Sobolev Active Contours

Ganesh Sundaramoorthi[1], Anthony Yezzi[1], and Andrea Mennucci[2]

[1] School of Electrical Engineering, Georgia Institute of Technology, Atlanta, USA
[2] Department of Mathematics, Scuola Normale Superiore, Pisa, Italy

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. This is 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 and tracking applications.

1 Introduction

Active contours, introduced by Kass *et al.* [1], have been widely used for the segmentation problem. One undesirable feature of Kass's model is that the energy used to derive a flow is dependent on parametrization. Formulations for geometric energies, which do not depend on the particular parametrization of the curve, were later introduced for edge-based active contours [2,3] and region-based active contours [4,5,6]. In order to define the notion of gradient of such energies, an inner product on the set of perturbations of a curve is needed. 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 a gradient. However, recent work in [7,8] has shown that the Riemannian metric on the space of curves induced by the H^0 inner product is pathological.

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

- First, 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

N. Paragios et al. (Eds.): VLSM 2005, LNCS 3752, pp. 109–120, 2005.

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 unsmooth instantly. Therefore, in geometric active contours models, a penalty on the curve's length is added to keep the curve smooth in addition to keeping the variational problem well-posed. However, this changes the energy that is being optimized.

- Second, H^0 gradients, evaluated at a particular point on the curve, depend locally on derivatives of the curve. Therefore, as the curve becomes unsmooth, 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 means, the H^0 gradient still depends on local image data. 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 collectively. This restricts the gradient at a particular point from "seeing" information located at other points of the curve, which implies sensitivity to noise and local features.
- Third, all geometric 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, a H^0 active contour does not remain smooth, one needs viscosity theory to define the flow.
- Fourth, 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 [9] cannot be used. This forces one to use particle methods to implement the flow. However, flows with higher than two derivatives are generally difficult to implement because of numerical artifacts.
- Lastly, as a specific example, the gradient ascent for arclength, i.e., backward heat flow, is not stable. 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 is unstable.

In this paper, we consider using inner products arising from Sobolev spaces to define gradients. Note that a first order Sobolev-like inner product defined on an equivalence class with respect to a group has been used in the context of shape analysis [10], but not for defining gradient flows.

2 General Theory

2.1 Structure on the Space of Curves

Let $\mathcal{C}$ denote the set of smooth embedded curves in $\mathbb{R}^2$, which is a differentiable manifold [11]. For $C \in \mathcal{C}$, we denote by $T_C\mathcal{C}$ the tangent space of $\mathcal{C}$ 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 now define inner products on $T_C\mathcal{C}$.

Definition 1. *Let $C \in \mathcal{C}$, L be the length of C, and $h, k \in T_C\mathcal{C}$. Let $\lambda > 0$. We assume h and k are parametrized by the arclength parameter of C.*

1. $\langle h, k \rangle_{H^0} := \frac{1}{L} \int_0^L h(s) \cdot k(s) \mathrm{d}s$
2. $\langle h, k \rangle_{H^1} := \langle h, k \rangle_{H^0} + \lambda L^2 \langle h', k' \rangle_{H^0}$
3. $\langle h, k \rangle_{\tilde{H}^1} := \overline{h} \cdot \overline{k} + \lambda L^2 \langle h', k' \rangle_{H^0}$

where $\overline{h} := \frac{1}{L} \int_0^L h(s) \mathrm{d}s$, and the derivatives are with respect to arclength.

Note that we have introduced length dependent scale factors for convenience of later computations. It is easy to verify that the above definitions are inner products. One can easily generalize the previous definitions to $\tilde{H}^n$ by simply replacing the first derivative with the n^{th} derivative, and to H^n by adding in components of the form $\left\langle h^{(m)}, k^{(m)} \right\rangle_{H^0}$ where $m \leq n$. Also, other versions of H^1 and $\tilde{H}^1$are possible [8]. We now define the notion of gradient of a functional $E : \mathcal{C} \to \mathbb{R}$.

