

# Sobolev gradients and joint variational image segmentation, denoising and deblurring

Miyoun Jung<sup>a</sup>, Ginmo Chung<sup>b</sup>, Ganesh Sundaramoorthi<sup>a</sup>, Luminita A. Vese<sup>a</sup>,  
and Alan L. Yuille<sup>a</sup>

<sup>a</sup>University of California, Los Angeles, U.S.A.

<sup>b</sup>Yale University, New Haven, U.S.A.

## ABSTRACT

We consider several variants of the active contour model without edges,<sup>4</sup> extended here to the case of noisy and blurry images, in a multiphase and a multilayer level set approach. Thus, the models jointly perform denoising, deblurring and segmentation of images, in a variational formulation. To minimize in practice the proposed functionals, one of the most standard ways is to use gradient descent processes, in a time dependent approach. Usually, the  $L^2$  gradient descent of the functional is computed and discretized in practice, based on the  $L^2$  inner product. However, this computation often requires theoretically additional smoothness of the unknown, or stronger conditions. One way to overcome this is to use the idea of Sobolev gradients.<sup>8,13,19</sup> We compare in several experiments the  $L^2$  and  $H^1$  gradient descents for image segmentation using curve evolution, with applications to denoising and deblurring. The Sobolev gradient descent is preferable in many situations and gives smaller computational cost.

**Keywords:** functional minimization, gradient descent, Sobolev gradients, image segmentation, image restoration, implicit representation.

## 1. INTRODUCTION

In this work we consider several variants of the active contour model without edges,<sup>4</sup> extended here to the case of noisy and blurry images, in a multiphase and a multilayer level set approach. Thus, the models jointly perform denoising, deblurring and segmentation of images. To minimize in practice the proposed functionals, one of the most standard ways is to use gradient descent processes, in a time dependent approach. Usually, the  $L^2$  gradient descent of the functional is computed and discretized in practice, based on the  $L^2$  inner product. However, this computation often requires theoretically additional smoothness of the unknown, or stronger conditions (for instance, having second order derivatives of the unknown in the space  $L^2$  of square integrable functions, when the functional to be minimized is of first order). In other words, the assumptions of the problem imply the existence of a weak solution only, while the  $L^2$  gradient descent assumes a strong solution.

To overcome this, J.W. Neuberger<sup>13</sup> has formalized in Lecture Notes in Mathematics, in 1997, the theory of Sobolev gradients for weak solutions in a function space approach. When the minimization is well posed, the specific Sobolev gradient imposed by the function space of the unknown is the right framework and is satisfied by the weak solution. However, the formulation of the  $L^2$  gradient requires strong solution.

For practical purposes, there are several choices of the Sobolev gradient to be used for the gradient descent method. More recently, in two different interesting works by G. Sundaramoorthi, A. Yezzi, A.C. Menicucci,<sup>19</sup> and by G. Charpiat, P. Maurel, J.-P. Pons, R. Keriven, O. Faugeras,<sup>8</sup> related ideas have been applied to curve evolution problems, showing cases when the use of Sobolev gradients is more beneficial than the  $L^2$  gradient, especially in the area of shape metrics.

In the discrete case, the Sobolev gradients can be interpreted as preconditioning of the original time-dependent  $L^2$  descent flow, or as descent flows obtained for different choices of matrix norms.

---

Further author information (send correspondence to M.J.):

M.J.: gomtaeng@math.ucla.edu, G.C.: senninha@math.ucla.edu, G.S.: ganeshs@ucla.edu

L.A.V.: lvese@math.ucla.edu, A.L.Y.: yuille@stat.ucla.edu

Following the approach of J.W. Neuberger,<sup>13</sup> and inspired by the above mentioned work, we will compute and compare here the  $L^2$  and Sobolev  $H^1$  gradients for the specific problem of joint segmentation, denoising and deblurring, in a variational level set approach. We will see that the Sobolev  $H^1$  gradient gives improved results in terms of quality and speed. Related works for segmentation using Sobolev gradients have also been proposed by G. Dogan, P. Morin and R. Nochetto,<sup>9</sup> although not in an implicit level set approach, and by L. Bar and G. Sapiro<sup>1</sup> in a recent parallel work. Another work using Sobolev gradients for image analysis is by W.B. Richardson.<sup>16</sup>

The outline of the paper is as follows. In the next subsection we present our framework of construction of the Sobolev  $H^1$  gradient, in a general fashion. In Section 2 we present several region-based variational segmentation models for joint object detection, denoising and deblurring, and we formulate their corresponding  $L^2$  and  $H^1$  gradient descents. Section 3 presents several experimental results and comparisons on synthetic and real images, obtained by the models described in Section 2. We end the paper with a short conclusion.

### 1.1 Illustration of the Sobolev $H^1$ gradient

We follow in this section Neuberger<sup>13</sup> and Renka.<sup>15</sup> Suppose  $\Omega$  is an open, bounded and connected subset of  $\mathbb{R}^n$  with Lipschitz boundary  $\partial\Omega$ , and  $F : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}$  is a  $C^1$  integrand. For illustration we consider the following minimization problem:

$$\arg \min_u \left\{ E(u) = \int_{\Omega} F(u, \nabla u) dx \right\}$$

where  $u \in H^1(\Omega) = \{u \in L^2(\Omega) : \nabla u \in (L^2(\Omega))^n\}$  is a Hilbert Sobolev space with its standard  $H^1$  inner product denoted by  $\langle \cdot, \cdot \rangle_{H^1}$ , and with the constraint  $u = g$  on  $\partial\Omega$  (in the sense of the trace operator), for a given  $g \in H^{1/2}(\partial\Omega)$ .

To minimize the functional  $E$  defined on  $H^1(\Omega)$ , we consider a gradient descent method. The gradient descent is based on the observation that if some real-valued function  $G(x)$  is well-defined and differentiable in a neighborhood of a point  $a$ , then  $G(x)$  decreases fastest if one goes from  $a$  in the direction of the negative gradient of  $G$  at  $a$ ,  $-\nabla G(a)$ . It follows that, if  $b = a - \gamma \nabla G(a)$  for  $\gamma > 0$  a small enough number, then  $G(a) \geq G(b)$ . With this observation, starting with a guess  $x_0$  for a local minimum of  $G$ , and considering the sequence  $x_0, x_1, x_2, \dots$  such that  $x_{n+1} = x_n - \gamma_n \nabla G(x_n)$ , we have  $G(x_0) \geq G(x_1) \geq G(x_2) \dots$ . Thus, we consider the following iteration to obtain a minimizer of  $E$ ,

$$u_{k+1} = u_k - \alpha_k \nabla E(u_k), \quad k = 0, 1, 2, \dots \quad (1)$$

with  $u_0 \in H^1(\Omega)$  and some means of choosing step-lengths  $\alpha_k$ .

