International Journal of Computer Vision 73(3), 345–366, 2007

DOI: 10.1007/s11263-006-0635-2

Sobolev Active Contours

GANESH SUNDARAMOORTHI AND ANTHONY YEZZI
School of Electrical Engineering, Georgia Institute of Technology, Atlanta, USA
ganeshs@ece.gatech.edu

ANDREA C. MENNUCCI
Scuola Normale Superiore, Pisa, Italy

Received May 26, 2006; Revised November 2, 2006; Accepted November 2, 2006

First online version published in January, 2007

Abstract. All previous geometric active contour models that have been formulated as gradient flows of various energies use the same L^2-type inner product to define the notion of gradient. Recent work has shown that this inner product induces a pathological Riemannian metric on the space of smooth curves. However, there are also undesirable features associated with the gradient flows that this inner product induces. In this paper, we reformulate the generic geometric active contour model by redefining the notion of gradient in accordance with Sobolev-type inner products. We call the resulting flows Sobolev active contours. Sobolev metrics induce favorable regularity properties in their gradient flows. In addition, Sobolev active contours favor global translations, but are not restricted to such motions; they are also less susceptible to certain types of local minima in contrast to traditional active contours. These properties are particularly useful in tracking applications. We demonstrate the general methodology by reformulating some standard edge-based and region-based active contour models as Sobolev active contours and show the substantial improvements gained in segmentation.

Keywords: active contours, gradient flows, Sobolev norm, global flows, shape optimization

1. Introduction

Active contours, introduced by Kass et al. (1987), have been widely used for the segmentation problem. The idea is to minimize an energy, defined on contours or curves living in the domain of the image, that contains an image based edge attraction term and a smoothness term, which becomes large when the curve is irregular. An evolution is derived to minimize the energy and thereby drive the contour towards the boundary of the object of interest in the image while maintaining smoothness. There have been many variations to original model of Kass et al.; for example, see Cohen (1991), and a survey in Blake and Isard (1998). 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 there have been geometric evolutions similar to the idea of Kass et al. in Caselles et al. (1993) and Malladi et al. (1995), which can be implemented by level set methods (Osher and Sethian, 1988). Thereafter, Kichenassamy et al. (1995) and Caselles et al. (1995) considered minimizing a geometric energy, which is a generalization of Euclidean arclength, defined on curves for the edge-detection problem. The authors derived *the gradient descent flow* in order to minimize the geometric energy.

In the works of Siddiqi et al. (1998) and Xu and Prince (1998), the authors build on the active contour edge-detection approach to segmentation.

In contrast to the edge-based approaches for active contours (mentioned above), *region-based* energies (Ronfard, 1994; Zhu et al., 1995; Yezzi et al., 1999; Chan and Vese, 2001) for active contours have provided many desirable features; for example, they provide 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 these approaches, an energy is designed to be minimized when the curve partitions the image into statistically distinct regions. In Mumford and Shah (1985) and Mumford and Shah (1989), the authors introduce and rigorously study a region-based energy that is designed to both extract the boundary of distinct regions while also smoothing the image within these regions. Subsequently, Tsai et al. (2001a) and Vese and Chan (2002) gave a curve evolution implementation of minimizing the energy functional considered by Mumford and Shah (1985) and Mumford and Shah (1989) in a level set framework (Osher and Sethian, 1988). Another work that makes use of a region-based energy is Paragios and Deriche (2002a, b), which also combines edge-based information. Like the edge-based methods, *the gradient descent flows* are calculated to minimize these energies.

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 *a-priori* 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. (2001b), Rousson and Paragios (2002), Chen et al. (2002), Cremers and Soatto (2003) and 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 prior term of the energy that is small when the curve is *close*, in some sense, to a pre-specified shape. The need for this type of prior term arose from several factors including the fact that some images contain limited information, the energies 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.

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 (2004), Charpiat et al. (2005a), Michor and Mumford (2003) and Yezzi and Mennucci (2005a), in the *shape analysis* literature have proposed different ways of defining this distance. However, Michor and Mumford, Yezzi and Mennucci (2003, 2005b) 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 distance induced by a Riemannian structure. Indeed, the notion of gradient of such energies relies on an inner product defined the set of perturbations of a curve. All of these previous works on geometric active contours use the same geometric L^2-type inner product, which we refer to as H^0, to define gradient. However, Michor and Mumford, Yezzi and Mennucci (2003, 2005a) have shown a surprising property: the Riemannian metric on the space of curves induced by the H^0 inner product is pathological, i.e., 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 as we will demonstrate in this paper. 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. By smooth in the spaces of curves, we mean that the surface formed by plotting the evolving curve as a function of time is smooth. 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. As a result, the curve becomes non-smooth instantly. Therefore, in geometric active contours models, a penalty on the curve's length is added to keep the curve smooth in addition to producing a variational problem that is well-posed. However, this changes the energy that is being optimized. Moreover, as we shall demonstrate, the length penalty is sometimes insufficient, and leads to other problems.
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 image data. These facts imply that the H^0 active contour 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. This implies, as we shall show, that the H^0-gradient of image dependent energies "encourages" points on the evolving curve to move "independently" to decrease energy rather than encouraging the points to move more collectively. This restricts the gradient at a particular point from "seeing" information located at other points of the curve. Therefore, the curve evolving according to the H^0 gradient flow is susceptible to local minima of the energy.
4. Many geometric active contours (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. If the energy depends on n derivatives of the curve, then the H^0 gradient has $2n$ derivatives of the curve. Since the corresponding level set flows with higher than two derivatives are not known to have a maximum principle, level set methods (Osher and Sethian, 1988) are difficult to use for the numerical implementation (Chopp and Sethian, 1999). Therefore, one may try to use particle methods to implement the flow. However, flows with higher than two derivatives are generally difficult to implement because of numerical artifacts.
6. Lastly, as a specific example, the H^0 gradient ascent for arclength, i.e., backward heat flow, is ill-posed. This is quite odd in an intuitive manner because there is nothing in the definition of length itself that indicates that a flow to increase length should be ill-posed. We shall see that this ill-posedness is a property of H^0 not just the energy itself.

In this paper, we consider using inner products arising from Sobolev spaces to define gradients. Unlike the research done in active contours for the past 20 years, which has focused on finding various *new energies* to deal with local minima and other problems, we present a *new way* of doing active contours by giving an alternative way of minimizing active contour energies.

By changing the Riemannian metric associated with the space of curves, we are able to regularize the minimizing flows associated with active contour energies without requiring the addition of regularization penalties in the aforementioned energies. Indeed, the change of metric does not affect the global minima of the energy, but it changes the notion of gradient, and the notion of "neighborhood of a curve". While previous local minimizers continue to be local minimizers (or at least critical points) in the Sobolev metrics, the definition of locality is completely different. As a result, Sobolev active contours are much more robust to the local minima which strongly influence previous, standard, active contours. Moreover, any existing active contour method can get the added benefits of this new approach with minimal changes to its existing implementation and with little extra computational time.

We would like to point out that Sobolev gradient methods have been used in the past; for example the book (Neuberger, 1997) (see also references within) presents the Sobolev gradient and applies it to the numerical solution of various physical problems. More recently, Burger (2003) has examined using various Sobolev gradients (especially fractional Sobolev gradients) to various inverse problems of boundary reconstruction, and examined rates of convergence among the various gradients considered. We Sundaramoorthi et al. (2005) and (independently) Charpiat et al. (2005b) have introduced the Sobolev method to active contour problems. Charpiat et al. (2005b) also go on to consider other "coherent" motions resulting from various inner products (see also the work of Mansouri et al. (2004) for a related idea).

2. General Theory

In the next sections, we begin by using principles from Riemannian geometry (do Carmo, 1992; Lang, 1999) and show how they fit into the framework of geometric active contours, which is essential to construct the Sobolev active contour.

2.1. Structure on the Space of Curves

Let M denote the set of smooth immersed curves in $\mathbb{R}^d$, where $d \geq 2$. We show how to make M a differentiable manifold. Let us give a precise definition of M:

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

where S^1 denotes the circle. Denote by F the Fréchet space $C^\infty(S^1, \mathbb{R}^d)$ of smooth vector fields on S^1 whose topology is defined by the following seminorms:

$$p_n(h) = \sup_{\theta \in S^1} |h^{(n)}(\theta)|, \quad n = 0, 1, \ldots$$

where $h \in F$. We now show

Proposition 2.1. *M is a differentiable manifold whose model space is the infinite dimensional space, F.*

Proof:

1. We first define charts from M to F. Let $c \in M$, then denote by $\Phi_c^{-1} : \tilde{U}_c \subset F \to U_c \subset M$ a chart at $c \in M$ defined by $\Phi_c^{-1}(h) = c + h$. We must define the neighborhood $\tilde{U}_c$, and show that $c + h \in M$ where $h \in \tilde{U}_c$. Choose

$$0 < \epsilon < \inf_{\theta \in S^1} |c'(\theta)|.$$

   Note the choice of ϵ is possible since $|c'(\theta)| \neq 0$ ($c \in M$), and the infimum cannot be zero since $|c'|$ is continuous on the compact set S^1. Now define $\tilde{U}_c := \{h \in F : p_1(h) < \epsilon\}$, which by definition of F is open. It is now clear that $c + h \in M$ for $h \in \tilde{U}_c$. It is also clear that Φ_c is a bijection.
2. We now show that the charts defined above are compatible. Let $c, \tilde{c} \in M$, $c \neq \tilde{c}$ and $U_c \cap U_{\tilde{c}} \neq \emptyset$. First note that $\Phi_c(U_c \cap U_{\tilde{c}}) = \{h \in F : p_1(h) < \epsilon_c, p_1(h - (\tilde{c} - c)) < \epsilon_{\tilde{c}}\}$, which is clearly open. Next, we see that $\Phi_{\tilde{c}} \circ \Phi_c^{-1} : \Phi_c(U_c \cap U_{\tilde{c}}) \to \Phi_{\tilde{c}}(U_c \cap U_{\tilde{c}})$ is given by

$$\left(\Phi_{\tilde{c}} \circ \Phi_c^{-1}\right)(h) = (c - \tilde{c}) + h,$$

   which is clearly a C^∞ diffeomorphism.

□

Remark 2.1. Note, we may consider $M = \{c \in C^1(S^1, \mathbb{R}^d) : c'(\theta) \neq 0 \forall \theta \in S^1\}$ and F to be the Banach space $C^1(S^1, \mathbb{R}^d)$ with topology defined by the norm $\|h\| = \sup_{\theta \in S^1}(|h(\theta)| + |h'(\theta)|)$. In this case, M is also a differentiable manifold. However, the tangent to the curve, c' is no longer in the model space. See Mennucci et al. (2006) for more details.

Remark 2.2. Let $\mathrm{Diff}(S^1)$ be the set of smooth automorphisms of S^1. Note that in our definition of the manifold M, the space of *immersed* curves, two different parameterizations of a curve (i.e., $c_1, c_2 \in M$ such that $c_1 = c_2 \circ \phi$ where $\phi \neq \mathrm{id}_{S^1}$ is in $\mathrm{Diff}(S^1)$) are considered two distinct elements of M. For computer vision applications one should consider the quotient space

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

this quotient models the space of *geometrical curves*, which are "curves up to a choice of parametrization". This space, however, is not a manifold (Michor and Mumford, 2003); but a slightly smaller quotient space

$$B_f := \mathrm{Imm}_f/\mathrm{Diff}(S^1)$$

of the freely immersed curves is a manifold. In our work, we will use M. The results we obtain, however, are *geometric* in that they do not depend on a particular parameterization of a curve; so they may be "projected" to B_f.

For $c \in M$, we denote by T_cM the tangent space of M at c (i.e., the space of valid perturbations of c). Since M is an open subspace of $C^\infty(S^1, \mathbb{R}^d)$, the tangent space may be identified with $C^\infty(S^1, \mathbb{R}^d)$ itself. It is easy to see that this identification satisfies all standard requirements and/or different choices of definitions for T_cM.

