fixed time, and so the plot on the left of Fig. 3, would simply be a horizontal line for all time. The right of Fig. 3 shows (when $n = 2$) the frequency for which $\tilde{g}_n(\cdot, t)$ (or $g_n(\cdot, t)$) is maximized versus time. This plot shows that the Sobolev active contour transitions through all possible frequencies.

The evolutions (H^0 and $\tilde{H}^2$) minimizing E_m are shown in Fig. 4. Notice that the H^0-evolution favors all frequency components the same, and so the coarse and fine features of the target curve are both detected at the same times. On the other hand, the Sobolev evolution changes in a coarse manner, detecting the coarse structure of the target curve, before finally detecting the fine "spiked" structure and deforming in a fine-scale manner.

Remark 3.2 *For real images, as long as the image data does not undergo drastic changes with respect to movements of the curve, this coarse-to-fine motion shown in this example also applies to real image-based evolutions (in particular, for many tracking applications). In the general case, we would expect this coarse-to-fine deformation to repeat during successive intervals of time.*

The next energy we consider is the following smoothing energy, $E_s : C^\infty(S^1, \mathbb{R}^2) \to \mathbb{R}^+$:

$$E_s(c) = \frac{1}{2} \int_{S^1} |c^{(m)}(u)|^2 \, \mathrm{d}u, \tag{11}$$

where $m \geq 1$.

Remark 3.3 *The geometric equivalent to* (11) *is*

$$E(c) = \frac{1}{2} \int_c |D_s^m c(s)|^2 \, \mathrm{d}s.$$

*In the case that $m = 1$, the energy is half of arc-length, and in the case $m = 2$, the energy is the elastic energy. The Sobolev flows for these energies were calculated in [**?**, **?**]. As we progress, we will illustrate the similar qualitative behavior between the parametric and geometric energies, which justifies our use of a parametric analysis to infer some qualitative behaviors in the geometric case.*

Note that $\nabla_{H^0} E_s(c) = (-1)^m c^{(2m)}$; therefore,

$$-\widehat{\nabla_{H^0} E_s}(l) = -(-1)^m (2\pi i l)^{2m} \widehat{c} = -(2\pi k)^{2m} \widehat{c}(k)$$

and also,

$$-\widehat{\nabla_{\tilde{H}^n} E_s}(l) = \begin{cases} 0 & \text{for } l = 0 \\ -\lambda^{-1}(2\pi l)^{2(m-n)} \widehat{c}(l) & \text{for } l \neq 0 \end{cases}.$$

We would like to compute the gradients as a function of time in Fourier domain. Note, $\partial_t \widehat{c}(l,t) = -\widehat{\nabla_{\tilde{H}^n} E_s}(l)$, so

$$\partial_t \widehat{c}(l,t) = \begin{cases} 0 & l = 0 \\ -\lambda^{-1}(2\pi l)^{2(m-n)} \widehat{c}(l) & l \neq 0 \end{cases},$$

which yields a solution of

$$\widehat{c}(l,t) = \begin{cases} \widehat{c}(0,0) & l = 0 \\ \exp\left(-\lambda^{-1}(2\pi l)^{2(m-n)} t\right) \widehat{c}(l,0) & l \neq 0 \end{cases}.$$

Hence, we see

$$\widehat{\nabla_{\tilde{H}^n} E_s}(l,t) = \begin{cases} 0 & l = 0 \\ \lambda^{-1}(2\pi l)^{2(m-n)} \exp\left(-\lambda^{-1}(2\pi l)^{2(m-n)} t\right) \widehat{c}(l,0) & l \neq 0 \end{cases}.$$

We make the following observations:

- For $n < m$, the rate of convergence of $\widehat{\nabla_{\tilde{H}^n} E_s}(l,\cdot)$ to 0 (also $\widehat{c}(l,\cdot)$ to 0) increases as $|l|$ increases. Using similar arguments as for E_m, one can see that this means that $\widehat{\nabla_{\tilde{H}^n} E_s}$ moves from fine to coarse scale perturbations. Note that this case includes the case when $m = 1$ and $n = 0$, which is linear heat flow. This flow is well-known to have a smoothing effect on the curve, and removes fine scale curve information before removing coarser scale information (i.e., $\widehat{\nabla_{\tilde{H}^n} E_s}$ moves from fine to coarse). Similarly, in the geometric case when $m = 1$ (i.e., the energy is the length of the curve) and $n = 0$, the evolution is non-linear geometric heat flow [**?**, **?**], which is well-known to have a fine to coarse smoothing effect.