In order that the iterates  $u_k$  remain in  $H^1(\Omega)$ , the gradients  $\nabla E(u_k)$  must be elements of  $H^1$  i.e. derivatives must be compatible with the metric (inner product and corresponding norm). More precisely, the directional derivative of  $E$  at  $u \in H^1(\Omega)$  in the direction  $h$  (defined in the usual way by  $E'(u)h = \lim_{\epsilon \rightarrow 0} \frac{E(u+\epsilon h) - E(u)}{\epsilon}$ ) defines the Sobolev (i.e.  $H^1$ ) gradient  $\nabla_S E(u) \in H^1$ :

$$E'(u)h = \langle \nabla_S E(u), h \rangle_{H^1}, \quad \forall h \in H_0^1(\Omega),$$

where  $H_0^1(\Omega) = \{h \in H^1(\Omega) : h = 0 \text{ on } \partial\Omega\}$ . We show next the existence of the Sobolev gradient  $\nabla_S E(u)$ . Let  $P$  be the (self-adjoint) orthogonal projection of  $L^2(\Omega) \times (L^2(\Omega))^n$  onto  $\left\{ Dv = \begin{pmatrix} v \\ \nabla v \end{pmatrix} : v \in H_0^1(\Omega) \right\}$  and define  $\pi \begin{pmatrix} f \\ g \end{pmatrix} = f$  for  $f \in L^2(\Omega)$ ,  $g \in (L^2(\Omega))^n$ . Then for  $u \in H^1(\Omega)$  and  $h \in H_0^1(\Omega)$ , we have

$$\begin{aligned} E'(u)h &= \int_{\Omega} F'(Du) Dh \\ &= \langle (\nabla F)(Du), Dh \rangle_{(L^2(\Omega))^{n+1}} \\ &= \langle (\nabla F)(Du), PDh \rangle_{(L^2(\Omega))^{n+1}} \\ &= \langle P(\nabla F)(Du), Dh \rangle_{(L^2(\Omega))^{n+1}} \\ &= \langle \pi P(\nabla F)(Du), h \rangle_{H^1}. \end{aligned}$$

Thus, the Sobolev gradient  $\nabla_S E(u)$  exists in  $H^1$  and has the form  $\nabla_S E(u) = \pi P(\nabla F)(Du)$ .

Now, we consider an ordinary  $L^2$  gradient associated with the  $L^2$  inner product, to show the critical benefit of the Sobolev gradient. Similarly, for  $u \in H^1$  and  $h \in H_0^1$ ,

$$\begin{aligned} E'(u)h &= \int_{\Omega} F'(Du) Dh \\ &= \langle \nabla F(Du), Dh \rangle_{(L^2(\Omega))^{n+1}} \\ &= \langle D^T \nabla F(Du), h \rangle_{L^2(\Omega)}. \end{aligned}$$

So, we obtain the  $L^2$  gradient,  $\nabla_{L^2} E(u) = D^T \nabla F(Du)$ , if it exists. However, we cannot guarantee that  $D^T \nabla F(Du)$  is in  $L^2(\Omega)$ . Specifically, for  $D^T \nabla F(Du) \in L^2(\Omega)$ , we need more smoothness on  $u$  i.e.  $u \in H^2(\Omega)$ , which might cause an ill-posed problem or numerical instability (non-existence of  $\nabla_{L^2} E(u)$ ). Moreover, if we would use the  $L^2$  gradient  $D^T \nabla F(Du)$  in (1), and assuming that  $u_k, u_{k+1} \in H^1(\Omega)$ , then we would need to have an even stronger condition on the minimizer  $u$ , that  $D^T \nabla F(Du) \in H^1(\Omega)$ . In short, standard  $L^2$  descent methods are lacking in integrity for minimizing functionals that involve derivatives.

Finally, we derive the computationally amenable expression for  $\nabla_S E(u)$  using the  $L^2$  gradient  $\nabla_{L^2} E(u)$  (applying a practical derivation method R.J. Renka<sup>15</sup>). By the definition of the directional derivative of  $E$  at  $u$  with respect to  $h$ , we have the equality

$$\begin{aligned} E'(u)h &= \langle \nabla_{L^2} E(u), h \rangle_{L^2(\Omega)} \\ &= \langle \nabla_S E(u), h \rangle_{H^1} \end{aligned}$$

and the Sobolev gradient  $\nabla_S E(u)$  can be written as

$$\begin{aligned} \langle \nabla_S E(u), h \rangle_{H^1} &= \langle D \nabla_S E(u), Dh \rangle_{L^2(\Omega) \times (L^2(\Omega))^n} \\ &= \langle D^T D \nabla_S E(u), h \rangle_{L^2(\Omega)}. \end{aligned}$$

By combining the above two equations, we obtain the Sobolev (i.e.  $H^1$ ) gradient  $\nabla_S E(u)$  of the form

$$\nabla_S E(u) = (D^T D)^{-1} \nabla_{L^2} E(u),$$

where  $D^T D u = (I - \Delta)u$ , thus using the Sobolev operator  $(I - \Delta)^{-1}$  with the corresponding boundary conditions.

## 2. APPLICATIONS TO IMAGE PROCESSING

We consider in this section  $L^2$  gradient descents of the region-based segmentation models for object detection and piecewise-constant denoising-deblurring. Following the previous section, we also (formally) formulate their corresponding  $H^1$  gradient descents and we give details of the numerical algorithm. We restrict our presentation to two-dimensional image data  $f : \Omega \rightarrow R$ , where  $\Omega$  is an open, bounded and connected subset in the plane.

We recall here that we use the operator  $(I - \Delta)^{-1}$  in the following sense: let  $G \in L^2(\Omega)$ , and let  $w \in H^1(\Omega)$  be the unique solution of the weak variational formulation of the problem

$$w - \Delta w = G \text{ in } \Omega, \quad \frac{\partial w}{\partial \vec{n}} \Big|_{\partial \Omega} = 0,$$

in other words,  $w$  is the unique  $H^1(\Omega)$  function that satisfies

$$\int_{\Omega} (wv + \nabla w \cdot \nabla v) dx = \int_{\Omega} Gv dx, \quad \forall v \in H^1(\Omega).$$

The existence and uniqueness of  $w$  is guaranteed by the Lax-Milgram Lemma. Thus, we can define  $(I - \Delta)^{-1}$  from  $L^2(\Omega)$  to  $H^1(\Omega)$ , which is considered as a smoothing or low-pass operator.

## 2.1 Binary 2-phase segmentation & denoising model

As a first example, we consider the binary, region-based active contour model without edges for image segmentation introduced in,<sup>3,4</sup> This model is solved using implicit representation of curves,<sup>14,18</sup> and can be seen as a particular case of the Mumford and Shah segmentation model<sup>12</sup>. It can be applied to the detection of objects whose boundaries are not necessarily defined by gradient and to the automatic detection of interior contours. Let  $f \in L^\infty(\Omega) \subset L^2(\Omega)$  be the given data to be segmented, and  $\phi : \Omega \rightarrow \mathbb{R}$  be a level set function (usually a Lipschitz continuous function) whose zero level set represents the evolving curve  $C = \{x \in \Omega : \phi(x) = 0\}$ . The functional minimization for two phase segmentation is given by<sup>4</sup>

$$\inf_{c_1, c_2, \phi} E(c_1, c_2, \phi) = \frac{\lambda_1}{2} \int_{\Omega} |f(x) - c_1|^2 H(\phi) dx + \frac{\lambda_2}{2} \int_{\Omega} |f(x) - c_2|^2 (1 - H(\phi)) dx + \mu \int_{\Omega} |\nabla H(\phi)| dx,$$

where  $\lambda_1 > 0$ ,  $\lambda_2 > 0$ ,  $\mu \geq 0$  are tuning parameters and  $H$  is the one-dimensional Heaviside function. The goal is to minimize  $E$  with respect to  $\phi$ ,  $c_1$ , and  $c_2$ , to obtain the desired contour (objects boundary), given by the zero level set of  $\phi$  that optimally partitions the domain  $\Omega$  into two regions, such that  $f(x) \approx c_1$  where  $\phi(x) > 0$  and  $f(x) \approx c_2$  where  $\phi(x) < 0$ . In the gradient descent approach, we start with an initial curve defined implicitly, that evolves to steady state. The geometric constraint on the curve is imposed by the last term of the functional (total variation of the characteristic function  $H(\phi)$ ), which corresponds to the length of  $C$  in two dimensions.