2.2. Inner Products on T_cM

We now define inner products on T_cM.

Definition 2.1. Let $c \in M$, L be the length of c, and $h, k \in T_cM$. Let $\lambda > 0$, and $n \in \mathbb{N}$. We assume h and k are parameterized by the arclength parameter of c.

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

where $\int \cdot \, \mathrm{d}s$ is the integral w.r.t the arclength parameter, $\mathrm{avg}(h) := \frac{1}{L}\int_c h(s)\,\mathrm{d}s$, and $h^{(n)}$ denotes the n^{th} derivative of h with respect to arclength.

It is easy to verify that the above definitions are inner products. Note that we have introduced length dependent scale factors so that the above inner products (and the corresponding norms) are independent of curve rescaling (we do not want a rescaling of the curve to change the inner product).

Remark 2.3. The most general scale invariant definition of H^n is

$$\langle h, k\rangle_{H^n} = \langle h, k\rangle_{H^0} + \sum_{j=1}^{n} \lambda_j L^{2j} \left\langle h^{(j)}, k^{(j)}\right\rangle_{H^0} \quad (1)$$

where $\lambda_j \geq 0$ for $j = 1, \ldots, n-1$, and $\lambda_n > 0$. We will see in Section 2.4 that it is unnecessary for our purposes to use this general definition.

Remark 2.4. Other definitions of H^n and $\tilde{H}^n$ are possible (Yezzi and Mennucci, 2005a). We illustrate the ideas on planar curves ($d = 2$). We recall from Remark 2.2 that one main goal of the present theory is to be *geometric*; that is, any result may be projected to the space B_f of *geometrical curves*. A possible choice for representing the tangent $T_c B_f$ is $T_c M/\sim$ where $h \sim k$ iff $h\cdot\mathcal{N} = k\cdot\mathcal{N}$ where $\mathcal{N}$ is the inward normal of c. The other definitions are as follows: Let $\alpha, \beta : S^1 \to \mathbb{R}$ be the normal components of two perturbations h, k (i.e., $\alpha = h\cdot\mathcal{N}$ and $\beta = k\cdot\mathcal{N}$) then

$$\langle [h], [k]\rangle_{H^{1^*}} := \frac{1}{L}\int_c \left[\alpha(s)\beta(s) + \lambda L^2 \alpha'(s)\beta'(s)\right] \mathrm{d}s \quad (2)$$

$$\langle [h], [k]\rangle_{H^{1^+}} := \frac{1}{L}\int_c \left[\alpha(s)\beta(s) + \lambda L^2 (\alpha(s)\mathcal{N}(s))' \cdot (\beta(s)\mathcal{N}(s))'\right] \mathrm{d}s. \quad (3)$$

The inner product in (3) can be simplified as follows

$$\langle [h], [k]\rangle_{H^{1^+}} = \frac{1}{L}\int_c \left[(1 + \lambda L^2 \kappa^2(s))\alpha(s)\beta(s) + \lambda L^2 \alpha'(s)\beta'(s)\right] \mathrm{d}s. \quad (4)$$

Ignoring the $\alpha'\beta'$ term and the length factors, (4) becomes the same inner product considered in (Michor and Mumford, 2003) for shape analysis.

One of the important consequences of defining an inner product on $T_c M$ is that, with some other smoothness and compatibility conditions, it produces a *Riemannian* structure on M. This means that specifying an inner product on $T_c M$ induces a distance between points on M. The authors in Michor and Mumford (2003) and Yezzi and Mennucci (2005a) consider the Riemannian structure induced from the H^0 inner product and show that it is pathological; they also propose alternative inner products to achieve a non-trivial Riemannian distance. We explore this idea in Mennucci et al. (2006) (see also Younes (1998)).

2.3. Gradient of Functionals on M

We now define the notion of gradient of a functional $E : M \to \mathbb{R}$.

Definition 2.2. Let $E : M \to \mathbb{R}$.

1. If $c \in M$ and $h \in T_c M$, then the *variation of* E in the direction h is
$$\mathrm{d}E(c)\cdot h = \frac{\mathrm{d}}{\mathrm{d}t} E(c + th)\bigg|_{t=0},$$
where $(c+th)(\theta) := c(\theta) + th(\theta)$ and $\theta \in S^1$. Note $\mathrm{d}E(c) \in T_c^* M$ is a linear operator on $T_c M$, $\mathrm{d}E(c)$ is called the *differential at* c.
2. Assume $\langle , \rangle_c$ is an inner product on $T_c M$. The *gradient of* E is a vector field $\nabla E(c) \in T_c M$ that satisfies
$$\mathrm{d}E(c)\cdot h = \langle h, \nabla E(c)\rangle_c$$
for all $h \in T_c M$.

We now give an intuitive interpretation of the gradient, which tells us the significance of this perturbation. We show that the gradient is the most efficient perturbation; that is, the gradient maximizes the change in energy per "cost" of perturbing the curve.

Proposition 2.2. *Let $\|\cdot\|_c$ be the norm induced from the inner product $\langle\cdot,\cdot\rangle_c$ on $T_c M$. Suppose $\mathrm{d}E(c) \neq 0$, and the gradient with respect to $\langle\cdot,\cdot\rangle_c$ exists; then the problem*

$$\sup_{\{h\in T_c M, \|h\|_c = 1\}} \mathrm{d}E(c)\cdot h = \sup_{\{k\in T_c M, k\neq 0\}} \frac{\mathrm{d}E(c)\cdot k}{\|k\|_c} \quad (5)$$

has a unique solution, $k = \nabla E(c) \in T_c M$, $h = k/\|k\|$.

Proof: Note $\mathrm{d}E(c)\cdot h = \langle \nabla E(c), h\rangle_c \leq \{\|\nabla E(c)\|_c\}\ \|h\|_c$ by the Cauchy-Schwartz inequality; moreover, $\mathrm{d}E(c)\cdot(\nabla E(c)) = \|\nabla E(c)\|_c^2$. □

If the gradient of E exists, then by the proposition above, we have that $k = \nabla E(c)$ attains the supremum on the right hand side of (5). Note for $\lambda \to +\infty$, translations have negligible norm with respect to other directions in the tangent space, that is, if h is a translation (i.e., it is constant w.r.t. s) and k is not a translation, then $\lim_{\lambda\to+\infty} \|h\|/\|k\| = 0$, both when $\|\cdot\|$ is the norm induced by H^n and when it comes from $\tilde{H}^n$. In light of the interpretation of the gradient as the perturbation that attains the supremum in (5), it follows that translations are favored for gradients in H^n and $\tilde{H}^n$ when λ is large (if they reduce energy).

Remark 2.5 (Comment on H^n for $n \geq 2$). Translations are favored for H^n and $\tilde{H}^n$ gradients when $\lambda \to +\infty$. This can be quite important for tracking applications where the object to be tracked is usually translating. One may wonder whether using higher order Sobolev inner products, H^n and $\tilde{H}^n$ for $n \geq 2$, will favor higher order polynomial motions of degree n. Note however, that any polynomial perturbation defined on S^1, the circle, must be constant to conform to periodic boundary conditions. Thus, higher than order $n = 1$ Sobolev gradients also favor just translations. In this sense, there is not an advantage, with respect to gaining higher dimensional preferred motions, in using higher order Sobolev gradients. However, one gains added regularity of the gradient flow in using higher order Sobolev gradients.

Remark 2.6. For the inner product on $T_cM/\sim$ defined in (2), it follows that dilations (i.e., perturbations of the form $h = \pm\mathcal{N}$) are favored for gradients when λ is large.

2.4. Relation Between H^n and $\tilde{H}^n$

We show that the norms associated with the inner products H^n and $\tilde{H}^n$, i.e.,

$$\|h\|_{H^n} = \sqrt{\int_0^L \frac{1}{L}|h(s)|^2 + \lambda L^{2n-1}|h^{(n)}(s)|^2\,\mathrm{d}s} \quad (6)$$

$$\|h\|_{\tilde{H}^n} = \sqrt{|\mathrm{avg}(h)|^2 + \lambda L^{2n-1}\int_0^L |h^{(n)}(s)|^2\,\mathrm{d}s} \quad (7)$$

are equivalent.

Proposition 2.3 (Equivalence of H^n and $\tilde{H}^n$). *The norms defined by* (6) *and* (7) *on T_cM are topologically equivalent; that is,*

$$\alpha\|h\|_{\tilde{H}^n} \leq \|h\|_{H^n} \leq \beta\|h\|_{\tilde{H}^n}$$

for all $h \in T_cM$, and $\alpha, \beta > 0$ are not a function of $c \in M$ and $h \in T_cM$.

Proof: We first derive a simple Poincaré inequality. For convenience, we will replace $h : S^1 \to \mathbb{R}^d$ with its periodic extension $h : \mathbb{R} \to \mathbb{R}^d$. We have

$$h(u) - h(0) = \int_0^u h'(s)\,\mathrm{d}s = -\int_u^L h'(s)\,\mathrm{d}s,$$

since h is periodic. Now,

$$\begin{aligned}
h(u) - h(0) &= \frac{1}{2}\left(\int_0^u h'(s)\,\mathrm{d}s - \int_u^L h'(s)\,\mathrm{d}s\right) \Rightarrow \\
\mathrm{avg}(h) - h(0) &= \frac{1}{2L}\int_0^L \left(\int_0^u h'(s)\,\mathrm{d}s \right. \\
&\qquad \left. - \int_u^L h'(s)\,\mathrm{d}s\right)\mathrm{d}u \Rightarrow \\
|\mathrm{avg}(h) - h(0)| &\leq \frac{1}{2L}\int_0^L \left(\int_0^u |h'(s)|\,\mathrm{d}s \right. \\
&\qquad \left. + \int_u^L |h'(s)|\,\mathrm{d}s\right)\mathrm{d}u \\
&= \frac{1}{2L}\int_0^L \left(\int_0^L |h'(s)|\,\mathrm{d}s\right)\mathrm{d}u \\
&= \frac{1}{2}\int_0^L |h'(s)|\,\mathrm{d}s
\end{aligned}$$

so that by replacing 0 with an arbitrary point, we have

$$\sup_u |h(u) - \mathrm{avg}(h)| \leq \frac{1}{2}\int_0^L |h'(s)|\,\mathrm{d}s. \quad (8)$$

Now, using (8) and Hölder's inequality, we see

$$\begin{aligned}
\sqrt{\int_0^L |h(s) - \mathrm{avg}(h)|^2\,\mathrm{d}s} &\leq \frac{1}{2}\sqrt{L}\sup_u |h(u) - \mathrm{avg}(h)| \\
&\leq \sqrt{L}\int_0^L |h'(s)|\,\mathrm{d}s \\
&\leq L\sqrt{\int_0^L |h'(s)|^2\,\mathrm{d}s},
\end{aligned}$$

which is the Poincaré inequality.

We now prove the equivalence of the two norms. By Hölder's inequality, we have that

$$|\text{avg}(h)|^2 \leq \frac{1}{L}\int_0^L |h(s)|^2\,\mathrm{d}s$$

so that $\|h\|_{\tilde{H}^1} \leq \|h\|_{H^1}$. On the other hand, note that

$$\begin{aligned}\frac{1}{L}\int_0^L |h(s)-\text{avg}(h)|^2\,\mathrm{d}s \\ = \frac{1}{L}\int_0^L |h(s)|^2\,\mathrm{d}s - |\text{avg}(h)|^2,\end{aligned}$$

so that

$$\begin{aligned}\|h\|_{H^1}^2 &= \int_0^L \frac{1}{L}|h(s)|^2 + \lambda L|h'(s)|^2\,\mathrm{d}s \\ &= \frac{1}{L}\int_0^L |h(s)-\text{avg}(h)|^2\,\mathrm{d}s \\ &\quad + \int_0^L \lambda L|h'(s)|^2\,\mathrm{d}s + |\text{avg}(h)|^2 \\ &\leq |\text{avg}(h)|^2 + L(1+\lambda)\int_0^L |h'(s)|^2\,\mathrm{d}s \\ &\leq ((1+\lambda)/\lambda)\|h\|_{\tilde{H}^1}^2\end{aligned}$$

One can iterate Poincaré's inequality to demonstrate the equivalence of H^n and $\tilde{H}^n$. □

The use of the Poincaré inequality can also be used to show the equivalence between the generic H^n norm (1) and the H^n norm we consider (6).