- For $n = m$, the rate of convergence of all frequency components of $\widehat{\nabla_{\tilde{H}^n} E_s}(l,\cdot)$ to 0 is the same for all l. The curve evolution is a simple rescaling of the contour about its parametric centroid. It is also apparent in this case that the curve evolution exists when the gradient *ascent* flow is considered. Similarly, as verified in [**?**, **?**] through direct computations, in the case of the geometric energy when $m = 1$ and $n = 1$, the evolution is a simple rescaling of the curve about the *geometric* centroid, and therefore also stable for the ascent.

- For $n > m$, the rate of convergence of $\widehat{\nabla_{\tilde{H}^n} E_s}(l,\cdot)$ to 0 (also $\widehat{c}(l,\cdot)$ to 0) decreases as $|l|$ increases. As in the case of the energy E_m, we can show that $\widehat{\nabla_{\tilde{H}^n} E_s}$ moves from coarse to finer scale motions.

- From the above statements, we see that the Sobolev gradient flows move in a more coarse to fine way than the H^0 gradient flow, and as the order of the Sobolev gradient increases, this coarse-to-fine motion is more pronounced.

4 Benefits of Sobolev Contours for Tracking

In this section, we outline the benefits of switching from the standard H^0 active contour evolution to a Sobolev active contour in tracking algorithms that use active contours.

The scale-space analysis of Sobolev active contours performed in Section 3 that shows a coarse-to-fine evolution of the contour also shows why Sobolev active contours are ideal for tracking. The fact that H^0 gradient flows change fine structure of the curve immediately when energetically favorable, and hence are easily attracted by undesirable local minima, is one reason for predicting motion and dynamics of the object being tracked. By predicting motion and dynamics of the moving object, a better estimate of the object's upcoming position can be attained thereby placing the initial guess hopefully closer to its desired final position. Many prediction schemes apply low dimensional global motions to the contour. Thus, the initial global motion followed by an H^0 flow is less likely than the naive tracker to get caught in an intermediate, undesirable local minimum of the energy. Notice that since Sobolev gradient flows naturally move from coarse to successively finer motions, the contour is less likely to be trapped by intermediate local minima caused by local features of the image, and is therefore likely to be less dependent on the prediction of motion and dynamics of the object. We also wish to emphasize that the transition from coarse to increasingly finer motions is automatic and continuous in comparison to other works (e.g., [**?**]) where the global motions must be deliberately specified, and the transition from the global motion to more local deformation is not continuous. Indeed, even discrete attempts to deliberately graduate from more global to more local motions are not trivial as one typically starts from translations, then rotations, then scale, but beyond this it becomes less clear and natural how to progress to finer scale deformations.

Another advantage of using Sobolev active contours for tracking is speed of convergence compared to standard H^0 active contours. While computing the $\tilde{H}^n$ gradient is slightly more computationally costly than computing the H^0 gradient, though both have the same order of complexity, we point out that without accurate prediction, the number of iterations in typical contour tracking applications required to update the active contour from frame to frame is usually much smaller with Sobolev active contours. Therefore the total computational time for processing between frames is significantly lower with Sobolev active contours. The reason is that the frame-to-frame motion of the object to be tracked is, as mentioned previously, usually dominated by more global motions: translations, scaling, and coarse scale deformations. Accordingly, a Sobolev active contour needs only a few iterations to lock onto the object in the next frame because the Sobolev gradient moves globally at first, preferring coarse scale motions in the first few iterations before proceeding to fine scale motions in later iterations. In contrast, standard H^0 active contours requires many more iterations since they immediately deform by local motions, significantly changing their initial shape (often to meaningless intermediate shapes), before deforming back to only slightly deformed, translated and scaled versions of their initial shape, and that is assuming they don't first get trapped into intermediate local minima!

We now illustrate the advantages discussed in the previous paragraphs with a simple synthetic image sequence (Fig. 5) in which we employ the naive tracker using the energy functional for geodesic active contours [**?**, **?**]:

$$E_{geo}(c) = \int_c \phi(c(s))\,\mathrm{d}s, \text{ where } \phi = \frac{1}{1 + \|\nabla I\|^2}. \tag{12}$$

Fig. 5 shows the tracking for both the H^0 gradient flow and the $\tilde{H}^1$ gradient flow. The flows

Figure 5: Simple tracking using geodesic active contours: Standard (H^0) active contour (left column) deforms the initialized contour greatly and is stuck in local minima, and Sobolev active contour (right column) moves in a global manner only slightly changing shape. In each frame, the initial curve (given by the contour detected in the previous frame) is blue, the intermediate curve is green, and the final detected curve is red.