First, keeping  $\phi$  fixed and minimizing the energy  $E(c_1, c_2, \phi)$  with respect to the unknowns  $c_1$  and  $c_2$ , we obtain the explicit expressions for these constants of  $\phi$  as the averages of  $f$  on each side of the curve,

$$c_1 = c_1(\phi) = \frac{\int_{\Omega} f(x) H(\phi(x)) dx}{\int_{\Omega} H(\phi(x)) dx}, \quad c_2 = c_2(\phi) = \frac{\int_{\Omega} f(x) (1 - H(\phi(x))) dx}{\int_{\Omega} (1 - H(\phi(x))) dx}.$$

Keeping now  $c_1$  and  $c_2$  fixed and formally minimizing the energy with respect to  $\phi$ , we can compute the  $L^2$  gradient of  $E$  with respect to  $\phi$ , denoted by  $\nabla_{L^2} E(c_1, c_2, \phi)$  associated with Neumann boundary condition  $\frac{\partial \phi}{\partial \vec{n}} = 0$  where  $\vec{n}$  is the normal to the boundary  $\partial\Omega$ .

Thus we obtain the  $L^2$  gradient with respect to  $\phi$ , which amounts to the following evolution equation for  $\phi(t, x)$ ,  $t > 0$ ,  $x \in \Omega$  with a given  $\phi(0, x) = \phi_0(x)$ :

$$\frac{\partial \phi}{\partial t}(t, x) = -\nabla_{L^2} E(c_1, c_2, \phi) = \delta(\phi) \left[ -\lambda_1 |f - c_1|^2 + \lambda_2 |f - c_2|^2 + \mu \operatorname{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) \right], \quad (2)$$

where  $\delta$  denotes the one-dimensional Dirac distribution (in practice, we substitute  $H$  and  $\delta$  by smooth approximations, as in<sup>4</sup>).

By applying the Sobolev operator  $(I - \Delta)^{-1}$  to  $\nabla_{L^2} E(c_1, c_2, \phi)$ , we derive the other evolution equation for  $\phi(t, x)$ ,  $t > 0$ ,  $x \in \Omega$  with a given  $\phi(0, x) = \phi_0(x)$ , that we call the  $H^1$  gradient descent:

$$\frac{\partial \phi}{\partial t}(t, x) = -\nabla_{H^1} E(c_1, c_2, \phi) = (I - \Delta)^{-1} \left\{ \delta(\phi) \left[ -\lambda_1 |f - c_1|^2 + \lambda_2 |f - c_2|^2 + \mu \operatorname{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) \right] \right\}. \quad (3)$$

The time-dependent flows (2) and (3) will be compared in practice in Section 3.

## 2.2 Joint segmentation, denoising & deblurring in a multilayer level set approach

We propose here a joint formulation for denoising, deblurring and piecewise-constant segmentation. For related work we refer the reader to,<sup>2,10,11,17</sup> We use a minimization approach and we consider the  $L^2$  and  $H^1$  gradient descents. To represent the unknown moving boundaries, we use the multilayer segmentation formulation introduced in<sup>5-7</sup> which is an extension of the binary case,<sup>3,4</sup> Let  $f = K * u + n$  be a given blurred noisy image, where  $K$  is a known blurring kernel (such as the Gaussian kernel) and  $n$  represents Gaussian additive noise of zero mean. We assume that the contours or jumps in the image  $u$  can be represented by the  $m$  distinct levels  $\{-\infty = l_0 < l_1 < l_2 < \dots < l_m < l_{m+1} = \infty\}$  of the same implicit (Lipschitz continuous) function  $\phi : \Omega \rightarrow \mathbb{R}$  partitioning  $\Omega$  into  $m + 1$  disjoint open regions  $R_j = \{x \in \Omega : l_{j-1} < \phi(x) < l_j\}$ ,  $1 \leq j \leq m + 1$ . Thus, we

recover the image  $u = c_1H(\phi - l_m) + \sum_{j=2}^{m-1} c_jH(\phi - l_{m-j+1})H(l_{m-j+2} - \phi) + c_{m+1}H(l_1 - \phi)$  by minimizing the following energy functional ( $\mu \geq 0$ ):

$$E(c_1, c_2, \dots, c_m, \phi) = \int_{\Omega} \left| f - K * \left( c_1H(\phi - l_m) + \sum_{j=2}^{m-1} c_jH(\phi - l_{m-j+1})H(l_{m-j+2} - \phi) + c_{m+1}H(l_1 - \phi) \right) \right|^2 dx + \mu \sum_{j=1}^m \int_{\Omega} |\nabla H(\phi - l_j)| dx$$

• In the binary case (one level  $m = 1$ ,  $l_1 = 0$ ) we assume the degradation model  $f = K * (c_1H(\phi) + c_2(1 - H(\phi))) + n$ , and we wish to recover  $u = c_1H(\phi) + c_2(1 - H(\phi))$  in  $\Omega$  together with a segmentation of  $f$ . The modified binary segmentation model incorporating the blur becomes:

$$E(c_1, c_2, \phi) = \inf_{c_1, c_2, \phi} \int_{\Omega} \left| f - K * (c_1H(\phi) + c_2(1 - H(\phi))) \right|^2 dx + \mu \int_{\Omega} |\nabla H(\phi)| dx.$$

We compute the Euler-Lagrange equations minimizing this energy with respect to  $c_1$ ,  $c_2$ , and  $\phi$ . Using alternating minimization, keeping first  $\phi$  fixed and minimizing the energy with respect to the unknown constants  $c_1$  and  $c_2$ , we obtain the following linear system of equations:

$$c_1 \int_{\Omega} k_1^2 dx + c_2 \int_{\Omega} k_1 k_2 dx = \int_{\Omega} f k_1 dx, \quad c_1 \int_{\Omega} k_1 k_2 dx + c_2 \int_{\Omega} k_2^2 dx = \int_{\Omega} f k_2 dx$$

with the notations  $k_1 = K * H(\phi)$  and  $k_2 = K * (1 - H(\phi))$ . Note that the linear system has a unique solution because the determinant of the coefficient matrix is not zero due to the Cauchy-Schwartz inequality  $(\int_{\Omega} k_1 k_2 dx)^2 \leq \int_{\Omega} k_1^2 dx \int_{\Omega} k_2^2 dx$ , where the equality holds if and only if  $k_1 = k_2$  for a.e.  $x \in \Omega$ . But clearly  $k_1 = K * H(\phi)$  and  $k_2 = K * (1 - H(\phi))$  are distinct, thus we have strict inequality.

Keeping now the constants  $c_1$  and  $c_2$  fixed and minimizing the energy with respect to  $\phi$ , we obtain the evolution equation by introducing an artificial time for the  $L^2$  gradient descent in  $\phi(t, x)$ ,  $t > 0$ ,  $x \in \Omega$

$$\begin{aligned} \frac{\partial \phi}{\partial t}(t, x) &= -\nabla_{L^2} E(c_1, c_2, \phi) \\ &= \delta(\phi) \left[ \left( \tilde{K} * f - c_1 \tilde{K} * (K * H(\phi)) - c_2 \tilde{K} * (K * (1 - H(\phi))) \right) (c_1 - c_2) + \mu \operatorname{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) \right]. \end{aligned} \quad (4)$$

where  $\tilde{K}(x) = K(-x)$ .

Similarly, the  $H^1$  gradient descent in  $\phi(t, x)$  for  $t > 0$ ,  $x \in \Omega$ , is

$$\begin{aligned} \frac{\partial \phi}{\partial t}(t, x) &= -\nabla_{H^1} E(c_1, c_2, \phi) \\ &= (I - \Delta)^{-1} \left\{ \delta(\phi) \left[ \left( \tilde{K} * f - c_1 \tilde{K} * (K * H(\phi)) - c_2 \tilde{K} * (K * (1 - H(\phi))) \right) (c_2 - c_1) \right. \right. \\ &\quad \left. \left. - \mu \operatorname{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) \right] \right\}. \end{aligned} \quad (5)$$

The time-dependent flows (4) and (5) will be compared in practice in Section 3.