Note that we have not established any relation between the geometry of the inner products H^n and $\tilde{H}^n$; however, in the next sections, we show that the gradients with respect to H^n and $\tilde{H}^n$ have similar properties.

3. H^1 and $\tilde{H}^1$ gradients

In this section, we describe how to compute first order Sobolev gradients from the H^0 gradient. Denote by $f = \nabla_{H^0}E(c)$ the gradient of E with respect to the H^0 inner product at c. We would like to compute first the H^1 gradient at c. Assuming $g = \nabla_{H^1}E(c)$ exists, we have for all $h \in T_cM$,

$$\begin{aligned}\mathrm{d}E(c)\cdot h &= \langle h, g\rangle_{H^0} + \lambda L^2\langle h', g'\rangle_{H^0} \\ &= \langle h, g - \lambda L^2 g''\rangle_{H^0}\end{aligned}$$

where we have integrated by parts and noted that we have periodic boundary conditions. Since gradients are unique (if they exist), we must have that

$$f(s) = g(s) - \lambda L^2 g''(s) \quad \text{where } s \in [0, L]. \tag{9}$$

Note that this is an ODE defined on $[0, L]$ with periodic boundary conditions, that is, all derivatives match on the boundary of $[0, L]$.

Now we take a similar approach to compute the $\tilde{H}^1$ gradient. Assuming $g = \nabla_{\tilde{H}^1}E(c)$ exists, we have

$$\begin{aligned}\mathrm{d}E(c)\cdot h &= \text{avg}(h)\cdot\text{avg}(g) + \lambda L^2\langle h', g'\rangle_{H^0} \\ &= \langle h, \text{avg}(g) - \lambda L^2 g''\rangle_{H^0}.\end{aligned}$$

Again by uniqueness, we have that $f = \text{avg}(g) - \lambda L^2 g''$. Noting periodic boundary conditions, we have that $\text{avg}(g) = \text{avg}(f)$, so

$$f(s) = \text{avg}(f) - \lambda L^2 g''(s) \quad \text{where } s \in [0, L] \tag{10}$$

and we have periodic boundary conditions.

3.1. *Solving the ODEs*

We want to first solve the ODE (9) for g. It suffices to solve (9) with the boundary conditions $g(0) = g(L)$ and $g'(0) = g'(L)$. One can show that $g(s) = \int_0^L k_\lambda(s, \hat{s})f(\hat{s})\,\mathrm{d}\hat{s}$, where $k_\lambda : [0, L]\times[0, L] \to \mathbb{R}$ satisfies the following conditions for all $s, \hat{s} \in (0, L)$

$$k_\lambda(s, \hat{s}) - \lambda L^2\frac{\partial^2 k_\lambda}{\partial s^2}(s, \hat{s}) = \delta(s - \hat{s}) \tag{11a}$$

$$k_\lambda(0, \hat{s}) = k_\lambda(L, \hat{s});\ \ \partial_s k_\lambda(0, \hat{s}) = \partial_s k_\lambda(L, \hat{s}) \tag{11b}$$

and δ denotes the Dirac distribution. It can be shown that the solution to the previous system is $k_\lambda(s, \hat{s}) = K_\lambda(|s - \hat{s}|)$, where $K_\lambda : \mathbb{R} \to \mathbb{R}$ is given by

$$K_\lambda(s) = \frac{\cosh\left(\frac{s-\frac{L}{2}}{\sqrt{\lambda}L}\right)}{2L\sqrt{\lambda}\sinh\left(\frac{1}{2\sqrt{\lambda}}\right)}, \quad \text{for } s \in [0, L], \tag{12}$$

and K_λ is periodically extended to all of $\mathbb{R}$. We may write

$$\boxed{\nabla_{H^1}E(s) = \int_c K_\lambda(\hat{s} - s)\nabla_{H^0}E(\hat{s})\,\mathrm{d}\hat{s} = (K_\lambda * \nabla_{H^0}E)(s)} \tag{13}$$

where the integral over c denotes any range of $\hat{s}$ that corresponds to one full period around the curve c (e.g. $[0,L]$, $[-L,0]$, $[-L/2,L/2]$, etc.).

We now solve the second ODE (10). It suffices to solve (10) with the boundary conditions $g(0) = g(L)$, $g'(0) = g'(L)$, and the relation $\text{avg}(f) = \text{avg}(g)$. Integrating (10) twice yields

$$g(s) = g(0) + sg'(0) - \frac{1}{\lambda L^2} \times \int_0^s (s - \hat{s})(f(\hat{s}) - \text{avg}(f))\,\mathrm{d}\hat{s}. \tag{14}$$

Using (14), applying the boundary conditions, and noting that $\text{avg}(g) = \text{avg}(f)$, after some manipulation, yields

$$g'(0) = -\frac{1}{\lambda L^3}\int_0^L s(f(s) - \text{avg}(f))\,\mathrm{d}s \quad \text{and}$$
$$g(0) = \int_0^L f(s)\tilde{K}_\lambda(s)\,\mathrm{d}s \tag{15}$$

where the formula for $\tilde{K}_\lambda$ is in (16).

Remark 2.6. We want to remark that the above yields a solution to the ODE (10) *without resorting to a convolution*; the same is true for the similar equation for the $\tilde{H}^n$ gradients. So, $\tilde{H}^n$ gradients may be computed from H^0 gradients in $O(N)$ time, where N is the number of sample points of the curve.

The Eq. (10) may be nonetheless solved using a convolution, where the kernel function $\tilde{K}_\lambda$ is given by

$$\tilde{K}_\lambda(s) = \frac{1}{L}\left(1 + \frac{(s/L)^2 - (s/L) + 1/6}{2\lambda}\right), \quad s \in [0, L]. \tag{16}$$

and it is (unsurprisingly) the same kernel that is used in (15). Note that $\tilde{K}_\lambda(0) = \tilde{K}_\lambda(L)$ and thus we may periodically extend $\tilde{K}_\lambda$ as before. In this case, we may rewrite, $g(0) = \int_c f(\hat{s})\tilde{K}_\lambda(\hat{s})\,\mathrm{d}\hat{s}$, where, again, the integral over c denotes any range of $\hat{s}$ that corresponds to one full period over c. Therefore,

$$\boxed{\nabla_{\tilde{H}^1} E(s) = \int_c \tilde{K}_\lambda(\hat{s} - s)\nabla_{H^0} E(\hat{s})\,\mathrm{d}\hat{s} = (\tilde{K}_\lambda * \nabla_{H^0} E)(s).} \tag{17}$$

3.2. Properties of the Kernels

Note the following formal properties of K_λ and $\tilde{K}_\lambda$:

$$K_\lambda''(s) = \frac{1}{\lambda L^2}(K_\lambda - \delta(s)) \quad \text{and}$$
$$\tilde{K}_\lambda''(s) = \frac{1}{\lambda L^2}\left(\frac{1}{L} - \delta(s)\right), \quad s \in [0, L). \tag{18}$$

The first property is just the relation in (11a), and the second is obtained through differentiation of $\tilde{K}_\lambda$. Using these relations, it is easy to see that $K_\lambda * f$ and $\tilde{K}_\lambda * f$ formally solve (9) and (10), respectively. Next, note that

$$\int_c K_\lambda(\hat{s})\,\mathrm{d}\hat{s} = 1 \quad \text{and} \quad \int_c \tilde{K}_\lambda(\hat{s})\,\mathrm{d}\hat{s} = 1 \tag{19}$$

for all $\lambda > 0$. Also observe that $K_\lambda \geq 0$ for all $\lambda > 0$, and that $\tilde{K}_\lambda \geq 0$ only when $\lambda \geq 1/24$. Finally, it is easy to verify that as $\lambda \to +\infty$, $K_\lambda \to 1/L$ and $\tilde{K}_\lambda \to 1/L$. See Fig. 1 for plots of K_λ and $\tilde{K}_\lambda$.

3.3. Properties of Sobolev Gradients

First note, from formulas (13) and (17), that the H^1 and $\tilde{H}^1$ gradients are geometric, i.e., they do not depend on a particular parameterization chosen for the curve. This is also evident from the definition of these inner products. The formulas (13) and (17) show that there may be a tangential component of the gradients; but these tangential components may be ignored when considering gradient flows. This is different from H^0 where if the energy is geometric, then the gradient will have only a normal component.

Because H^1 and $\tilde{H}^1$ gradients are given by integrals of the H^0 gradient, given in formulas (13) and (17), integration by parts and the relations in (18) imply that two derivatives of the curve can be moved to derivatives on the kernels. This means that H^1 and $\tilde{H}^1$ gradients involve two fewer derivatives of the curve than H^0 gradients involve. Note that H^0 gradients have twice the number of derivatives of the curve as is defined in the energy E to be optimized. Thus, fourth order evolution equations of curves in H^0 may reduce to second order equations in H^1 and $\tilde{H}^1$. A similar remark can be made for H^n and $\tilde{H}^n$ gradients; these gradients require $2n$ less derivatives of the curve than the H^0 gradient requires.

The property that the integral of both the kernels is unity (19) implies that the H^1 gradient can be interpreted as a weighted average of the H^0 gradient; the

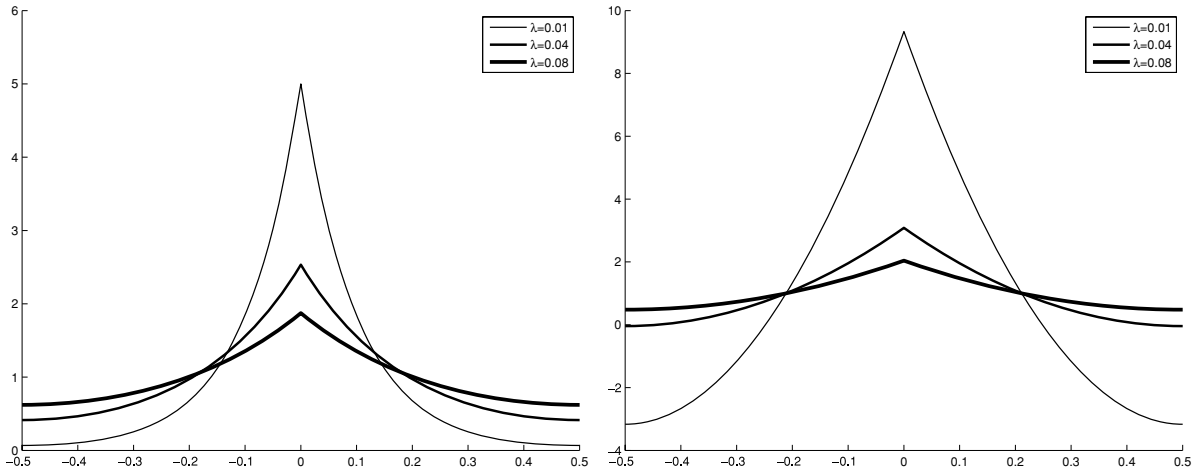


Figure 1. Plots of K_λ (left) and $\tilde{K}_\lambda$ (right) for various λ with $L = 1$. The plots show the kernels over one period.

same interpretation holds for $\tilde{H}^1$ when $\lambda > 1/24$. In light of this weighted average interpretation, we see that Sobolev gradients are less sensitive to noise and local features in the image than H^0 gradients are. Moreover, the property that the kernels approach $1/L$ as $\lambda \to +\infty$ shows that, in this case, the H^1 and $\tilde{H}^1$ gradients approach pure translations equal to the average value of the H^0 gradient, as expected from the interpretation of gradient noted in Section 2.1.