Figure 6: Graphs showing number of iterations to converge and total time for convergence versus set symmetric difference (SSD) of initial region and desired object in percent (scaled by desired object area).

are run until convergence in each frame. Note that the H^0 active contour deforms its initial shape greatly to react to local information. Hence the contour changes shape and must re-deform back to its initial shape. However, the contour gets trapped in an undesirable local minimum. The Sobolev active contour, on the other hand, only changes shape slightly while moving in an overall translation. This means that the number of iterations until convergence for the H^0 active contour is much greater than the Sobolev active contour, and therefore the computational time is also much greater. See Fig. 6 for a simple quantitative analysis of the number of iterations and computational times. In this simulation, we segment the object shown in Fig. 5 when the initial contour is a translated and a slightly deformed version of the object. We quantify the difference by using the set symmetric difference between the desired object and the initial contour. From the graph in Fig. 6, we see that the number of iterations and the computational time is significantly lower for the $\tilde{H}^1$ active contour.

5 Experiments

We now demonstrate significant performance gains by replacing standard H^0 active contours with their Sobolev counterparts for the *exact same* detection energy in a variety of tracking scenarios on real videos, both when using the *naive tracker* as well as when tracking with a predictor. These experiments give evidence to support our claim that the Sobolev metric rather than the traditional metric should be used in tracking applications that make use of active contours.

Note that in the next experiments, we use the $\tilde{H}^1$ active contour and the algorithm described in [**?**] that is independent of the parameter λ in the definition of the $\tilde{H}^1$ inner product. The algorithm evolves by the translation component of the $\tilde{H}^1$ gradient until this term becomes zero followed by one iteration of the deformation component, which is geometrically independent of λ, and the process is iterated.

Fig. 7 shows the results for a sequence in which a man is walking on a street. The sequence is heavily corrupted by noise (Gaussian noise: $\mu = 0, \sigma^2 = 0.3$). The tracking is done using the naive tracker (no prediction) where the detection energy is the Chan-Vese energy functional [**?**]:

$$E_{cv}(c) = \int_{c_{in}} (I - u)^2 \,\mathrm{d}A + \int_{c_{out}} (I - v)^2 \,\mathrm{d}A + \alpha L(c), \tag{13}$$

where

$$u = \frac{\int_{c_{in}} I \,\mathrm{d}A}{\int_{c_{in}} \mathrm{d}A}, \text{ and } v = \frac{\int_{c_{out}} I \,\mathrm{d}A}{\int_{c_{out}} \mathrm{d}A},$$

Figure 7: Tracking of a person in a noisy image sequence with a region-based (Chan-Vese) energy with H^0 (top row), with H^0 translation favored (middle row), and with $\tilde{H}^1$ (bottom row) active contours.

and $\alpha \geq 0$ specifies a penalty on the length (used for curve regularity for the H^0 active contour) $L(\cdot)$ of the curve. The top row shows the standard H^0 ($\alpha = 5000$) active contour, and the bottom row shows the Sobolev $\tilde{H}^1$ ($\alpha = 0$) active contour. Sobolev active contours in particular favor translations; therefore in order to show that the translation favoring property of Sobolev active contours is not solely responsible for the pleasing tracking results, but more generally it is the coarse-to-fine property, we also show the results of tracking where the energy minimization is performed using an H^0 inner product that has a heavily weighted translation component in the middle of Fig. 7 (note the advantages of the Sobolev technique over such explicitly favored groups was discussed at the end of Section 1.1); a similar result is obtained for an affine favored H^0 gradient. We have used an alternating algorithm between a translation and the H^0 gradient minus the translation, to avoid picking the weight on the translation, but the result is similar to Fig. 7 (middle row). Note that such an inner product was considered in [**?**]. The contours are evolved until convergence between frames. After a few frames, the H^0 active contour gets stuck in noise and loses track of the person. The translation favored H^0 contour initially does better than the H^0 active contour, but soon loses track of the person, becoming stuck in noise. Note that the translation, in the first frames, initially pushes the contour in the vicinity of the person, but then the active contour immediately detects the fine scale noise since H^0 minus the translation does not favor coarser motions, and cannot more accurately detect the person. The Sobolev active contour, because of its more global initial motions (translations and other coarse motions), avoids the intermediate local minima caused by noise and keeps tracking the person. Due to the high noise level, however, the precise shape of the person is not captured in all of the cases. Note that one may try different weights, α, to obtain better results for the H^0 and the translation favored H^0 active contours; however, we have found that lower values (than $\alpha = 5000$) produce results that are more easily trapped in noise and higher values simply shrink the curve to a point. It should be noted that although a length penalty for the Sobolev $\tilde{H}^1$ is not needed (and indeed does not make the curve more regular [**?**, **?**]), a length penalty of $\alpha = 5000$, does not significantly affect the results shown. Of course a much higher α would also shrink the curve to a point.