• In the case of two distinct levels  $l_1 < l_2$  of the level set function  $\phi$  ( $m = 2$ ), we wish to recover a piecewise-constant image of the form  $u = c_1H(\phi - l_2) + c_2H(l_2 - \phi)H(\phi - l_1) + c_3H(l_1 - \phi)$  and a segmentation of  $f$ ,

assuming the degradation model  $f = K * (c_1 H(\phi - l_2) + c_2 H(l_2 - \phi) H(\phi - l_1) + c_3 H(l_1 - \phi)) + n$ , by minimizing

$$\begin{aligned} \inf_{c_1, c_2, c_3, \phi} E(c_1, c_2, c_3, \phi) &= \int_{\Omega} \left| f - K * (c_1 H(\phi - l_2) + c_2 H(l_2 - \phi) H(\phi - l_1) + c_3 H(l_1 - \phi)) \right|^2 dx \\ &\quad + \mu \sum_{j=1}^2 \int_{\Omega} |\nabla H(\phi - l_j)| dx. \end{aligned} \quad (6)$$

Similar to the previous binary model with blur, for fixed  $\phi$ , the unknown constants are computed by solving the linear system of three equations:

$$\begin{aligned} c_1 \int k_1^2 dx + c_2 \int k_1 k_2 dx + c_3 \int k_1 k_3 dx &= \int f k_1 dx \\ c_1 \int k_1 k_2 dx + c_2 \int k_2^2 dx + c_3 \int k_2 k_3 dx &= \int f k_2 dx \\ c_1 \int k_1 k_3 dx + c_2 \int k_2 k_3 dx + c_3 \int k_3^2 dx &= \int f k_3 dx \end{aligned}$$

where  $k_1 = K * H(\phi - l_2)$ ,  $k_2 = K * (H(l_2 - \phi) H(\phi - l_1))$ , and  $k_3 = K * H(l_1 - \phi)$ .

For fixed  $c_1$ ,  $c_2$  and  $c_3$ , by minimizing the functional  $E$  with respect to  $\phi$ , we obtain the  $L^2$  gradient descent for  $\phi(t, x)$ ,  $t > 0$ ,  $x \in \Omega$ :

$$\begin{aligned} \frac{\partial \phi}{\partial t}(t, x) &= -\nabla_{L^2} E(c_1, c_2, c_3, \phi) \\ &= \tilde{K} * (f - K * (c_1 H(\phi - l_2) + c_2 H(l_2 - \phi) H(\phi - l_1) + c_3 H(l_1 - \phi))) (c_1 \delta(\phi - l_2) \\ &\quad + c_2 H(l_2 - \phi) \delta(\phi - l_1) - c_2 H(\phi - l_1) \delta(l_2 - \phi) - c_3 \delta(l_1 - \phi)) \\ &\quad + \mu \operatorname{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) (\delta(\phi - l_1) + \delta(\phi - l_2)). \end{aligned} \quad (7)$$

Similarly, the corresponding  $H^1$  gradient descent for  $\phi(t, x)$ ,  $t > 0$ ,  $x \in \Omega$ , is

$$\begin{aligned} \frac{\partial \phi}{\partial t}(t, x) &= -\nabla_{H^1} E(c_1, c_2, c_3, \phi) = -(I - \Delta)^{-1} \nabla_{L^2} E(c_1, c_2, c_3, \phi) \\ &= (I - \Delta)^{-1} \left\{ \tilde{K} * (f - K * (c_1 H(\phi - l_2) + c_2 H(l_2 - \phi) H(\phi - l_1) + c_3 H(l_1 - \phi))) (c_1 \delta(\phi - l_2) \right. \\ &\quad + c_2 H(l_2 - \phi) \delta(\phi - l_1) - c_2 H(\phi - l_1) \delta(l_2 - \phi) - c_3 \delta(l_1 - \phi)) \\ &\quad \left. + \mu \operatorname{div} \left( \frac{\nabla \phi}{|\nabla \phi|} \right) (\delta(\phi - l_1) + \delta(\phi - l_2)) \right\}. \end{aligned} \quad (8)$$

The time-dependent flows (7) and (8) will be compared in practice in Section 3.

### 2.3 Algorithm for solving the $H^1$ gradient descent

By applying the spatial Sobolev operator  $(I - \Delta)^{-1}$  to  $\nabla_{L^2} E(\phi)$ , we derived the evolution equation for  $\phi(t, x)$ ,  $t > 0$ ,  $x \in \Omega$  using the  $H^1$  gradient,  $\nabla_{H^1} E(\vec{c}, \phi)$ :

$$\frac{\partial \phi}{\partial t}(t, x) = -\nabla_{H^1} E(\vec{c}, \phi) = -(I - \Delta)^{-1} \nabla_{L^2} E(\vec{c}, \phi),$$

where  $\vec{c} = (c_1, c_2)$  or  $\vec{c} = (c_1, c_2, c_3)$ . Now we briefly present the details of the numerical algorithm for this evolution equation ( $H^1$  gradient descent). First, we rewrite the above equation as

$$(I - \Delta) \frac{\partial \phi}{\partial t}(t, x) = -\nabla_{L^2} E(\vec{c}, \phi).$$

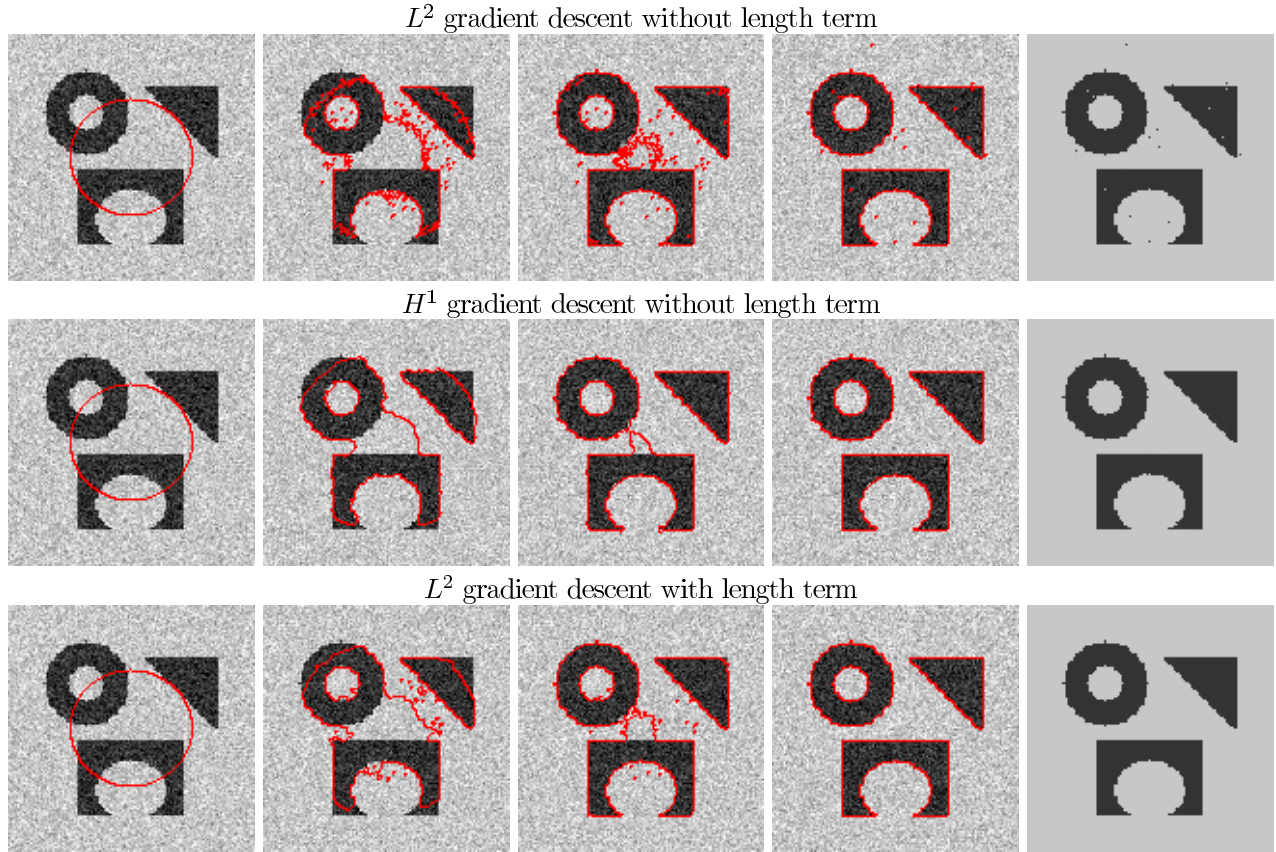


Figure 1. Segmentation & denoising of a synthetic image using the binary segmentation model without blur: curve evolution using  $L^2$  and  $H^1$  gradient descents, and the final segmented denoised image  $u = c_1 H(\phi) + c_2(1 - H(\phi))$ . Top row:  $L^2$  gradient descent (2) without length term,  $\mu = 0$ ; middle row:  $H^1$  gradient descent (3) without length term,  $\mu = 0$ ; bottom row:  $L^2$  gradient descent (2) with length term,  $\mu = 20 \cdot 255^2$ .