While local minimizers in the H^0 metric continue to be local minimizers in the H^1 metric (but not vice-versa), the definition of locality is completely different. As a result, Sobolev active contours are much more robust to the local minima which strongly influence H^0 active contours. This property arises due to the fact that Sobolev active contours are able to exploit much more relevant information contained in the initial contour about the shape of the desired contour. This is because the Sobolev active contour is more resistant to changing the local structure of the initial curve since local perturbations induce high derivatives in the flow field compared to more global motions which are smoother. Thus, if one starts with a contour that is smooth on a local scale, the cost to perturb the initially smooth contour into a "noisy" version of the same contour is enormous. This renders a very large distance between a smooth and non-smooth contour even though the two contours may be very close in other senses, such as in the Hausdorff distance. As such, local minimizers due to noise are no longer "local" to the initial contour if the initial contour is smooth and the local minimizer is not. Switching to the Sobolev metric has the effect of pushing many undesirable minimizers so far away from the initial contour that they are no longer able to influence the gradient flow, which moves towards a more desirable minimizer that is now much closer to the initial contour.

Remark 3.2. One may wonder whether any smoothing kernel can used to smooth out the H^0-gradient to form a smoother perturbation, which is robust to noise. The answer may be yes, but it is not guaranteed to reduce the energy in question. The fact that the kernels we've derived come directly from a gradient means that the energy is guaranteed to be reduced.

The previous comment also relates to various numerical methods for energy descent algorithms where the H^0 gradient is modified by an operator known as a pre-conditioner to yield a better descent direction for faster convergence to a solution. For example, the Newton method uses the inverse Hessian of the energy as the pre-conditioner (see for example, Hintermüller and Ring (2004) and Burger and Osher (2005) for applications to active contours). The disadvantage of this approach is that one has to be careful to make sure the resulting algorithm reduces the energy; even for the Newton algorithm, the inverse Hessian pre-conditioner is not guaranteed to reduce the energy, in general.

3.4. *Advantages of $\tilde{H}^1$ over H^1*

There is a computational advantage of using the $\tilde{H}^1$ gradient as opposed to the H^1 gradient as mentioned in

Remark 2.6. Another advantage of $\tilde{H}^1$ over H^1 is that we can eliminate the dependence on the parameter λ when implementing $\tilde{H}^1$ gradient flows. Observe from the kernel definition in (16) that $\tilde{K}_\lambda$ is a sum of two terms: one that depends on λ and another that does not. Thus, the $\tilde{H}^1$ gradient is a sum of two components: one that depends on λ by a simple scale factor, and another that is independent of λ. The component that does not depend on λ is $\text{avg}(\nabla_{H^0} E)$, which is a just a translation. The other component is a complex deformation. An algorithm to implement an approximate version of the $\tilde{H}^1$ gradient flow is to first evolve the curve by the translation component until this component becomes zero, then to evolve the curve by the deformation component, and the process is repeated until convergence. Note that λ does not need to be chosen for evolving the deformation component because λ only changes the speed of the curve, not the geometry. Therefore, this algorithm also gives a way of separating the (rigid) motion of the curve from the deformation. Separating the motion from deformation has particular importance in tracking applications (Soatto and Yezzi, 2002). Note that this algorithm is equivalent to optimizing the energy first according to the $\tilde{H}^1$ inner product as $\lambda \to +\infty$, and then according to $\tilde{H}^1$ with any λ (as there is no dependence on λ after the first step).

4. Some Sobolev Gradient Flows

In this section, we simplify the formulas (13) and (17) for some common geometric energies defined on curves in the plane ($c : S^1 \to \mathbb{R}^2$), note some interesting properties, and compare these with the usual H^0 gradients. A question that arises when considering these gradient flows is whether an initial curve $c \in M$ in the manifold of curves, stays in the manifold of curves (this also relates to existence of a solution for the PDE). The manifold of curves consists of curves being immersed and regular. Since the flows we consider are geometric and we represent the curve by arclength parameterization, the curves always remain immersed by representation if their derivative is defined. The question of regularity is a difficult one in general, and we do not address it in this paper (but for a simple result in Proposition 4.1).

In what follows, we use K to denote either the kernel (12) or (16), and ∇_1 will denote either the H^1 or $\tilde{H}^1$ gradient; when the distinction is needed, we will use the subscript λ on the kernels, and write H^1 or $\tilde{H}^1$.

4.1. Length and Weighted Length

We consider the geodesic active contour model (Caselles et al., 1995; Kichenassamy et al., 1995). The energy is

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

where $\phi : \mathbb{R}^2 \to \mathbb{R}^+$. Then the gradient with respect to H^0 is

$$\nabla_{H^0} E = L(\nabla\phi(c) \cdot \mathcal{N})\mathcal{N} - L\phi(c)\kappa\mathcal{N}$$

where $\mathcal{N}$ is the unit inward normal, and κ is the curvature. We omit the argument in $\phi(c)$, for simplicity, and write c' for arc parameter derivative. Let us first note that $\nabla_{H^0} E = L\nabla\phi - L(\phi c')'$. Integrating by parts we find that

$$\begin{aligned}\frac{1}{L}\nabla_1 E &= \nabla\phi * K - (\phi c')' * K \\ &= \nabla\phi * K - (\phi_s c) * K' - (\phi c) * K'',\end{aligned}$$

where $\phi_s := (\phi \circ c)'$. Using the relations in (18), we find that

$$\nabla_{\tilde{H}^1} E = \frac{\phi c - \text{avg}(\phi c)}{\lambda L} - L(\phi_s c) * \tilde{K}'_\lambda + L\nabla\phi * \tilde{K}_\lambda. \tag{20}$$

The above does not require that the curve be twice differentiable, and thus we may prove that

Proposition 4.1. *Suppose that $\phi \in C^1$, $\phi > 0$; then the gradient flow*

$$\frac{dc}{dt} = -\nabla_{\tilde{H}^1} E(c)$$

exists for small (positive) times; if $\liminf_{x\to\infty} \phi(x)$ $|x| > 0$, then it exists for all (positive) times.

Proof: The proof is based on the inequality

$$|\nabla_{\tilde{H}^1} E(s)| \le M_r(1+r)\frac{1}{\lambda} + LM_r r\frac{1}{2\lambda} + LM_r \quad \forall s$$

where the ball of radius r centered at 0 contains c, and

$$\frac{1}{M_r}E(c) \le L \le \frac{1}{m_r}E(c)$$

where

$$M_r := \sup_{|x|<r}(\max\{\phi(x), |\nabla\phi(x)|\}) \text{ and}$$
$$m_r := \inf_{|x|<r}\phi(x);$$

the above (by using the Euler method) implies existence, in the class of curve evolutions $c(\theta, t)$ that are Lipschitz in both variables.

Details will be available in a forthcoming paper. □

Of particular interest is when $\phi = 1$, that is $E = L$, the length of the curve. We see that

$$\nabla_{\tilde{H}^1} L = \frac{c - \text{avg}(c)}{\lambda L}.$$

It is interesting to notice that the H^1 and $\tilde{H}^1$ gradient flows are stable for both ascent and descent while the H^0 gradient flow is only stable for descent. Note that the $\tilde{H}^1$ gradient flow constitutes a simple rescaling of the curve about its centroid. While the H^0 gradient descent smooths the curve, the $\tilde{H}^1$ gradient descent (or ascent) has neither a beneficial nor a detrimental effect on the regularity of the curve.

4.2. Area and Weighted Area

We consider region-based active contour models; for example, Yezzi et al. (1999) and Chan and Vese (2001). The energy is

$$E(c) = \int_{c_{in}} \phi \, \mathrm{d}A,$$

where c_{in} denotes the region enclosed by the closed curve c, $\phi : \mathbb{R}^2 \to \mathbb{R}$ and $\mathrm{d}A$ is the area form; this energy is defined only for embedded curves. The gradient with respect to H^0 is

$$\nabla_{H^0} E = -L\phi\mathcal{N} = -L\phi Jc'$$

where J is a rotation by 90° matrix. Integrating by parts we find that

$$\frac{1}{L}\nabla_1 E = -(\phi Jc') * K = (\phi_s Jc) * K + (\phi Jc) * K'. \tag{21}$$

For the $\tilde{H}^1$ gradient, this simplifies to

$$\nabla_{\tilde{H}^1} E = \frac{J}{\lambda L^2}\int_0^L (\phi c(\cdot + \hat{s}) - \text{avg}(\phi c))\hat{s}\,\mathrm{d}\hat{s} + (\phi_s Jc) * \tilde{K}_\lambda. \tag{22}$$

Of particular interest is when $\phi = 1$, that is $E = A$, the area enclosed by the curve. We see that $\nabla_1 A = (Jc) * K'$. This simplifies to the gradient ascent/descent

$$C_t(s) = \pm\frac{J}{\lambda L^2}\int_0^L (C(s + \hat{s}) - \text{avg}(C))\hat{s}\,\mathrm{d}\hat{s} \tag{23}$$

in the $\tilde{H}^1$ gradient case.

4.3. Elastic Energy

Consider the elastic energy defined by

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

where L is the length of c, and κ is the signed curvature. This may serve as a regularizer that does not favor smaller length curves, unlike the standard curve shortening term. It is also the term that is commonly omitted in the original snakes model of Kass et al. since it leads to a fourth order gradient flow.

We now derive the $\tilde{H}^1$ flow by first calculating the H^0- gradient. Let us write a family of curves as $C : \mathbb{R}^+ \times S^1 \to \mathbb{R}^2$, then we may write

$$E(t) = \int_C C_{ss} \cdot C_{ss}\,\mathrm{d}s.$$

Let us first compute some intermediate formulas. Let $f : \mathbb{R}^+ \times S^1 \to \mathbb{R}$, then

$$\begin{aligned}\frac{\partial}{\partial t}\frac{\partial}{\partial s}f &= \frac{\partial}{\partial t}\frac{1}{\|C_p\|}\frac{\partial}{\partial p}f \\ &= -\frac{C_{tp}\cdot C_p}{\|C_p\|^3}\frac{\partial}{\partial p}f + \frac{1}{\|C_p\|}\frac{\partial}{\partial p}\frac{\partial}{\partial t}f \\ &= -C_{ts}\cdot C_s\frac{\partial}{\partial s}f + \frac{\partial}{\partial s}\frac{\partial}{\partial t}f,\end{aligned}$$

or more conveniently,

$$f_{st} = f_{ts} - (C_{ts}\cdot C_s)f_s. \tag{24}$$

We now compute

$$\frac{\partial}{\partial t}(C_{ss}\cdot C_{ss}) = 2C_{sst}\cdot C_{ss},$$

but using (24) twice, we see

$$\begin{aligned} C_{sst} &= C_{sts} - (C_{ts}\cdot C_s)C_{ss} \\ &= \frac{\partial}{\partial s}(C_{ts} - (C_{ts}\cdot C_s)C_s) - (C_{ts}\cdot C_s)C_{ss} \\ &= C_{tss} - (C_{tss}\cdot C_s + C_{ts}\cdot C_{ss})C_s \\ &\quad -(C_{ts}\cdot C_s)C_{ss} - (C_{ts}\cdot C_s)C_{ss} \\ &= C_{tss} - (C_{tss}\cdot C_s + C_{ts}\cdot C_{ss})C_s - 2(C_{ts}\cdot C_s)C_{ss}, \end{aligned}$$

and since $C_s\cdot C_{ss} = 0$, we have

$$\frac{\partial}{\partial t}(C_{ss}\cdot C_{ss}) = 2(C_{tss}\cdot C_{ss}) - 4(C_{ts}\cdot C_s)(C_{ss}\cdot C_{ss}). \tag{25}$$