Definition 2. *Let $E : \mathcal{C} \to \mathbb{R}$.*

1. *If $C \in \mathcal{C}$ and $h \in T_C\mathcal{C}$, 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 $\theta \in S^1$.*
2. *Assume $\langle , \rangle_C$ is an inner product on $T_C\mathcal{C}$. The* gradient of E *is a vector field $\nabla E(C) \in T_C\mathcal{C}$ that satisfies $\mathrm{d}E(C) \cdot h = \langle h, \nabla E(C) \rangle_C$ for all $h \in T_C\mathcal{C}$.*

For each $C \in \mathcal{C}$, note that $\mathrm{d}E(C)$ is a linear operator on $T_C\mathcal{C}$. If $\mathrm{d}E(C)$ is bounded, then the notion of operator norm can be defined. The operator norm of $\mathrm{d}E(C)$ with respect to an inner product $\langle , \rangle_C$, which induces a norm $\| \cdot \|_C$, is

$$\|\mathrm{d}E(C)\| = \sup_{h \in T_C\mathcal{C} \setminus \{0\}} \frac{|\mathrm{d}E(C) \cdot h|}{\|h\|_C}. \tag{1}$$

If the gradient of E exists, then by the Cauchy-Schwartz inequality, we have that $h = \nabla E(C)$ attains the supremum on the right hand side of (1). Note for $\lambda \to +\infty$, translations have norm approaching zero with respect to the norm induced by H^1 and $\tilde{H}^1$. In light of the interpretation of the gradient as the perturbation that attains the supremum in (1), it follows that translations are favored for gradients in H^1 and $\tilde{H}^1$ as $\lambda \to +\infty$ if they reduce energy.

2.2 Relation Between H^1 and $\tilde{H}^1$

We show that the norms associated with the inner products H^1 and $\tilde{H}^1$, *i.e.*,

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

are equivalent.

We first derive a simple Poincaré inequality: from $h(u) - h(v) = \int_v^u h'(s)\mathrm{d}s$ we derive that $\sup_u |h(u) - \overline{h}| \leq \int_0^L |h'(s)|\mathrm{d}s$, and then

$$\sqrt{\int_0^L |h(s) - \overline{h}|^2 \mathrm{d}s} \leq \sqrt{L} \sup_u |h(u) - \overline{h}| \leq \sqrt{L} \int_0^L |h'(s)|\mathrm{d}s \leq L\sqrt{\int_0^L |h'(s)|^2 \mathrm{d}s},$$

which is the Poincaré inequality.

We now prove the equivalence of the two norms. By Hölder's inequality, we have that $|\overline{h}|^2 \leq \frac{1}{L}\int_0^L |h(s)|^2 \mathrm{d}s$ so that $\|h\|_{\tilde{H}^1} \leq \|h\|_{H^1}$. On the other hand, note that, $\frac{1}{L}\int_0^L |h(s) - \overline{h}|^2 \mathrm{d}s = \frac{1}{L}\int_0^L |h(s)|^2 \mathrm{d}s - |\overline{h}|^2$, so that

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

where $\vee$ denotes maximum. Note that we have not established any relation between the geometry of the inner products H^1 and $\tilde{H}^1$; however, in the next sections, we show that the gradients from H^1 and $\tilde{H}^1$ have similar properties.

2.3 Comment on H^n for $n \geq 2$

As alluded to in Section 2.1, translations are favored for H^1 and $\tilde{H}^1$ 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 of using higher order Sobolev gradients. However, one gains added regularity of the gradient flow in using higher order Sobolev gradients.

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 $h \in T_C\mathcal{C}$,

$$\mathrm{d}E(C) \cdot h = \langle h, g \rangle_{H^0} + \lambda L^2 \langle h', g' \rangle_{H^0} = \left\langle h, g - \lambda L^2 g'' \right\rangle_{H^0}$$

where we have integrated by parts and noted that we have periodic boundary conditions. Since gradients are unique (if they exist), in particular, the H^0 is unique, we must have that

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

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

$$\mathrm{d}E(C) \cdot h = \overline{h} \cdot \overline{g} + \lambda L^2 \left\langle h', g' \right\rangle_{H^0} = \left\langle h, \overline{g} - \lambda L^2 g'' \right\rangle_{H^0}.$$

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

$$f(s) = \overline{f} - \lambda L^2 g''(s) \quad \text{where } s \in [0, L] \tag{3}$$

and we have periodic boundary conditions.