In the discrete case, let  $f_{i,j} \approx f(x_i, y_j)$ ,  $\phi_{i,j}^n \approx \phi(n\Delta t, x_i, y_j)$  with  $(x_i, y_j)$  the discrete points,  $\Delta t$  the time step,  $n \geq 0$ , and let  $\Delta x$  and  $\Delta y$  be the space steps. Then for each  $n > 0$  (given  $\phi^n$ ),

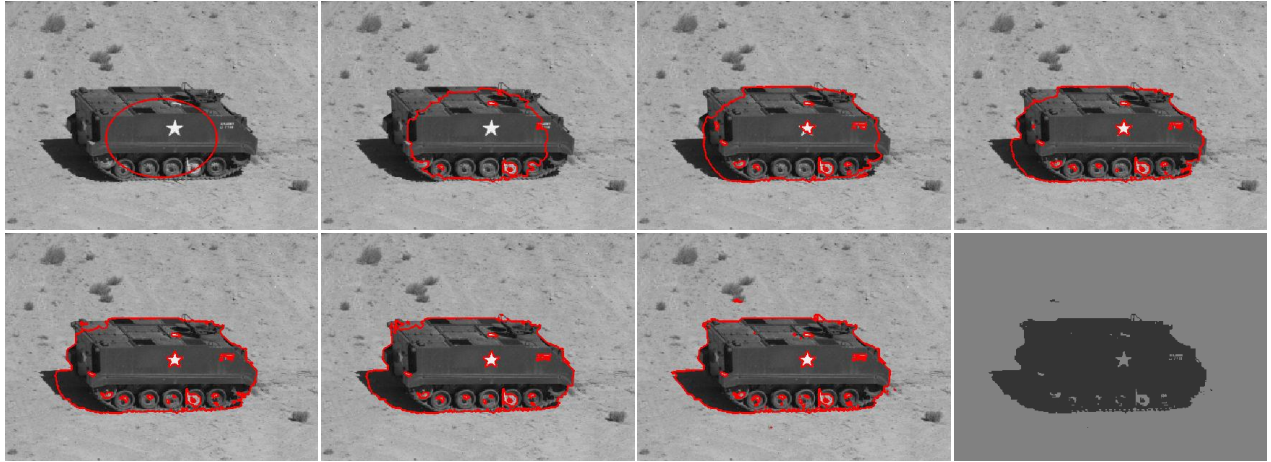
- evaluate the unknown constants  $c_i = c_i(\phi^n)$
- discretize by finite differences and evaluate  $G^n := -\nabla_{L^2} E(\vec{c}, \phi^n)$
- introduce the notation  $w$  (will correspond to  $w = \frac{\phi^{n+1} - \phi^n}{\Delta t}$ )
- solve  $(I - \Delta)w = G^n$  with  $\frac{\partial w}{\partial \vec{n}}|_{\partial\Omega} = 0$  in  $w$  using the semi-implicit scheme: start with  $w^0 = 0$  (or the previous  $w$ ), iterate for  $l = 1, 2, \dots$  until reaching the steady state solution  $w$ :

$$w_{i,j}^{l+1} - \left\{ \frac{w_{i+1,j}^l - 2w_{i,j}^{l+1} + w_{i-1,j}^l}{\Delta x^2} + \frac{w_{i,j+1}^l - 2w_{i,j}^{l+1} + w_{i,j-1}^l}{\Delta y^2} \right\} = G_{i,j}^n$$

- update  $\phi^{n+1} = \phi^n + \Delta t \cdot w$

We note that other more optimized methods could be used for inverting the operator  $I - \Delta$ .

$L^2$  gradient descent



$H^1$  gradient descent

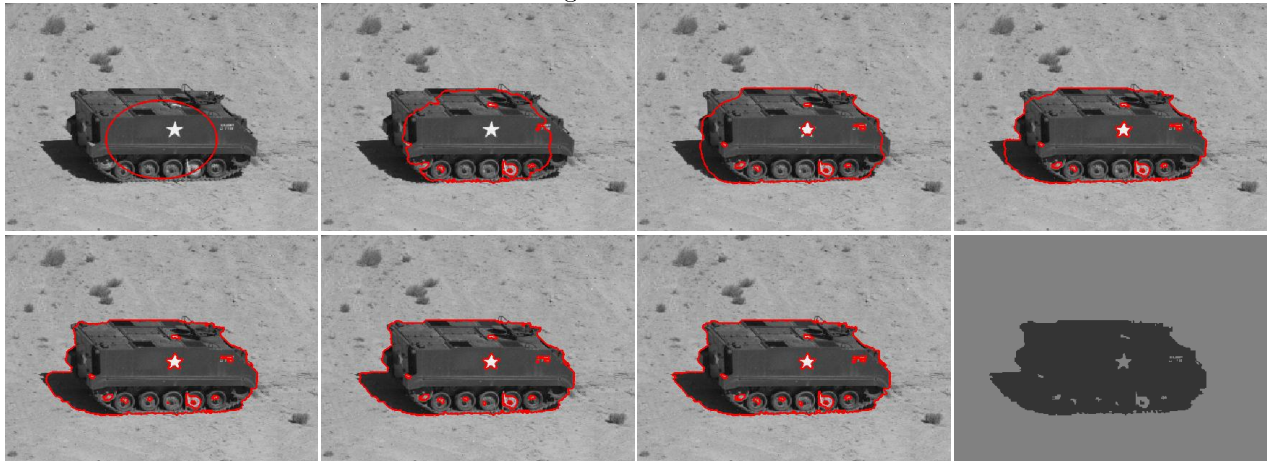


Figure 2. Segmentation of a real image using the binary model without blur. Top, from left to right, top to bottom: curve at iterations 0, 50, 100, 200, 300, 500, 1000 using  $L^2$  gradient descent (2) with  $\mu = 5 \cdot 255^2$ , and the final segmented image  $u$ . Bottom, from left to right, top to bottom: curve at iterations 0, 5, 10, 20, 30, 40, 70 using  $H^1$  gradient descent (3) with  $\mu = 5 \cdot 255^2$ , and the final segmented image  $u$ .

Image	$L^2$	$\Delta t$	Iterations	Total time (s)	$H^1$	$\Delta t$	Iterations	Total time (s)
Fig. 1	$\mu > 0$	0.005	700	13.4371	$\mu = 0$	0.1	30	2.4868
Fig. 2	$\mu > 0$	0.02	1000	106.4395	$\mu > 0$	0.25	70	28.0375
Fig. 3	$\mu > 0$	0.005	300	15.9263	$\mu > 0$	0.02	100	11.7915
Fig. 4	$\mu > 0$	0.002	400	11.5187	$\mu > 0$	0.01	100	5.89979
Fig. 5	$\mu > 0$	0.01	2000	203.1195	$\mu > 0$	0.05	500	124.1719
Fig. 6	$\mu > 0$	0.01	500	29.9126	$\mu > 0$	0.02	200	23.4868

Table 1. Summary of the results and time comparison of  $L^2$  and  $H^1$  gradient descents.