Now,

$$\begin{aligned} E'(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\int_0^1 C_{ss}\cdot C_{ss}\|C_p\|\,\mathrm{d}p \\ &= \int_0^1 \frac{\partial}{\partial t}(\|C_p\|)C_{ss}\cdot C_{ss}\,\mathrm{d}p \\ &\quad + \int_C \frac{\partial}{\partial t}(C_{ss}\cdot C_{ss})\,\mathrm{d}s. \end{aligned}$$

By substituting (25), we have

$$\begin{aligned} E'(t) &= \int_C (C_{ts}\cdot C_s)(C_{ss}\cdot C_{ss})\,\mathrm{d}s \\ &\quad + 2\int_C (C_{tss}\cdot C_{ss}) - 2(C_{ts}\cdot C_s)(C_{ss}\cdot C_{ss})\,\mathrm{d}s \\ &= -\int_C 3(C_{ts}\cdot C_s)(C_{ss}\cdot C_{ss})\,\mathrm{d}s \\ &\quad + 2\int_C (C_{tss}\cdot C_{ss})\,\mathrm{d}s \\ &= \int_C (2(C_{tss}\cdot C_{ss}) - 3(C_{ts}\cdot C_s)(C_{ss}\cdot C_{ss}))\,\mathrm{d}s. \end{aligned}$$

Integrating by parts, we find

$$\begin{aligned} E'(t) = \frac{1}{L}\int_C C_t\cdot(&2L\partial_{ss}(C_{ss}) \\ &+ 3L\partial_s((C_{ss}\cdot C_{ss})C_s))\ \mathrm{d}s. \end{aligned}$$

Hence,

$$\nabla_{H^0}E(c) = 2Lc_{ssss} + 3L\partial_s((c_{ss}\cdot c_{ss})c_s). \tag{26}$$

Computing the Sobolev gradient from (26), we have

$$\begin{aligned} \nabla_{\tilde{H}^1}E &= K * \nabla_{H^0}E \\ &= 2LK'' * (c_{ss}) - 3LK' * ((c_{ss}\cdot c_{ss})c_s) \end{aligned}$$

Hence, for the $\tilde{K}_\lambda$ kernel we have

$$\nabla_{\tilde{H}^1}E = -\frac{2}{\lambda L}\kappa\mathcal{N} - 3LK' * (\kappa^2\mathcal{T}) \tag{27}$$

since $\mathrm{avg}(\kappa\mathcal{N}) = 0$, and $\mathcal{T}$ is the unit tangent to c. Notice that the corresponding $\tilde{H}^1$ gradient flow is second order, although it is an integral-PDE.

4.4. Comparison of H^0 and H^1, $\tilde{H}^1$

We notice several advantages of the gradients flows for H^1 and $\tilde{H}^1$ gradients as compared with H^0 gradients. First note that both the expressions for edge-based and region-based active contour gradients with respect to H^1 and $\tilde{H}^1$ (20), (21) do not involve any derivatives of the curve. This is in contrast to H^0, which requires two derivatives for geodesic active contours and one derivative for region-based active contours. Hence, these Sobolev flows are defined for non-smooth curves, e.g., polygons, without the need to resort to numerical techniques based on viscosity solutions of the corresponding level set equations. Moreover, Proposition 4.1 shows that we do not need viscosity theory to define the $\tilde{H}^1$ flow at least for the weighted length energy. Note that the expression in (20) does not require any more derivatives of ϕ than the expression for H^0 does. This is not the case for (21), which requires a derivative of ϕ. However, since ϕ_s is contained within a convolution, the possible noise generated by ϕ_s is mitigated. Alternatively, the original expressions (13) and (17) may be used if a derivative of ϕ is not desired to be computed.

Notice the expressions of Sobolev gradients for the elastic energy (27) only require two derivatives of the curve; this is in contrast to the H^0 gradient, which requires four derivatives of the curve. Since there is no maximum principle for fourth order equations, the H^0 gradient descent of the elastic energy is difficult to implement using level set methods (Chopp and Sethian, 1999; Droske and Rumpf, 2004). A particle method can

be used; however, this is prone to numerical problems. Note that the integral-PDE (27) may not have a maximum principle, but in our numerical implementation with level set methods (see Section 5.2), we bypass this issue for the following reason. The local term certainly has a maximum principle, and for the global term, we perform *extensions* of this quantity from the zero level set, which is done, for example, in image-based evolutions. Therefore, we expect that level sets do not collide, that the level set function gradient doesn't become ill-defined, and that the level set evolution is modeling the curve evolution. These are issues that need to be dealt with for the H^0 gradient flow using a level set method.

5. Numerical Implementation

In this section, we describe the numerical scheme used to simulate Sobolev active contours defined in the plane. Sobolev active contours are naturally suited for a parametric or a marker particle based implementation. This is because computing Sobolev gradients requires computing an integral around the curve, which is straightforward to compute if one has an ordered set of sample points of the curve.

5.1. Multiple Curves Evolution

In this section, we consider evolving multiple curves according to gradient flows for the metrics defined in Section 2.2. We consider defining the flows for curves that undergo topological changes. Consider a family of embedded curves $c = (c_i)$ in the plane and of displacements $h = (h_i)$, where $c_i, h_i : S^1 \to \mathbb{R}^d$ and $i = 1, \ldots, N = N(c)$. We may define a metric on multiple curves $\langle , \rangle_c$ using a choice of some metric $\langle , \rangle_{c_i}$ that was defined for a single curve in Section 2.2. We define the inner product on multiple curves to be

$$\langle h, k \rangle_c := \sum_i \langle h, k \rangle_{c_i}.$$

Consider an energy, E, on multiple curves defined by

$$E(c) := \sum_{i=0}^{N(c)} E(c_i),$$

then, the variation of E for several embedded curves is

$$\mathrm{d}E(c) \cdot h = \sum_{i=0}^{N(c)} \langle \nabla E(c_i), h_i \rangle_{c_i},$$

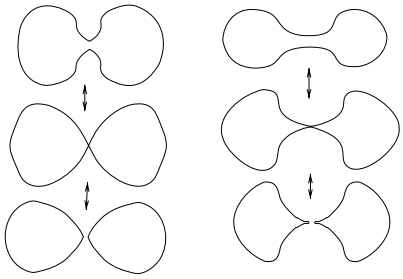


Figure 2. Some illustrations of topological changes.

for some given inner product. Thus, we consider evolving the curves according to the curve evolution

$$\partial_t c_i = -\nabla E(c_i) \quad i = 1, \ldots, N(c). \tag{28}$$

Clearly, if the curves are moved according to the above equation, and the curves do not undergo topological changes then the evolution reduces the energy, E.

We now consider the possibility of topological changes of c, that is merging and splitting of curves in c. Two ways of topological changes are pictorially shown in Fig. 2. On the left, the curves develop kinks; on the right, they develop cusps. Note that other topological changes may be possible.

We consider whether the evolution of (28) reduces the energy, E of the curve c, even when topology changes occur. First, we note a problem for actually defining the evolution in (28) for Sobolev metrics. Note that during the instant in which the curve undergoes a topology change an ambiguity arises in how to parameterize the curve at the instant of self-intersection. Since the Sobolev inner product depends upon the arclength parameterization of the curve this leads to an ambiguity in defining the gradient and the resulting curve evolution. For example, on the right side of Fig. 2, at the instant of topology change, there is an ambiguity over whether to treat c as a single curve or as two separate curves. In the case of the H^0 gradient, this ambiguity is not important since in both cases the H^0 gradient is the same at all points of the curve except the point of self-intersection as the H^0 does not depend upon the curve globally as does the Sobolev gradient. Since gradient descent flows are designed to minimize their corresponding energy functional as quickly as possible, there is a very natural solution to the ambiguity that arises in the case of the Sobolev gradient at the moment a topological change occurs. In theory, one may simply

compute the derivative of the energy corresponding to either interpretation of the curve (each of which results in a different gradient under the Sobolev metric) and choose the one with the larger derivative as it will lead to a faster instantaneous descent of the energy. In practice, though, an even simpler numerical procedure may be employed based on the same essential idea. We will outline the procedure below for the case of level set methods.

When employing level set methods for Sobolev active contours (as described in Section 5.2), the contour extracted at any discrete moment in time from the discrete spatial grid always exhibits the simple topology of an embedded curve(s), which bypasses the ambiguity discussed above. Thus a topological change occurs more abruptly after one discrete evolution step in which the curve passes from one topology to another, never passing through the ambiguous intermediate configuration in which the curve no longer remains embedded. Thus, there is never a need to resolve the Sobolev gradient of the curve in such configurations. Instead, noting that there is typically a discontinuity in the Sobolev gradient flow before and after topology change, we may simply verify that the energy has decreased after the topological change. If E is continuous, there is no reason to expect repeated oscillations in the topology of the curve even though the Sobolev gradient flow may exhibit a discontinuity across the topology change. The weighted length and area energies defined in Section 4 satisfy this continuity condition (since these functionals remain unambiguous through a topology change even though their Sobolev gradients do not). Indeed on the right-hand side of Fig. 2, these energies remain constant for the curve just before and just after the topology change. On the other hand, energies that depend on local derivatives of the curve such as the elastic energy defined in Section 4 may not be continuous through the topology change. As such there is no guarantee that the energy E decreases after an evolution step that creates a topological change. If the energy decreases, it makes sense to continue evolving. If instead an increase is detected, then one should stop evolving the curve and restore its prior configuration. Otherwise, oscillations may occur.

While the elastic energy pose a potential problem for Sobolev metrics (and indeed even the H^0 metric) in the event of certain types of topological changes (those for which the elastic energy increases), they pose an even greater problem for the standard H^0 metric as discussed in Section 4.4.

5.2. Level Set Method

We first describe a straightforward numerical method to extend curve evolutions that depend on integrals around the curve to level set evolutions, introduced in Osher and Sethian (1988). The algorithm for updating the level set function, $\Psi : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$, for most Sobolev active contours is

1. Compute polygon(s) estimate of zero level set of Ψ from the narrowband, U_n, a small thickening of the zero level set.
2. Calculate and interpolate H^0-gradient to polygon(s) estimate.
3. Compute Sobolev gradient on polygon(s) estimate by using one of the formulas for the H^1 or $\tilde{H}^1$-gradients in terms of the H^0-gradient shown in Section 3.1.
4. Extend polygon(s) forces to narrowband band of level set function domain.
5. Evolve Ψ by the transport equation $\Psi_t = -\nabla\Psi \cdot \vec{F}$, where $\vec{F} : U_n \subset \mathbb{R}^2 \to \mathbb{R}$ is the extended Sobolev gradient to the narrowband region, U_n.

The computational complexity of the entire algorithm for extracting the polygon(s), and computing the extensions is linear in the size of the narrowband, U_n.

In most of the cases of various energy functionals, the H^0 gradient of the energy must be computed on the narrowband U_n in the usual fashion; or in our case, we directly compute the H^0-gradient to the polygon estimate of $\{\Psi = 0\}$. In other cases (such as for the elastic energy), the H^0 gradient does not need to be computed, but only certain expressions that do not involve the convolutions. For example, in the elastic energy, the term $\kappa^2\mathcal{T}$ should be computed at every point of the polygon estimate. Note that in computing quantities such as the normal vector, and curvature of $\{\Psi = 0\}$, we compute them directly from Ψ using standard formulas and interpolate them to the polygon estimate. The Sobolev gradient is then computed using a convolution or simple integral with the formulas in Section 3.1. Note that we must compute separate convolutions or integrals on each polygon extracted from Ψ (or connected component of Ψ). We use a standard Riemann sum to compute the curve integrals, and the polygon estimate is used to determine the arclength measure, $\mathrm{d}s$. After the Sobolev gradient is computed on the polygon estimate, it can be extended to U_n. We call the extended Sobolev gradient $\vec{F}$. Note that we are extending the gradient so that

$\nabla\Psi \cdot \nabla F_i = 0$ where $i = 1, 2$ and $\vec{F} = (F_1, F_2)$. That is, we assume that all the level sets of Ψ are moving by the same speed function, which means that the level sets of Ψ do not collide. Now the level set evolution can be computed by

$$\Psi_t(x) = -\nabla\Psi(x) \cdot \vec{F}(x).$$

5.3. *Level Set Method Without Polygon Extraction*

In this section, we describe an approach where we can directly compute (without polygon extraction) an approximation to the Sobolev gradient from the H^0 gradient that is defined on the narrowband, U_n, of the level set function. This may be of use since the polygon extraction step is cumbersome.