To quantify the robustness to noise of the Sobolev active contour versus the traditional H^0 active contour as seen in the previous experiment, we have conducted experiments with a synthetically generated image sequence (so that the ground truth is known) in which various degrees of noise are added. The sequence is binary images in which a square is translating (to represent motion) and changing its area slightly (to represent deformation). In the first experiment, we use the naive tracker and the detection energy (12). In this case the square is translated by three pixels and the length of the side is randomly changed by ± 2 pixels when compared with the square in the previous frame. The segmentation error for various degrees of Gaussian noise ($\mu = 0$, and standard deviation σ specified) using both H^0 and Sobolev $\tilde{H}^1$ active contours is shown in the left of Fig. 8. Similar experiments are done using the energy (13), but the square is translated by 17 pixels and the length is adjusted by a random ± 5 pixels. Results are shown in the right of Fig. 8. The results for the H^0 active contour are shown for the best value of α chosen for the given noise levels. Note that our measure of error is one-half of the number of false positive classified pixels plus

Figure 8: Plots of error for tracking a square that is translating and slightly changing its area with various degrees of noise (Gaussian mean 0, standard deviation σ). Left: Using geodesic active contours, right: using the Chan-Vese model. Note the difference in the scales of each plot; in particular, the plots show that the results of using Sobolev active contours is vastly better than the corresponding H^0 active contour.

(a) Tracking using the H^0 active contour.

(b) Tracking using the Sobolev ($\tilde{H}^1$) active contour.

Figure 9: Tracking of a sea-creature at the sea bottom using an energy which incorporates the mean intensity and variance information inside and outside the contour.

false negative classified pixels divided by the ground truth number of pixels of the object. In both cases of the detection energies chosen, the Sobolev active contour does significantly better than the corresponding H^0 active contour.

In the next experiment (Fig. 9), we demonstrate that the Sobolev active contour is useful not only for noisy situations, but in other cases where one is trying to track an object in a cluttered or textured environment where the object shares some visual characteristics with the background. In this experiment, we track a sea creature at the bottom of an ocean using the naive tracker and the detection energy,

$$E_{cv+\sigma}(c) = (1-\beta)E_{cv}(c) - \frac{1}{2}\beta(\sigma_u^2 - \sigma_v^2)^2$$

where

$$\sigma_u^2 = \frac{\int_{c_{in}} (I-u)^2 \, \mathrm{d}A}{\int_{c_{in}} \mathrm{d}A} \text{ and } \sigma_v^2 = \frac{\int_{c_{out}} (I-v)^2 \, \mathrm{d}A}{\int_{c_{out}} \mathrm{d}A}.$$

Since the mean values of some regions in the background are closer to the mean value inside the creature rather than to other regions of the background, a first order Chan-Vese energy is not enough to capture the object, and thus, we incorporate second order information. For this experiment, we chose $\beta = 0.6$ although different β produced similar results. In Fig. 9, we see that the H^0 active contour tracks the object for some time, but when the object's statistics look closer to the light part of the background than the dark part, the contour leaks into the background. On the other hand, because the Sobolev active contour moves globally before gradually changing its fine structure, the contour is able to avoid the distracting features of the background and get a rough approximation of the object before detecting finer features of the object, and thereby locking into a more desirable local minimum than the H^0-active contour.

Figure 10: Tracking of a car under an occlusion using the Mumford-Shah energy with H^0 (top) and $\tilde{H}^1$ active contours.