3.1 Solving the ODEs

We want to solve first the ODE (2) for g. It suffices to solve (2) 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 $s, \hat{s} \in (0, L)$

$$k_\lambda(s, \hat{s}) - \lambda L^2 \frac{\partial^2 k_\lambda}{\partial s^2}(s, \hat{s}) = \delta(s - \hat{s}) \tag{4a}$$

$$k_\lambda(0, \hat{s}) = k_\lambda(L, \hat{s}); \;\; \partial_s k_\lambda(0, \hat{s}) = \partial_s k_\lambda(L, \hat{s}); \;\; k_\lambda(\hat{s}+, \hat{s}) = k_\lambda(\hat{s}-, \hat{s}), \tag{4b}$$

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

$$K_\lambda(s) = \frac{\cosh\left(\frac{s - \frac{L}{2}}{\sqrt{\lambda} L}\right)}{2L\sqrt{\lambda} \sinh\left(\frac{1}{2\sqrt{\lambda}}\right)}, \quad \text{for } s \in [0, L], \tag{5}$$

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

$$\nabla_{H^1} E(s) = \int_C K_\lambda(\hat{s} - s) \nabla_{H^0} E(\hat{s}) \mathrm{d}\hat{s} = (K_\lambda * \nabla_{H^0} E)(s) \tag{6}$$

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 (3). It suffices to solve (3) with the boundary conditions $g(0) = g(L)$, $g'(0) = g'(L)$, and the relation $\overline{f} = \overline{g}$. Integrating (3) twice yields

$$g(s) = g(0) + s g'(0) - \frac{1}{\lambda L^2} \int_0^s (s - \hat{s})(f(\hat{s}) - \overline{f}) \mathrm{d}\hat{s}. \tag{7}$$

Using (7), applying the boundary conditions, and noting that $\overline{g} = \overline{f}$, after some manipulation, yields

$$g'(0) = -\frac{1}{\lambda L^3}\int_0^L s(f(s) - \overline{f})\mathrm{d}s \quad \text{and} \quad g(0) = \int_0^L f(s)\tilde{K}_\lambda(s)\mathrm{d}s \tag{8}$$

where the kernel function $\tilde{K}_\lambda$ is given by

$$\tilde{K}_\lambda(s) = \frac{1}{L}\left(1 + \frac{(s/L)^2 - (s/L) + 1/6}{2\lambda}\right), \quad s \in [0, L]. \tag{9}$$

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. Now if we shift the arclength parameterization of the curve, we obtain a convolution formula for g at any point s. Therefore,

$$\nabla_{\tilde{H}^1} E(s) = \int_C \tilde{K}_\lambda(\hat{s} - s)\nabla_{H^0} E(\hat{s})\mathrm{d}\hat{s} = (\tilde{K}_\lambda * \nabla_{H^0} E)(s). \tag{10}$$

3.2 Properties of the Kernels

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

$$K_\lambda''(s) = \frac{1}{\lambda L^2}\left(K_\lambda - \delta(s)\right) \text{ and } \tilde{K}_\lambda''(s) = \frac{1}{\lambda L^2}\left(\frac{1}{L} - \delta(s)\right), \quad s \in [0, L). \tag{11}$$

The first property is just the relation in (4a), 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 (2) and (3), respectively. Next, note that

$$\int_C K_\lambda(\hat{s})\mathrm{d}\hat{s} = 1 \quad \text{and} \quad \int_C \tilde{K}_\lambda(\hat{s})\mathrm{d}\hat{s} = 1 \tag{12}$$

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 (6) and (10), that the H^1 and $\tilde{H}^1$ gradients are geometric, *i.e.*, they do not depend on a particular parametrization chosen for the curve. This is also evident from the definition of these inner products. The formulas (6) and (10) 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.

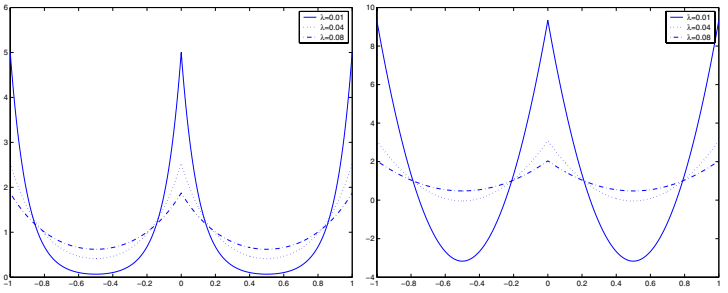

Fig. 1. Plots of K_λ (left) and $\tilde{K}_\lambda$ (right) for various λ with $L = 1$. The plots show the kernels over two periods.

Because H^1 and $\tilde{H}^1$ gradients are given by integrals of H^0, given in formulas (6) and (10), integration by parts and the relations in (11) 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 (12) implies that the H^1 gradient can be interpreted as a weighted average of the H^0 gradient; the 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 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.