Our approach is to approximate the convolution integrals (13) and (17) with region integrals evaluated within the narrowband U_n. The arc-distance between points on the curve (i.e., $s - \hat{s}$) that is required by the formulas (13) and (17) can be approximated by the use of the eikonal equation (Rouy and Tourin, 1992). Consider an embedded curve, $c \in C^2(S^1, \mathbb{R}^2)$, and the function $F : \mathbb{R}^2 \to \mathbb{R}^+$ defined as

$$F(x) = \begin{cases} +\infty & \text{for } x \notin \{c\} \\ 1 & \text{for } x \in \{c\} \end{cases}$$

where $\{c\} = \{c(\theta) : \theta \in S^1\}$. Define $u : \mathbb{R}^2 \to \mathbb{R}^+$ by

$$u(c(s)) := \inf_{\gamma \in \Gamma} \int_0^1 F(\gamma(\theta))\|\gamma'(\theta)\| \, \mathrm{d}\theta,$$

where $\Gamma := \{\gamma : [0, 1] \to \mathbb{R}^2, \gamma(0) = c(\hat{s}), \gamma(1) = c(s)\}$; then clearly, we see that $u(c(s))$ equals the minimum distance between the points $c(s)$ and $c(\hat{s})$ along the curve c. Because of the symmetry of the kernels in (12) and (16), it suffices to use the quantity $u(c(s))$ as a substitute for $s - \hat{s}$ in the convolutions formulas (13) and (17). As in Rouy and Tourin (1992), we may solve for the viscosity solution of the eikonal equation

$$\|\nabla u(x)\| = F(x), \quad u(c(\hat{s})) = 0 \tag{29}$$

to obtain the desired solution of u at all points along the curve, c. As we will solve this equation numerically on a grid, we consider the following approximation to (29):

$$\|\nabla u(x)\| = 1 + |\Psi(x)|/\epsilon, \quad u(c(\hat{s})) = 0 \tag{30}$$

where $\epsilon > 0$ is chosen small enough, and Ψ is the level set function with $\Psi(c(s)) = 0$ for all s.

Next, note the co-area formula for a Borel measurable function $f : \mathbb{R}^2 \to \mathbb{R}$, a Borel set $A \subset \mathbb{R}^2$ such that $\nabla\Psi(x) \neq 0$ for $x \in A$:

$$\int_A f(x)\|\nabla\Psi(x)\| \, \mathrm{d}x = \int_{\Psi(A)} \int_{\Psi^{-1}(t)} f(x) \, \mathrm{d}\mathcal{H}^1(x) \, \mathrm{d}t$$

where $\mathrm{d}\mathcal{H}^1$ is the one-dimensional Hausdorff measure, that is, arc-length measure. We use this formula to convert the contours integrals of interest to integrals over the domain of the level set (or the narrowband region). Thus, we find by the co-area formula and the Lebesgue differentiation theorem,

$$\int_c H(c(s))K(\hat{s} - s) \, \mathrm{d}s \approx \int_A H(x)K(u(x))\delta_a(\Psi(x))\|\nabla\Psi(x)\| \, \mathrm{d}x$$

where $\delta_a : \mathbb{R} \to \mathbb{R}$ is a smooth approximation to the Dirac distribution, $H : \mathbb{R}^2 \to \mathbb{R}$ is some function, u is the solution to (30), and $A \subset \mathbb{R}^2$, such that $\{c\} \subset A^o$, the interior of A. According to the previous comments, we propose the following theorem:

Theorem 5.1. *Suppose that $c : S^1 \to \mathbb{R}^2$ is a C^2 curve embedded in the plane, Ψ is zero only on the image of c, and $\nabla\Psi(c(\theta)) \neq 0$ for $\theta \in S^1$; then,*

$$\lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_A |\nabla\Psi(x)| Z(|\Psi(x)|/\epsilon) \, \mathrm{d}x = \textit{length of the curve} \tag{31}$$

where $Z : \mathbb{R}^+ \to \mathbb{R}^+$ is a positive continuous decreasing functions s.t. $Z(0) = \int_0^\infty Z = 1$ and $\lim_{x \to \infty} xZ(x) = 0$ (for example $Z(x) = \exp(-x)$); the integral on the left hand side is on a compact set A such that the curve is contained in $\mathring{A}$, the interior of A.

Fix a continuous $H : \mathbb{R}^2 \to \mathbb{R}$, a kernel $K : \mathbb{R}^+ \to \mathbb{R}$, and suppose that $\lim_{x \to \infty} K(x)/x = 1$. Let $\epsilon > 0$, and $c(\bar{s})$ be a point in the curve and let $u : \mathbb{R}^2 \to \mathbb{R}$ be the viscosity solution of

$$|\nabla u(x)| - (1 + |\Psi(x)|/\epsilon) = 0, \quad u(c(\hat{s})) = 0 \tag{32}$$

then

$$\lim_{\epsilon\to 0}\frac{1}{2\epsilon^2}\int_A |\nabla\Psi(x)|\,H(x)\,K(u(x))\,Z(|\Psi(x)|/\epsilon^2)\,\mathrm{d}x = \int_c H(c(s))\,K(|s-\hat{s}|)\,\mathrm{d}s \tag{33}$$

where $|s-\hat{s}|$ is the shortest arclength distance between $c(s)$ and $c(\hat{s})$ along c.

To solve the eikonal equation in (32) numerically, we use the fast marching method (Adalsteinsson and Sethian, 1995). We discretize on a $N\times N$ grid, then the integrals above are numerically substituted by a sum; if $\epsilon\to 0$ while N is kept constant, then clearly the numerical versions of (31) and (33) would converge to zero rather than to the desired result; so we propose the following numerical scheme for the approximation of (31):

$$\frac{1}{2\epsilon N^2}\sum_{i,j}^{N}|\nabla\Psi(x_{i,j})|Z(|\Psi(x_{i,j})|/\epsilon) \approx \text{length of the curve} \tag{34}$$

where $x_{i,j}$ are points of the discretization of the domain of Ψ to an $N\times N$ grid, and $\epsilon\to 0$, $N\to+\infty$ and $N\epsilon\to+\infty$. Similarly, we use the method by Adalsteinsson and Sethian (1995) to solve for a numerical approximation $u^N_{i,j}$ to (32); then (33) becomes

$$\frac{1}{2\epsilon^2N^2}\sum_{i,j}^{N}|\nabla\Psi(x_{i,j})|H(x_{i,j})K\left(u^N_{i,j}\right)Z(|\Psi(x_{i,j})|/\epsilon^2)\,\mathrm{d}x \approx \int_c H(c(s))\,K(|s-\bar{s}|)\,\mathrm{d}s. \tag{35}$$

6. Experiments

In this section, we show some simulations of Sobolev active contours used for segmentation of static images and some simple image sequences.

We tried both the method of Section 5.2 and 5.3; the latter is simpler to implement, but is difficult to tune w.r.t. the choice of ϵ and N. The following numerical experiments use the polygon(s) estimate of the zero level set, as in Section 5.2.

In all the simulations done below, the results for the Sobolev active contours are done with the $\tilde{H}^1$ inner product ($\lambda = 10$ unless stated otherwise, although a wide range of λ gives similar results). Using the H^1 inner product gives visually similar results as to what are shown. We consider two energies to illustrate the advantages of Sobolev active contours over H^0 active contours. The edge-based energy we consider is

$$E_e(c)=\int_c \phi(c(s))\,\mathrm{d}s,\ \text{where}\ \phi=\frac{1}{1+\|\nabla I\|^2}, \tag{36}$$

which was proposed by Caselles et al., Kichenassamy et al. (1995, 1995). The region-based energy we consider is

$$E_r(c)=\int_{c_{in}}(I-u)^2\,\mathrm{d}A+\int_{c_{out}}(I-v)^2\,\mathrm{d}A+\alpha L(c), \tag{37}$$

where

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

and $\alpha\geq 0$ specifies a penalty on the length, $L(\cdot)$, of the curve. This is the piecewise-constant model of Mumford and Shah (1989) (see also Chan and Vese (2001)). The use of the length penalty is partly to keep the evolving H^0 contour smooth, and avoid undesirable local minimum of the first terms caused by noise. This term has a smoothing effect since the H^0 gradient flow of L is curvature flow, which has smoothing properties. As we saw in Section 4.1, the length penalty for the Sobolev flow is futile in terms of giving smoothness to the contour.

6.1. Merging and Splitting

In Fig. 3, we demonstrate the experimental evidence for the ideas presented in Section 5.1. In this experiment, we segment an image using the region-based energy (37), and a Sobolev active contour. We see that the active contour can change topology to achieve the global minimum.

6.2. Noisy Square Segmentation

In Fig. 4, we show the results of an experiment in which we segment a noisy square image with the region-based

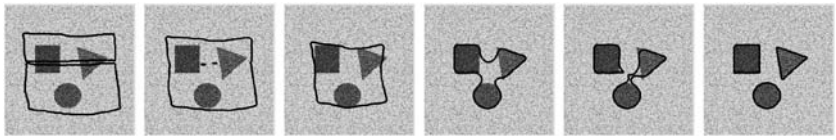
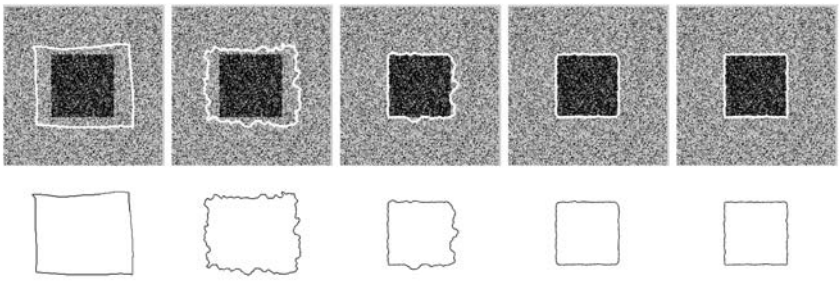
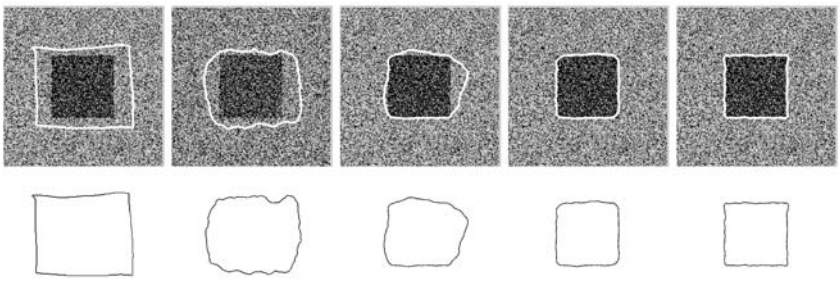


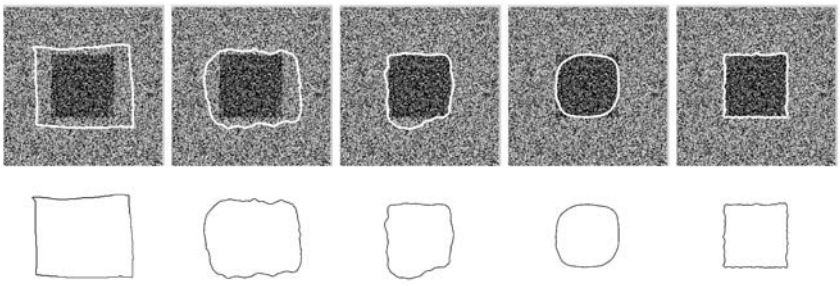
Figure 3. Segmentation of various shapes using a Sobolev active contour for a region-based energy (37). This illustrates the ability of Sobolev active contours to cope with topological changes: merging and splitting.