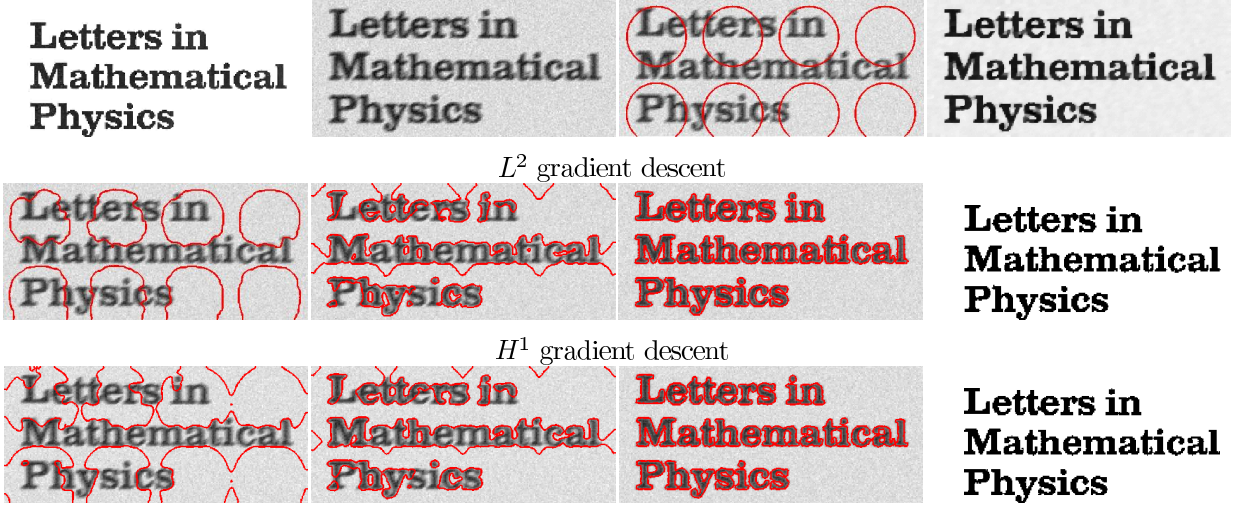


Figure 3. Joint segmentation, denoising and deblurring using the binary level set model. Top row: (from left to right) original image, degraded image (blurred with motion blur kernel of length 10, oriented at an angle  $\theta = 25^\circ$  w.r.t. the horizon and contaminated by Gaussian noise with  $\sigma_n = 10$ ), initial curve, and restored image (SNR=27.3701) of noisy blurry image using RO model<sup>17</sup> (shown for comparison). Second row:  $L^2$  gradient descent (4) curve evolution at iterations 50, 100, 300 with  $\mu = 5 \cdot 255^2$ , and the restored image  $u$  (SNR=28.1827). Bottom row:  $H^1$  gradient descent (5) curve evolution at iterations 10, 20, 100 with  $\mu = 5 \cdot 255^2$ , and the restored image  $u$  (SNR=29.5093).  $(c_1, c_2)$ : original image  $\approx (62.7525, 259.8939)$ , restored  $u$  using  $L^2 = (61.9194, 262.7795)$  and  $H^1 = (62.6023, 261.7207)$ .

### 3. EXPERIMENTAL RESULTS

We present in this section several experimental results and comparisons between the use of the  $L^2$  and  $H^1$  gradient descents, for segmentation, joint with denoising and deblurring. The results are summarized in Table 1, in terms of parameters and time comparison.

In Fig. 1 we compare the binary segmentation models (2) and (3) on a noisy synthetic image. Note that in the first two rows of Fig. 1, we set  $\mu = 0$  (no length term). While the  $L^2$  gradient descent without length term produces a noisy result and a “noisy” curve evolution, the  $H^1$  gradient descent without length term produces a very smooth curve evolution, even smoother than the one with the  $L^2$  gradient descent with length term, with the reduced total computational cost (this effect may be desirable in some applications, such as tracking or medical imaging).

Fig. 2 also compares (2) and (3), both with length term  $\mu > 0$ , on a real image. We observe again not only that the  $H^1$  gradient descent gives faster speed and smoother curve evolution than the  $L^2$  one, but also that it provides a better (more desirable) final segmentation.

Fig. 3 tests the joint segmentation and restoration model for the binary case with a noisy blurry image, and it shows the results using the  $L^2$  gradient descent (4) and the  $H^1$  gradient descent (5), with length term  $\mu > 0$ . In addition, we compare our proposed model with the well-known image recovery model proposed by Rudin and Osher.<sup>17</sup> First, we see that the proposed model gives better result than RO model according to SNR (signal-to-noise ratio) and visually. Moreover, in this case, the restored image using  $H^1$  gradient descent is better than the one using  $L^2$  gradient descent in the sense of SNR, and the  $H^1$  gradient descent gives smaller total computational time.

In Figures 4, 5 and 6, we compare the  $L^2$  (7) and  $H^1$  (8) gradient descents in the joint segmentation and restoration model for the multilayer case with noisy blurry images, applied to one synthetic image and two medical images. For all these results, the  $H^1$  gradient descent gives faster computations and smoother curve evolution. Especially, for the cell and brain images, the  $H^1$  gradient descent gives more desirable final curves, which leads to visually better recovered images than the  $L^2$  one. In both examples, while the  $H^1$  gradient descent can give finer final curves by decreasing the parameter  $\mu$ , the  $L^2$  one cannot provide smoother final curves like