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 since the formulas (7), (8) give the $\tilde{H}^1$ gradient as a single integral without convolution, as opposed to the necessary convolution for H^1. 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 (9) 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 $\overline{\nabla_{H^0} E}$, which is a just a translation. The other component is a complex deformation. An algorithm to implement 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 [12].

4 Some Sobolev Gradient Flows

In this section, we simplify the formulas (6) and (10) for some common geometric energies, note some interesting properties, and compare these with the usual H^0 gradients. In what follows, we use K to denote either the kernel (5) or (9), 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 [2,3]. 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 $\nabla_{H^0} E = L(\nabla\phi \cdot \mathcal{N})\mathcal{N} - L\phi\kappa\mathcal{N}$ where $\mathcal{N}$ is the unit inward normal, and κ is the curvature. Let us first note that $\nabla_{H^0} E = L\nabla\phi - L(\phi C')'$. Integrating by parts we find that

$$\frac{1}{L}\nabla_1 E = \nabla\phi * K - (\phi C')' * K = \nabla\phi * K - (\phi_s C) * K' - (\phi C) * K'',$$

where $\phi_s(\hat{s}) := \mathrm{d}/\mathrm{d}\hat{s}\phi(C(\hat{s}))$. Using the relations in (11), we find that

$$\nabla_{\tilde{H}^1} E = \frac{\phi C - \overline{\phi C}}{\lambda L} - L(\phi_s C) * \tilde{K}'_\lambda + L\nabla\phi * \tilde{K}_\lambda. \tag{13}$$

Of particular interest is when $\phi = 1$, that is $E = L$, the length of the curve. We see that $\nabla_{\tilde{H}^1} L = \frac{C - \overline{C}}{\lambda L}$. 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 [5,6]. 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. The gradient with respect to H^0 is $\nabla_{H^0} E = -L\phi\mathcal{N} = -L\phi JC'$ and J is a rotation by 90^o matrix. Integrating by parts we find that

$$\frac{1}{L}\nabla_1 E = -(\phi JC') * K = (\phi_s JC) * K + (\phi JC) * K'. \tag{14}$$

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

$$\nabla_{\tilde{H}^1} E = \frac{J}{\lambda L^2} \int_0^L \Big(\phi C(\cdot + \hat{s}) - \overline{\phi C}\Big)\hat{s}\,\mathrm{d}\hat{s} + (\phi_s JC) * \tilde{K}_\lambda. \tag{15}$$

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

$$C_t(s) = \pm\frac{J}{\lambda L^2} \int_0^L \Big(C(s + \hat{s}) - \overline{C}\Big)\hat{s}\mathrm{d}\hat{s} \tag{16}$$

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

4.3 Elastic Energy

Consider the geometric version of elastic energy defined by $E(C) = \int_C \kappa^2 \mathrm{d}s = \int_C \|C''\|^2 \mathrm{d}s$. It can be shown that the H^0 gradient is $\nabla_{H^0} E = L(2C^{(3)} + 3(C'' \cdot C'')C')'$. Thus, we find that

$$\frac{1}{L}\nabla_1 E = \Big(2C^{(3)} + 3(C'' \cdot C'')C'\Big)' * K = 2C'' * K'' - 3(C'' \cdot C'')C' * K'.$$

For the kernel $\tilde{K}_\lambda$, this simplifies to

$$\frac{1}{L}\nabla_{\tilde{H}^1} E = -\frac{2C''}{\lambda L^2} - 3(C'' \cdot C'')C' * \tilde{K}'_\lambda, \tag{17}$$

and thus, the gradient descent, up to a scale factor depending on λ and L, is

$$C_t = C_{ss} + \frac{3}{2}\int_s^{s+L} (C_{ss} \cdot C_{ss})C_s \left(\frac{\hat{s} - s}{L} - \frac{1}{2}\right) \mathrm{d}\hat{s}. \tag{18}$$

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 $\tilde{H}^1$ (13), (14) 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, the Sobolev flows are defined for polygons, without the use of viscosity theory. Note that the expression in (13) does not require any more derivatives of ϕ than the expression for H^0 does. This is not the case for (14), 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 (6) and (10) may be used if a derivative of ϕ is not desired to be computed. Notice the expressions of Sobolev gradients for the elastic energy (17) 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 of elastic energy cannot be implemented using level set methods. Thus, a particle method must be used; however, this is prone to numerical problems.