(a) $\rho = 0.5$. Left to right: Initialization, $H^0 : \alpha = 20\,000, 50\,000, 90\,000$, Sobolev.



(b) $\rho = 0.6$. Left to right: Initialization, $H^0 : \alpha = 50\,000, 120\,000, 160\,000$, Sobolev.

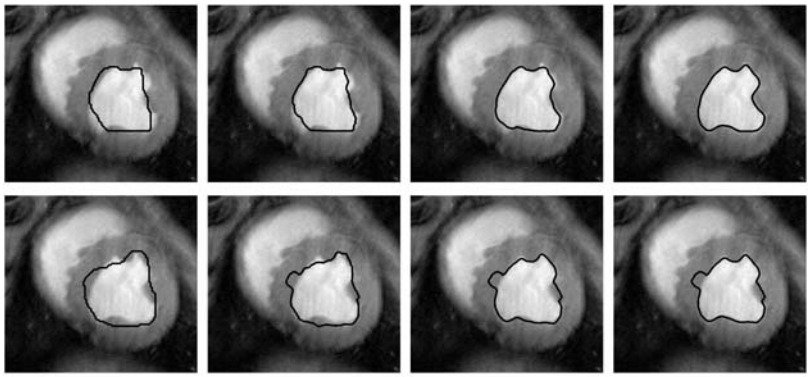


(c) $\rho = 0.7$. Left to right: Initialization, $H^0 : \alpha = 50\,000, 100\,000, 300\,000$, Sobolev.

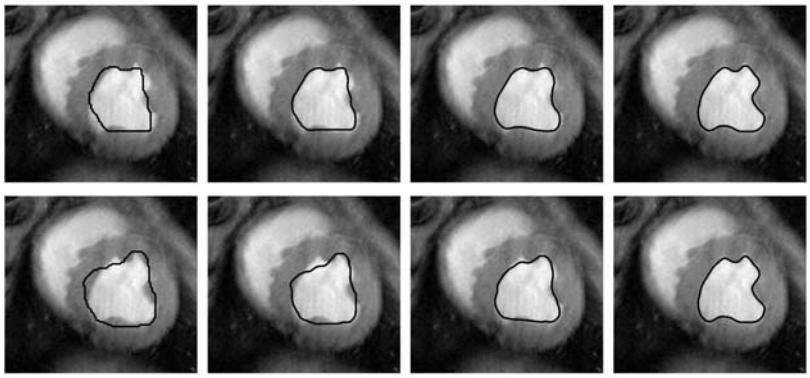
Figure 4. Segmentation of square binary image with salt and pepper noise of various densities. The experiment shows the results with H^0 (of various degrees of regularization, α), and the Sobolev active contour.

energy (37). We added salt and pepper noise with densities of $\rho = 0.5, 0.6, 0.7$ to the binary image. In this experiment, we compare the results obtained from using the usual H^0 active contour with the result obtained from using a Sobolev active contour.

First, we explore the effects of various weights, α, on the length penalty in (37) for the H^0 active contour. In Fig. 4, we see that with using a small α, the contour becomes stuck in the noise, at an intermediate local minimum of E_r (37). The image second from the right



(a) H^0 active contour.



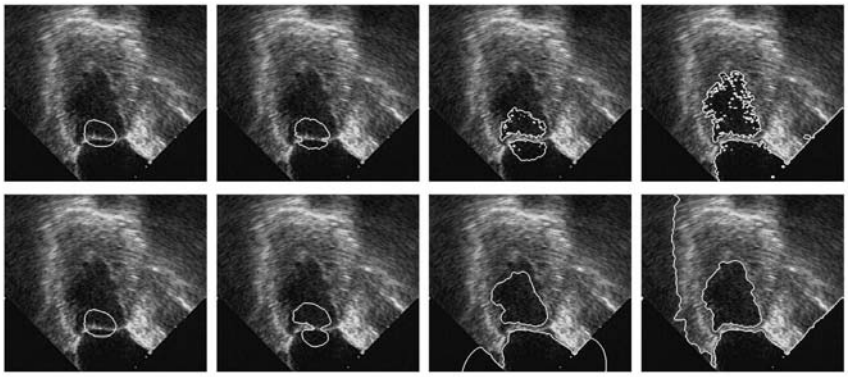
(b) Sobolev active contour.

Figure 5. Segmentation of a cardiac image using the edge-based energy in (36) (two different initializations).

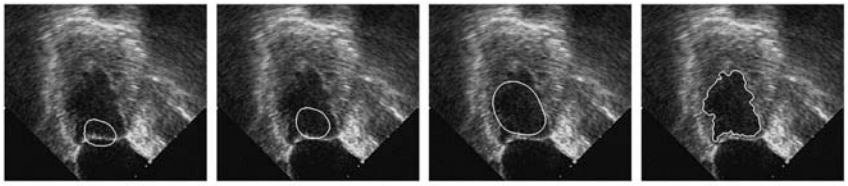
in each row is the result of the H^0-active contour with a minimum α just high enough so that the contour is not stuck in noise at a local minimum. In Fig. 4(a), we see that with relatively low noise, the result of the H^0-active contour with just high enough α to overcome the local minimum captures the desired square accurately. As the noise increases, a higher α is needed to overcome local minima; however, because of such a high length penalty and therefore higher smoothness, the active contour is unable to capture the fine-scale structure of the desired object. In Fig. 4(c), we see that using the minimal α to overcome the noise results in the contour failing to capture the corners of the desired square object.

Notice that for each of the noise levels, the Sobolev active contour without any length penalty ($\alpha = 0$) or additional regularization terms in the energy moves in such a way as to avoid any local minimum of the energy, E_r, as it moves in a global fashion first before resorting to finer scale deformations (the final segmentation of the square images are the last image in each row of Fig. 4). The end result captures the square accurately and looks to be the global minimum of the region-based energy. The result is independent of λ, the weighting on the derivative component of the $\tilde{H}^1$ inner product since translations do not optimize the energy. Notice that the converged Sobolev contour becomes more rugged as the noise increases. This is because the image is corrupted by the noise, and the original boundary of the square is no longer the global minimum of energy. We should point out that the Sobolev active contour gives smoothness in the contour *flow*; it does not guarantee smoothness in the final contour, which is determined by the energy that is being minimized.

Besides obtaining a more accurate segmentation, the Sobolev active contour converges much faster than the H^0 active contour with just high enough length penalty, α, to get passed local minima. The reason for this is



(a) H^0 active contour with low ($\alpha = 10$, top) and with high ($\alpha = 5000$) regularization.



(b) Sobolev active contour (no regularization).

Figure 6. Segmentation of an ultrasound image using the region-based energy (37).

because of the small step size needed for stability of the curvature term arising from the length penalty in the level set implementation. A crude estimate for the Sobolev active contour shows that the step size is less than $0.5/255^2 = 130\,050^{-1}$, whereas for the H^0-active contour with $\alpha = 300\,000$, as in Fig. 4(c), a step size of less than $\min\{0.5/300\,000 = 600\,000^{-1}, 0.5/255^2 = 130\,050^{-1}\} = 600\,000^{-1}$ is needed. This is a factor of about 5 for this crude estimate.

6.3. *Segmentation of Real Images*

We illustrate the advantages of using Sobolev active contours over H^0 with the same energy on real medical images in Figs. 5, 6, and 7.

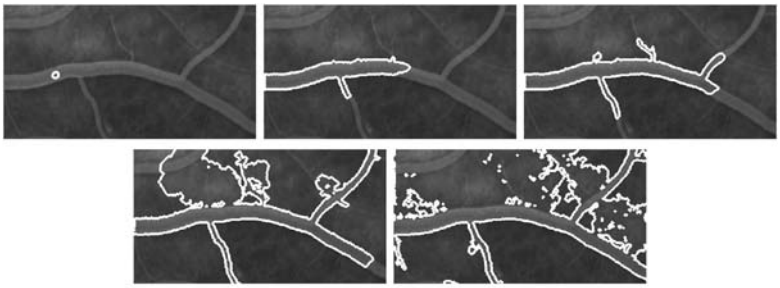
In Fig. 5, we use the edge-based energy, E_e (36) to segment a cardiac image. We show the results with two different initializations using both H^0 and Sobolev active contours. Because of the complicated texture of the image, the edge-based energy is riddled with local minima. Because the Sobolev active contour moves more globally, it is less likely than the H^0 active contour to become stuck in local minima, as shown in Fig. 5.

In Fig. 6, we segment an ultrasound image using the region-based energy (37). In Fig. 6(a), the results are shown using an H^0 active contour with two different weightings, α, on the length penalty. In both cases, extraneous features of the image are detected. In Fig. 6(b), we see that the Sobolev contour moves initially according to global motions (translation and dilation), and finally detects the more fine features of the image when more global motions cannot optimize the energy. Thus, the contour is able to avoid irrelevant local features that disturbs the H^0 active contour.

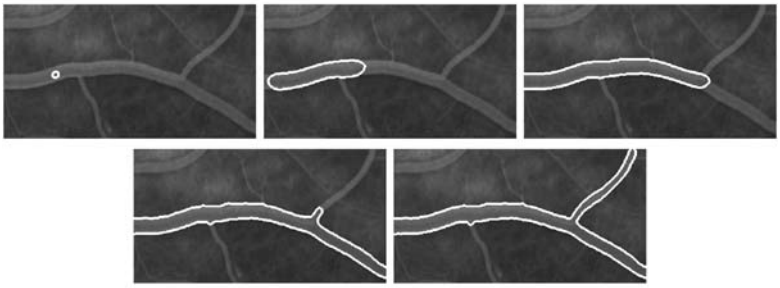
In Fig. 7, we segment a vessel image using a the region-based energy (37). As we see, the Sobolev active contour is less affected by local features, which causes the H^0 active contour to leak into an irrelevant region of the image.

6.4. *Segmentation of Simple Image Sequences*

We now illustrate the robustness of the Sobolev active contour for segmenting simple synthetic image sequences (Figs. 8 and 9). The image sequences are simply formed by translating a square object. We successively segment frames of the image by an active contour whose initialization in the current frame is the final segmentation in the previous frame. In the experiment shown in Fig. 8, we employ the edge-based energy (36). In the experiment in Fig. 9, we use the region-based energy (37). The segmentation evolutions



(a) H^0 active contour with length penalty.



(b) Sobolev active contour (no length penalty).

Figure 7. Segmentation of a vessel image using the region-based energy (37) with both the H^0 active contour and the Sobolev active contour. The Sobolev active contour is able to avoid distracting fine features of the image and therefore does not leak into an irrelevant region of the image.