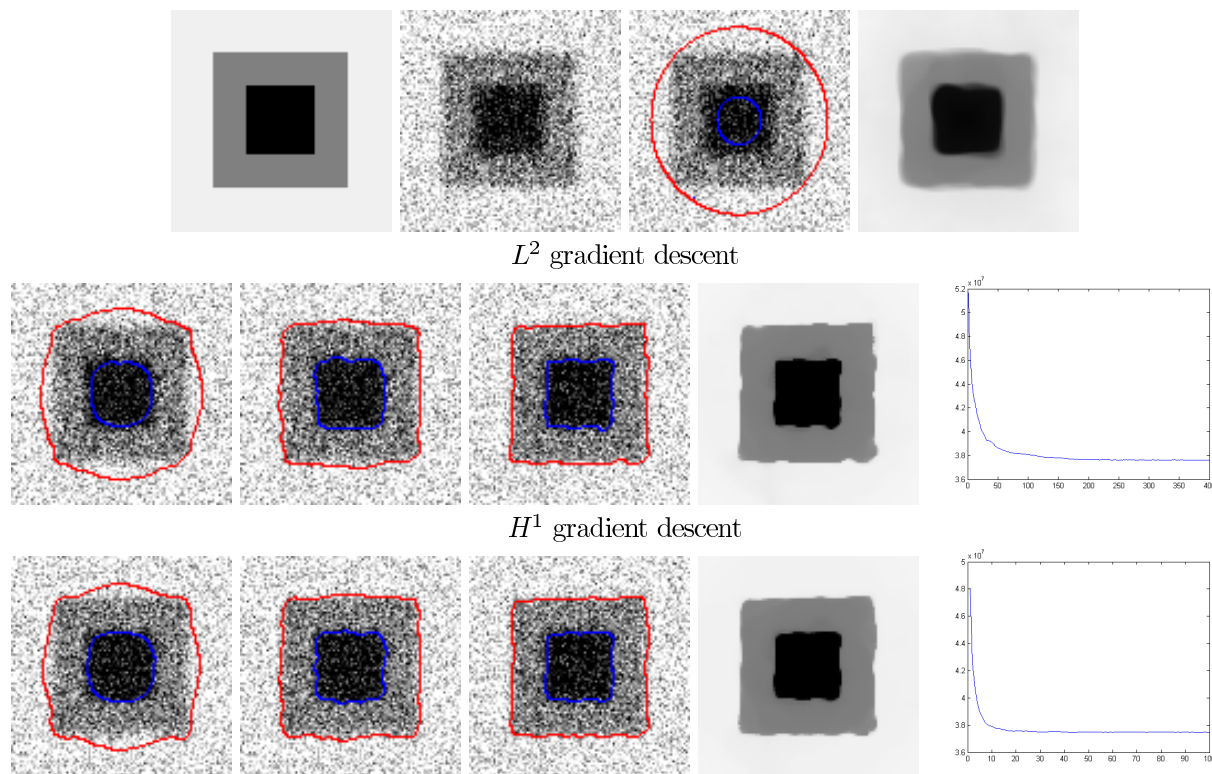


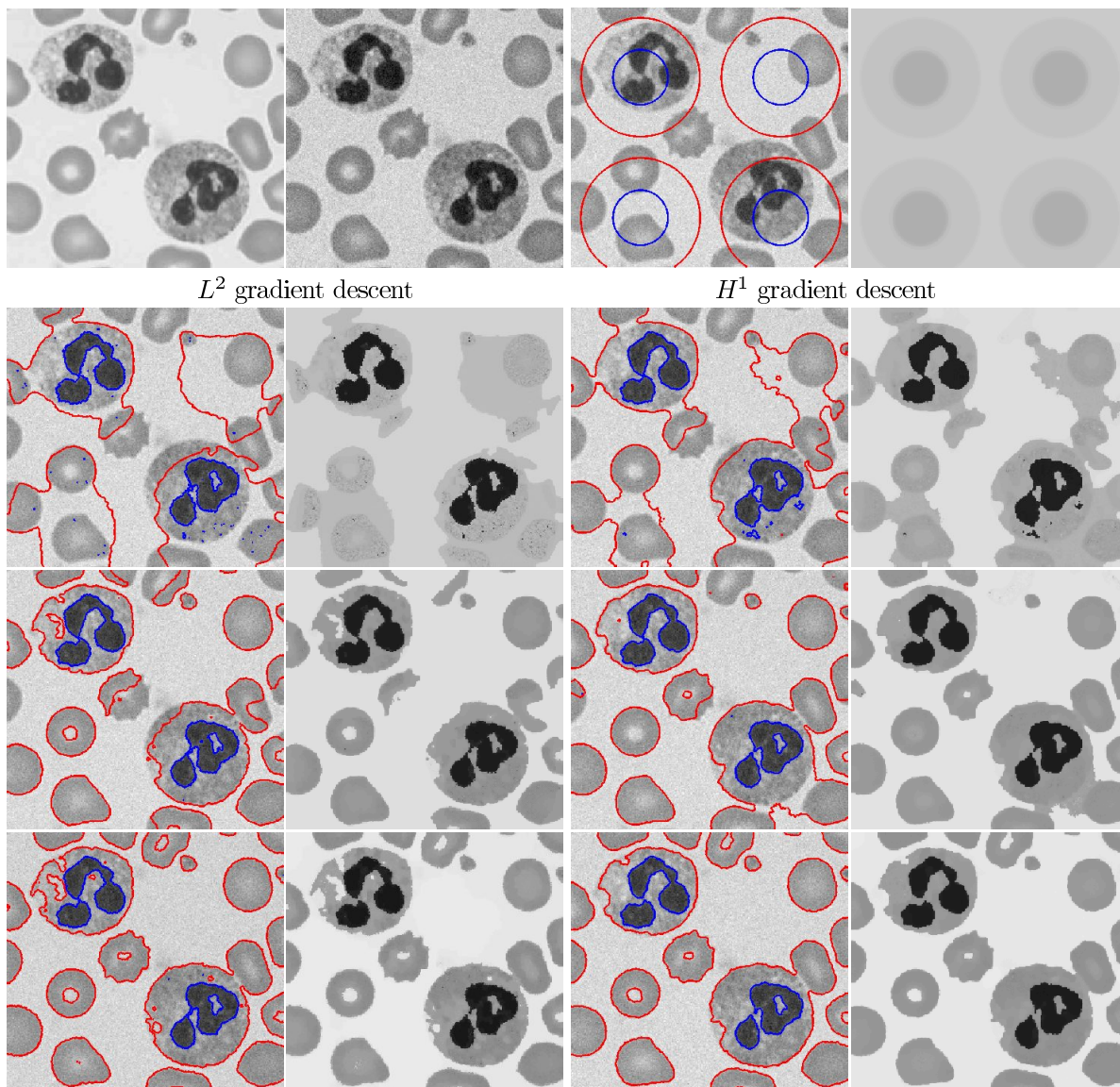
Figure 4. Joint segmentation and restoration using the multilayer segmentation model with one level set function and two levels  $l_1 = 0$ ,  $l_2 = 30$ . Top row: original image, image blurred with Gaussian kernel with  $\sigma_b = 3$  and contaminated by Gaussian noise  $\sigma_n = 60$ , initial curves, and restored image (SNR=26.6349) of noisy blurry image using RO model<sup>17</sup> (shown for comparison). Second row:  $L^2$  gradient descent (7) curve evolution at iterations 5, 50, 400 with  $\mu = 5 \cdot 255^2$ , the restored image  $u$  (SNR=33.6663), and plot of energy functional (6). Bottom row:  $H^1$  gradient descent (8) curve evolution at iterations 2, 10, 100 with  $\mu = 5 \cdot 255^2$ , the restored image  $u$  (SNR=35.7167), and plot of energy functional (6).

the ones obtained by  $H^1$  gradient, which may be caused by the numerical instability due to the relation of  $\mu$  and  $\Delta t$ . Moreover, according to the plots of energy functionals in Fig. 4, we observe that the  $H^1$  gradient descent converges faster than the  $L^2$  one.

Before ending this section, we briefly discuss the choice of parameter  $\mu$ . As the parameter  $\mu$  increases, a smoother curve can be obtained while the time step  $\Delta t$  needs to become smaller for the numerical stability. However,  $\Delta t$  cannot be too small because a very small time step might produce other numerical errors, and moreover the  $L^2$  gradient descent needs already a small  $\Delta t$  for an explicit scheme. In addition, a very small change in the parameter  $\mu$  gives very different results using the  $L^2$  gradient, while for the  $H^1$  gradient descent, a change of the parameter  $\mu$  produces very similar results. Thus, the  $L^2$  gradient descent is more sensitive to the choice of parameter  $\mu$ .

#### 4. CONCLUSION

We have considered several minimization problems for joint segmentation, denoising and deblurring. For the computation of their minimizers, we have compared the  $L^2$  and  $H^1$  gradient descents. Although the  $H^1$  gradient descent requires more operations per main iteration, the total computational cost is reduced, since fewer main iterations are necessary (roughly speaking, the path evolution of the curve from the initial position to the final one is smaller for the  $H^1$  gradient descent). Depending on the objective, other Sobolev gradients  $H^s$  can be used for functional minimization in image processing and computer vision.



$L^2$  gradient descent

$H^1$  gradient descent

Figure 5. Joint segmentation and restoration using the multilayer segmentation model with one level set function and two levels  $l_1 = 0$ ,  $l_2 = 30$ . Top row: original image, image blurred with Gaussian kernel with  $\sigma_b = 0.01$  and contaminated by Gaussian noise  $\sigma_n = 10$ , initial curve, initial  $u$ . From second to the bottom rows: curve evolution and restored  $u$  using (left two columns)  $L^2$  gradient descent (7) with  $\mu = 0.012 \cdot 255^2$ , and (right two columns)  $H^1$  gradient descent (8) with  $\mu = 0.03 \cdot 255^2$ .