5 Simulations

In this section, we show some simulations of Sobolev active contours used for segmentation and tracking. In all the simulations done below, the results for the Sobolev active contours are done with the H^1 inner product with $\lambda = 10$, although higher λ produces similar results with higher regularity. Using $\tilde{H}^1$ gives the same results as to what are shown.

Figure 2 shows snapshots of evolutions that use the edge-based energy [2,3] for both the H^0 (top) and the H^1 (bottom) active contours to segment a noisy image. The initial contour is a shifted version of the true object with a slightly different radius. Notice that the H^0 active contour learns local features instantly, and therefore becomes stuck at an undesirable local minimum. On the other hand, the Sobolev H^1 active contour moves according to a global motion first, then when it cannot reduce energy by moving in a global manner, it begins to learn local features. As a result, the Sobolev active contour overcomes any undesirable local minima created by the noise. Figure 3 shows a similar experiment

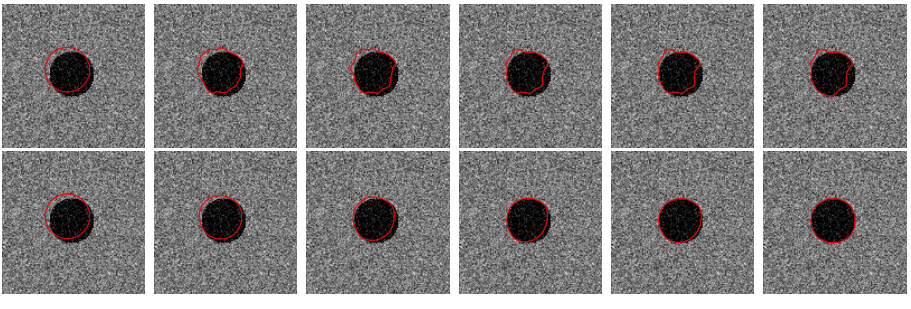

Fig. 2. Segmentation of a noisy image using edge-based H^0 active contour (top), and edge-based Sobolev active contour (bottom)

using the region-based energy [6]. The top row of the figure shows the result using the H^0 active contour. Note that a curvature term is added to keep the curve smooth. A curvature-data term ratio of 2000 to 1 is used for the H^0 active contour. Notice that although global statistics are used to define the energy, points on the H^0 active contour move independently without knowledge of other points on the curve. Thus, the curve becomes unsmooth instantly, being susceptible to local features, even though a high curvature weighting is used. The flow converges to an undesirable local minimum. On the contrary, the H^1 flow preserves the shape of the initial contour as it translates until translations are no longer favorable to reduce energy. Then the contour deforms from a square shape to a circular shape as fine scale features of the image are learned. Note that there was no curvature term added to the Sobolev active contour; the regularity is achieved solely through the inner product definition.

Figures 4 and 5 show examples of tracking a square that translates using both H^0 active contours (in red) and Sobolev active contours (in blue). The

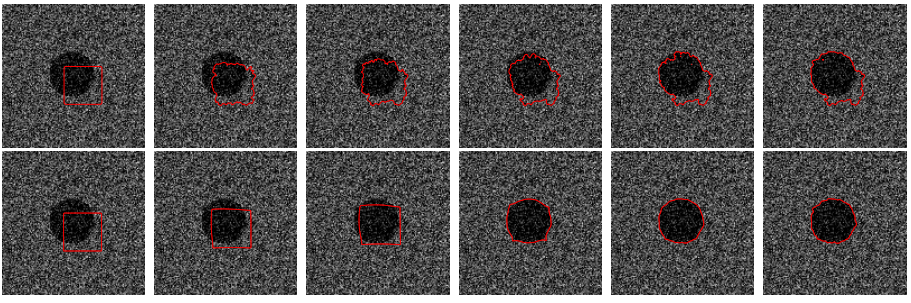

Fig. 3. Segmentation of a noisy image using region-based H^0 active contour with a length penalty (top), and region-based Sobolev active contour (bottom)

segmentation result from the previous frame is used as the initial contour for the next frame. The segmentation evolutions are run until convergence of both contours. The first example in Fig. 4 shows the result using the edge-based energy. The second example in Fig. 5 shows the result using the region-based Chan-Vese energy. A curvature regularizer for the H^0 region-based active contour at a ratio of 1000 to 1 was used to compensate for noise. 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 contours do not have this problem and successfully track the object.