In the next experiment (in Fig. 10), we show that the Sobolev active contour can offer improvements over the traditional metric for tracking an object through partial occlusions. In particular, we track a car that moves under a lamp post. The energy functional used for the active contours is the Mumford-Shah functional [?]:

$$E_{ms}(c, f, g) = \int_{c_{in}} (I - f)^2 \, \mathrm{d}A + \int_{c_{out}} (I - g)^2 \, \mathrm{d}A \tag{14}$$

where f and g are smooth functions defined inside (resp. outside) the curve. The functional is minimized jointly in c, f, and g (see [?, ?, ?] for implementation details). A fixed number of iterations (300) is used to evolve the curve at each frame. The top row shows the H^0 active contour, which is thrown off as soon as the contour hits the lamp post. This is because each point of the H^0 active contour moves in a direction independent from the other points. Hence, the points close to the lamp post do not want to move past the post. On the other hand, the Sobolev $\tilde{H}^1$ flow moves globally first, and hence does not get stuck on the lamp post and continues to track the car, although at the end, the contour misses the outer parts of the car.

In the last experiment, we illustrate that the Sobolev active contour can improve the traditional active contour even when a prediction step is used to obtain the overall global motion (indeed an affine motion is predicted). The experiment (in Fig. 11) tries to address the problem with the previous experiment by using a predictor and observer/estimator. We use the detection/prediction algorithm considered in [?]. The detection step involves a simultaneous segmentation and rigid registration (i.e., affine) of three consecutive frames using the Mumford-Shah functional (14) and a fixed (300) number of iterations. For the prediction, a constant acceleration model is assumed for the parameters of the rigid registration. The measurements that the estimator uses to estimate the contour and its registrations are just the results of the detection step. A Kalman gain is used to determine if more weight is put on the measured contour versus the model prediction. As can be seen in Fig. 11, the H^0 active contour prefers coarse scale and fine scale perturbations equally and therefore the contour immediately becomes distracted by the pole, which is detected by fine scale perturbations, and the estimator/predictor is of little help. Note that a higher regularization penalty may be used; however, like the previous experiments, the length penalty, in addition to restricting the deformation into the pole, also shrinks the curve and a significant portion of the car is not detected. On the other hand, for the Sobolev $\tilde{H}^1$, the estimator/predictor greatly improves the result, as the shape is more accurately captured. Note that if the detection step is iterated for a very large number of iterations, the Sobolev active contour will also be distracted by the pole in this experiment. However, in real-time tracking applications it is often the case that the detection is not run until convergence; therefore, it is nice to know that with a limited number of iterations, the Sobolev active contour detects the coarse deformations, which are more essential than the fine deformations. In both cases (with and without the predictor/estimator), it is clear that simply replacing the standard H^0 active contour with the Sobolev active contour greatly improves the tracking performance.

Figure 11: Tracking a car under an occlusion using estimation with Mumford-Shah energy functional for the detection. H^0 (top) and $\tilde{H}^1$ (bottom) active contours.

6 Conclusion

We have shown that Sobolev active contours move successively from coarse to fine scale motions in a continuous manner through a scale-space type analysis. This property gives more justification for using the Sobolev framework. We have shown that this property, along with others, makes Sobolev active contours natural for tracking, and experiments have shown that the Sobolev technique is beneficial over the standard technique both when tracking with or without a predictor. The property of coarse to fine motion, as we saw, implies that Sobolev active contours take fewer iterations (and also less time) to converge to the desired local minimum than H^0 active contours. This is important for real-time tracking systems where a more efficient detection scheme with better accuracy is beneficial. Note that existing tracking algorithms, which use active contours, need not be modified; nor does the energy functional for the active contour, just a simple addition of a procedure to compute the Sobolev active contours is necessary, which is straight forward to obtain from the original active contour.

In this article, we have analyzed Sobolev gradient flows for curves and showed important properties of these flows, which are quite useful for tracking applications. The next step is to extend these ideas to surface evolutions, where there are many applications such as tracking of the heart in ultrasound sequences. Although the Sobolev method extends to surfaces, there are no equivalent convolution formulas (and in particular no simple integral solution as in $\tilde{H}^n$) for Sobolev gradients with respect to surfaces, and they are computationally expensive to compute. Future work is to formulate computationally feasible methods for computing the Sobolev gradients for surfaces, or consider alternate definitions of the Sobolev metric that lead to fast computations of the corresponding gradient.