### ACKNOWLEDGMENTS

This work has been supported in part by the Institute for Pure and Applied Mathematics (IPAM), by the National Institutes of Health (NIH) through the NIH Roadmap for Medical Research Grant U54 RR021813 entitled Center for Computational Biology (CCB), and by the National Science Foundation Grants DMS-0714945 and DMS-0312222.

### REFERENCES

- [1] L. Bar and G. Sapiro, "Generalized Newton method for energy formulations in image processing", IMA preprint series 2195, 2008.

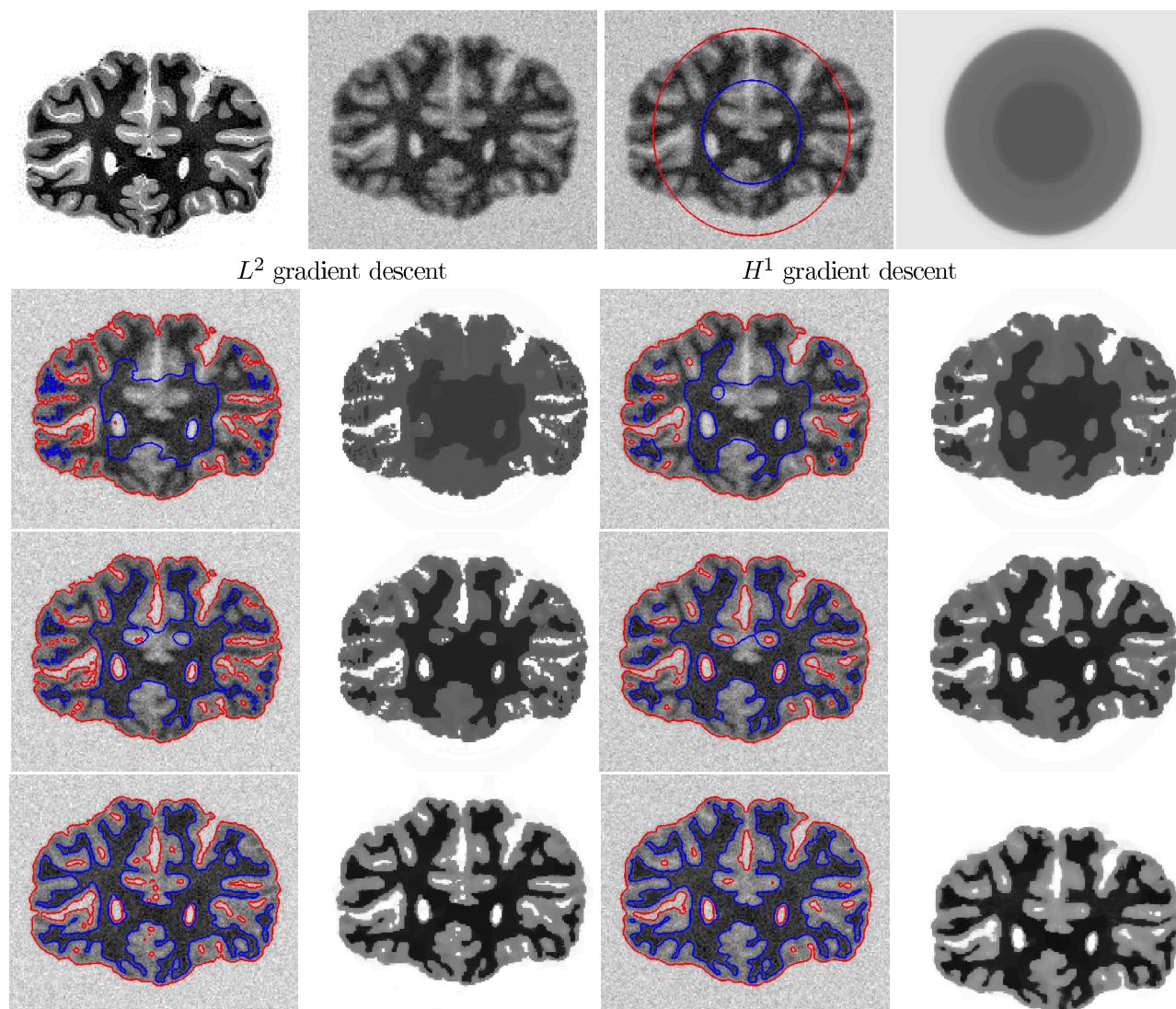


Figure 6. (Similar with Fig. 5) Top row: original image, image blurred with Gaussian kernel with  $\sigma_b = 1$  and contaminated by Gaussian noise  $\sigma_n = 20$ , initial curve, initial  $u$ . From second to bottom rows: curve evolution and restored  $u$  using (left two columns)  $L^2$  gradient descent (7),  $\mu = 0.02 \cdot 255^2$ , and (right two columns)  $H^1$  gradient descent (8),  $\mu = 0.027 \cdot 255^2$ .  $(c_1, c_2, c_3)$ : original image  $\approx (12.7501, 125.3610, 255.6453)$ , restored  $u$  using  $L^2 = (22.4797, 136.9884, 255.0074)$  and  $H^1 = (16.7475, 132.5393, 255.0811)$ .

- [2] L. Bar, N. Sochen, and N. Kiryati, "Semi-blind image restoration via Mumford-Shah regularization", IEEE TIP 15(2): 483-493, 2006.
- [3] T. Chan and L. Vese, "An active contour model without edges", LNCS 1682: 141-151, 1999.
- [4] T.F. Chan and L.A. Vese, "Active contours without edges", IEEE TIP 10(2): 266-177, 2001.
- [5] J. Chung, "A unifying framework for piecewise-constant image segmentation and deblurring", Ph.D. Thesis, UCLA, 2007.
- [6] J. Chung and L. Vese, "Energy Minimization Based Segmentation and Denoising Using a Multilayer Level Set Approach", LNCS 3757: 439-455, 2005.
- [7] J. Chung and L.Vese, "Image segmentation using a multilayer level-set approach", to appear in Computing and Visualization in Science, 2009.

- [8] G. Charpiat, P. Maurel, J.-P. Pons, R. Keriven, O. Faugeras, "Generalized Gradients: Priors on Minimization Flows", *IJCV* 73(3): 325 - 344, 2007.
- [9] G. Dogan, P. Morin and R. Nochetto, "A Variational Shape Optimization Approach for Image Segmentation with a Mumford-Shah Functional", *SIAM Journal on Scientific Computing* 30(6): 3028-3049, 2008.
- [10] D. Kim, A. Tsai, M. Cetin and A. Willsky, "A curve evolution based variational approach to simultaneous image restoration and segmentation", *Proc. IEEE ICIP*, Vol. 1: 109-112, 2002.
- [11] C. Li, C.-Y. Kao, J.C. Gore, Z. Ding, "Implicit Active Contours Driven by Local Binary Fitting Energy", *CVPR* 2007.
- [12] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems", *Commun. Pure Appl. Math.*, 42(5): 577-685, 1989.
- [13] J.W. Neuberger, "Sobolev gradients and differential equations", *Springer Lecture notes in Mathematics*, Vol. 1670, 1997.
- [14] S. Osher, R. Fedkiw, "Level set methods and dynamic implicit surfaces", Springer, New York, 2003.
- [15] R.J.Renka, "A simple explanation of the Sobolev gradient method", 2006 (<http://www.cse.unt.edu/renka/papers/sobolev.pdf>).
- [16] W.B. Richardson, "Sobolev gradient preconditioning for image processing PDEs", *Commun Numer. Methods in Engineering*, 24(6): 493-504, 2008.
- [17] L. Rudin and S. Osher, "Total variation based image restoration with free local constraints", *Proc. IEEE ICIP*, Vol. I: 31-35, 1994. Austin, TX.
- [18] J.A. Sethian, "Level Set Methods and Fast Marching Methods Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science", Cambridge University Press, 1999.
- [19] G. Sundaramoorthi, A. Yezzi, and A. C. Mennucci, "Sobolev active contours", *IJCV* 73(3):345-366, 2007.