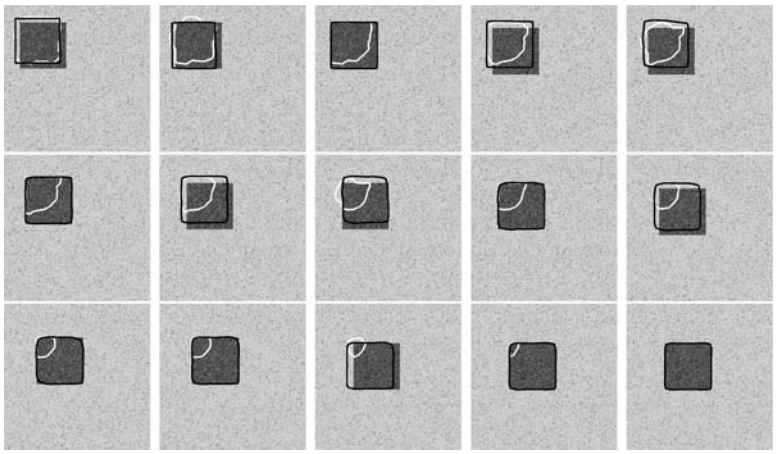


Figure 8. Left to right, top to bottom: Tracking a moving square in a noisy environment (Gaussian noise, $\mu = 0, \sigma^2 = 0.01$) using H^0 active contour (white) and Sobolev active contour (black). The edge-based energy (36) is used.

are run until convergence of both contours (H^0 in white and Sobolev in black). Notice the H^0 active contours becomes stuck in a undesirable local minima after the initial movement of the object and soon lose track of the object. The Sobolev active contour does not have this problem and it successfully tracks the object by mostly translating and deforming only slightly.

7. Conclusion

In summary, we have observed that much of the literature on active contours uses the concept of *gradient flow* to minimize energies, but it has always been assumed (knowingly or unknowingly) that the inner product on curve perturbations, on which the gradient

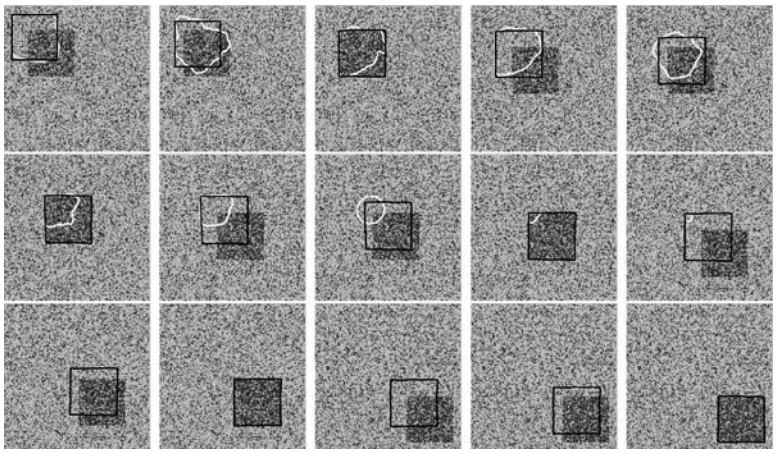


Figure 9. Left to right, top to bottom: Tracking a moving square in a noisy environment (Gaussian noise, $\mu = 0, \sigma^2 = 0.3$) using the region-based energy (37) and the H^0 active contour (white) with $\alpha = 10000$ and Sobolev active contour (black) using no regularization ($\alpha = 0$).

depends, is the H^0 inner product. We have introduced using Sobolev inner products on the set of perturbations of a curve. We have demonstrated the general methodology for computing Sobolev gradients, which requires integrating the H^0 gradient. The procedure requires very little extra computational time and little change of existing computer code when compared with the H^0 gradient flow, in particular for the $\tilde{H}^n$ family of norms, as remarked in 3.1. It was demonstrated by theory and experiments that Sobolev gradient flows are global flows, in which a single point on the curve depends on all other points of the curve, and the flows deform locally after global motions (e.g., translations) can no longer optimize the energy. This particular property shows that the Sobolev method gives a smooth *flow* (not necessarily a smooth contour), which in many cases helps avoid certain undesirable local minimum of active contour energies that disturb the (local) H^0 flows. Explicit formulas for Sobolev gradient flows of typical energies found in the active contour literature, which were derived, showed many interesting properties of Sobolev active contours. One of the notable properties that these explicit formulas shows is that the Sobolev method regularizes the corresponding H^0 gradient flow by reducing the order of the PDE. In particular, derivatives of the curve need not be defined for region-based and edge-based energies, and the elastic energy, which results in a fourth order H^0 flow is reduced to a second order flow using a first order Sobolev flow.

Acknowledgments

The authors would like to acknowledge the support of the following grants: NSF CCR-0133736, NIH/NINDS R01-NS-037747 and an Airforce MURI grant.

References

Adalsteinsson, D. and Sethian, J. 1995. A fast level set method for propagating interfaces. *J. Comp. Phys.*, 118:269–277.

Blake, A. and Isard, M. 1998, *Active Contours*. Springer Verlag.

Burger, M. 2003. A framework for the construction of level set methods for shape optimization and reconstruction. *Interfaces Free Boundaries*, 5:301–329.

Burger, M. and Osher, S. 2005. A survey on level set methods for inverse problems and optimal design. *Eur. J. Appl. Math.*, 16.

Caselles, V., Catte, F., Coll, T., and Dibos, F. 1993. A geometric model for edge detection. *Num. Mathematik*, 66:1–31.

Caselles, V., Kimmel, R., and Sapiro, G. 1995. Geodesic active contours. In: *Proc. of the IEEE Int. Conf. on Computer Vision*, Cambridge, MA, USA, pp. 694–699.

Chan, T. and Vese, L. 2001. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277.

Charpiat, G., Faugeras, O.D., and Keriven, R. 2005a. Approximations of shape metrics and application to shape warping and empirical shape statistics. *Foundations Comput. Math.*, 5(1):1–58.

Charpiat, G., Keriven, R., Pons, J., and Faugeras, O. 2005b. Designing spatially coherent minimizing flows for variational problems base d on active contours. In: *ICCV*.

Chen, Y., Tagare, H., Thiruvenkadam, S., Huang, F., Wilson, D., Gopinath, K., Briggs, R., and Geiser, E. 2002. Using prior shapes in geometric active contours in a variational framework. *Int. J. Comput. Vision*, 50(3):315–328.

Chopp, D.L. and Sethian, J.A. 1999. Motion by intrinsic laplacian of curvature. *Interfaces Free Boundaries*, 1:107–123.

Cohen, L.D. 1991. On active contour models and ballons. *Comput. Vision, Graphics, and Image Processing: Image Processing*, 53(2).

Cremers, D. and Soatto, S. 2003. A pseuso distance for shape priors in level set segmentation. In: *IEEE Int. Workshop on Variational, Geometric and Level Set Methods*, pp. 169–176.

do Carmo, M. 1992. *Riemannian Geometry*. Birkhäuser Boston.

Droske, M. and Rumpf, M. 2004. A level set formulation for the willmore flow. *Interfaces and Boundaries*, 6(3):361–378.

Hintermüller, M. and Ring, W. 2004. An inexact Newton-CG-type active contour approach for the minimization of the mumford-shah functional. *J. Math. Imaging Vision*, 20(1):19–42.

Kass, M., Witkin, A., and Terzopoulos, D. 1987. Snakes: Active contour models. *Int. J. Comput. Vision*, 1:321–331.

Kichenassamy, S., Kumar, A., Olver, P., Tannenbaum, A., and Yezzi, A. 1995. Gradient flows and geometric active contour models. In: *Proc. of the IEEE Int. Conf. on Comput. Vision*, pp. 810–815.

Lang, S. 1999. *Fundamentals of Differential Geometry*. Springer-Verlag.

Leventon, M., Grimson, E., and Faugeras, O. 2000. Statistical Shape influence in geodesic active contours. In: *IEEE Conf. on Comp. Vision and Patt. Recog.*, vol. 1, pp. 316–323.

Malladi, R., Sethian, J., and Vemuri, B. 1995. Shape modeling with front propagation: a level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:158–175.

Mansouri, A.-R., Mukherjee, D.P., and Acton, S.T. 2004. Constraining active contour evolution via Lie Groups of transformation. *IEEE Transactions on Image Processing*, 13(6):853–863.

Mennucci, A.C.G., Yezzi, A., and Sundaramoorthi, G. 2006. Sobolev–type metrics in the space of curves. *Preprint, arXiv:math.DG/0605017*.

Michor, P. and Mumford, D. 2003. Riemannian geometries on the space of plane curves. *ESI Preprint 1425, arXiv:math.DG/0312384*.

Mio, W. and Srivastava, A. 2004. Elastic-string models for representation and analysis of planar shapes. In: *CVPR*, vol. 2, pp. 10–15.

Mumford, D. and Shah, J. 1985. Boundary detection by minimizing functionals. In: *Proc. IEEE Conf. Computer Vision Pattern Recognition*.

Mumford, D. and Shah, J. 1989. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685.

Neuberger, J.W. 1997. *Sobolev Gradients and Differential Equations*. Lecture Notes in Mathematics #1670. Springer.

Osher, S. and Sethian, J. 1988. Fronts propagating with curvature-dependent speed: algorithms based on the Hamilton-Jacobi equations. *J. Comp. Phys.*, 79:12–49.

Paragios, N. and Deriche, R. 2002a. Geodesic active regions: A new paradigm to deal with frame partition problems in computer vision. *International Journal of Visual Communication and Image Representation, Special Issue on Partial Differential Equations in Image Processing, Computer Vision and Computer Graphics*, 13(2):249–268.

Paragios, N. and Deriche, R. 2002b. Geodesic active regions and level set methods for supervised texture segmentation. *Int. J. Comput. Vision*, 46(3):223.

Raviv, T.R., Kiryati, N., and Sochen, N. 2004. Unlevel-set: Geometry and prior-based segmentation. In: *Proc. European Conf. on Computer Vision*.

Ronfard, R. 1994. Region based strategies for active contour models. *Int. J. Comput. Vision*, 13(2):229–251.

Rousson, M. and Paragios, N. 2002. Shape Priors for Level Set Representations. In: *Proc. European Conf. Computer Vision*, vol. 2, pp. 78–93.

Rouy, E. and Tourin, A. 1992. A viscosity solutions approach to shape-from-shading. *SIAM J. Numerical Anal.*, 29(3):867–884.

Siddiqi, K., Lauzière, Y.B., Tannenbaum, A., and Zucker, S. 1998. Area and length minimizing flows for shape segmentation. *IEEE Transactions on Image Processing*, 3(7):433–443.

Soatto, S. and Yezzi, A.J. 2002. DEFORMOTION: Deforming motion, shape average and the joint registration and segmentation of images. In: *ECCV*, vol. 3, pp. 32–57.

Sundaramoorthi, G., Yezzi, A., and Mennucci, A. 2005. Sobolev active contours. In: *VLSM*, pp. 109–120.

Tsai, A., Yezzi, A., and Willsky, A.S. 2001a. Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. *IEEE Transactions on Image Processing*, 10(8):1169–1186.

Tsai, A., Yezzi, A.J., III, W.M.W., Tempany, C., Tucker, D., Fan, A., Grimson, W.E.L., and Willsky, A.S. 2001b. Model-based curve evolution technique for image segmentation. In: *CVPR*, vol. 1, pp. 463–468.

Vese, L.A. and Chan, T.F. 2002. A multiphase level set framework for image segmentation using the mumford and shah model. *Int. J. Comput. Vision*, 50(3):271–293.

Xu, C. and Prince, J.L. 1998. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7(3):359–369.

Yezzi, A. and Mennucci, A. 2005a. Metrics in the space of curves. *Preprint, arXiv:math.DG /0412454*.

Yezzi, A., Tsai, A., and Willsky, A. 1999. A statistical approach to snakes for bimodal and trimodal imagery. In: *Int. Conf. on Comput. Vision*, pp. 898–903.

Yezzi, A.J. and Mennucci, A. 2005b. Conformal metrics and true "Gradient flows" for curves. In: *ICCV*, pp. 913–919.

Younes, L. 1998. Computable elastic distances between shapes. *SIAM J. Appl. Math.*, 58(2):565–586.

Zhu, S.C., Lee, T.S., and Yuille, A.L. 1995. Region competition: Unifying snakes, region growing, Energy/Bayes/MDL for multi-band image segmentation. In: *ICCV*, pp. 416–423.