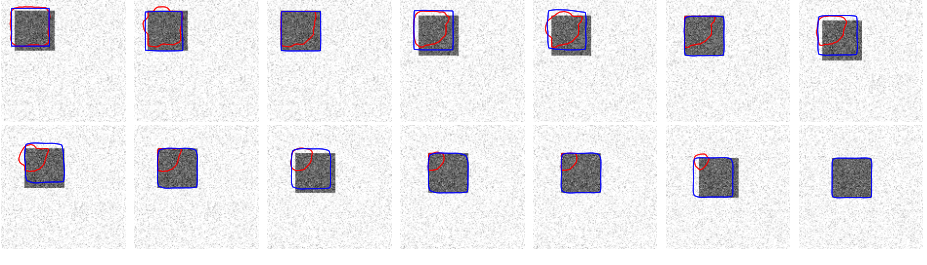

Fig. 4. Tracking a moving square in a noisy environment using edge-based H^0 active contour (red) and Sobolev active contour (blue)

6 Conclusion

We have introduced using Sobolev inner products on the set of perturbations of a curve rather than the traditional H^0 inner product used in all previous works on geometric active contours. We have demonstrated the general methodology for computing Sobolev gradients, and derived various flows with respect to Sobolev inner products. We have shown that Sobolev flows are smooth in the space of curves, are not as dependent on local image information as H^0

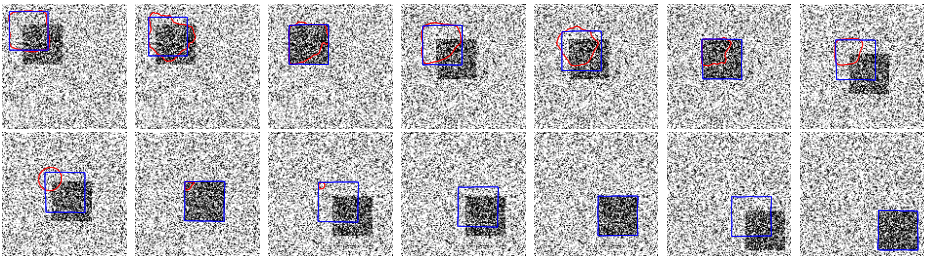

Fig. 5. Tracking a moving square in a noisy environment using region-based H^0 active contour (red) and Sobolev active contour (blue)

flows, are global motions which deform locally after moving globally, and do not require derivatives of the curve to be defined for region-based and edge-based energies.

Acknowlegdements. We thank George Roberts for aiding with simulations.

References

1. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International Journal of Computer Vision **1** (1987) 321–331
2. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. In: Proceedings of the IEEE Int. Conf. on Computer Vision, Cambridge, MA, USA (1995) 694–699
3. Kichenassamy, S., Kumar, A., Olver, P., Tannenbaum, A., Yezzi, A.: Gradient flows and geometric active contour models. In: Proceedings of the IEEE Int. Conf. on Computer Vision. (1995) 810–815
4. Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. Comm. Pure Appl. Math. **42** (1989) 577–685
5. Yezzi, A., Tsai, A., Willsky, A.: A statistical approach to snakes for bimodal and trimodal imagery. In: Int. Conf. on Computer Vision. (1999) 898–903
6. Chan, T., Vese, L.: Active contours without edges. IEEE Transactions on Image Processing **10** (2001) 266–277
7. Michor, P., Mumford, D.: Riemannian geometries on the space of plane curves. ESI Preprint 1425, arXiv:math.DG/0312384 (2003)
8. Yezzi, A., Mennucci, A.: Metrics in the space of curves. Preprint, arXiv:math.DG /0412454 (2005)
9. Osher, S., Sethian, J.: Fronts propagating with curvature-dependent speed: algorithms based on the Hamilton-Jacobi equations. J. Comp. Physics **79** (1988) 12–49
10. Younes, L.: Computable elasic distances between shapes. SIAM J. Appl. Math. **58** (1998) 565–586
11. Kriegl, A., Michor, P.: The Convenient Setting of Global Analysis. Volume 53 of Mathematical Surveys and Monographs. American Mathematical Society (1997)
12. Soatto, S., Yezzi, A.J.: DEFORMOTION: Deforming motion, shape average and the joint registration and segmentation of images. In: ECCV (3). (2002) 32–57