SIAM J. APPL. MATH.
Vol. 66, No. 5, pp. 1632–1648

ALGORITHMS FOR FINDING GLOBAL MINIMIZERS OF IMAGE SEGMENTATION AND DENOISING MODELS*

TONY F. CHAN†, SELIM ESEDOĞLU‡, AND MILA NIKOLOVA§

Abstract. We show how certain nonconvex optimization problems that arise in image processing and computer vision can be restated as convex minimization problems. This allows, in particular, the finding of global minimizers via standard convex minimization schemes.

Key words. denoising, segmentation

AMS subject classifications. 94A08, 65K10

DOI. 10.1137/040615286

1. Introduction.

Image denoising and segmentation are two related, fundamental problems of computer vision. The goal of denoising is to remove noise and/or spurious details from a given, possibly corrupted, digital picture while maintaining essential features such as edges. The goal of segmentation is to divide the image into regions that belong to distinct objects in the depicted scene.

Approaches to denoising and segmentation based on the calculus of variations and partial differential equations (PDEs) have had great success. One important reason for their success is that these models are particularly well suited to imposing geometric constraints (such as regularity) on the solutions sought. Among the best known and most influential examples are the Rudin–Osher–Fatemi (ROF) total variation–based image denoising model [22] and the Mumford–Shah image segmentation model [17].

Denoising models such as the ROF model can be easily adapted to different situations. An interesting scenario is the denoising of shapes: Here, the given image is binary (representing the characteristic function of the given shape), and the noise is in the geometry of the shape: Its boundary might be very rough, and the user might be interested in smoothing out its boundary, and perhaps removing small, unnecessary connected components of the shape. This task is a common first step in many object detection and recognition algorithms.

A common difficulty with many variational image processing models is that the energy functional to be minimized has local minima (which are not global minima). This is a much more serious drawback than nonuniqueness of global minimizers (which is also a common phenomenon) because local minima of segmentation and denoising models often have completely wrong levels of detail and scale: whereas global minimizers of a given model are usually all reasonable solutions, the local minima tend to be blatantly false. Many solution techniques for variational models are based on gradient descent, and are therefore prone to getting stuck in such local minima. This

*Received by the editors September 17, 2004; accepted for publication (in revised form) November 21, 2005; published electronically June 19, 2006.
http://www.siam.org/journals/siap/66-5/61528.html

†Mathematics Department, UCLA, Los Angeles, CA 90095 (TonyC@college.ucla.edu). The research of this author was supported in part by NSF contract DMS-9973341, NSF contract ACI-0072112, ONR contract N00014-03-1-0888, and NIH contract P20 MH65166.

‡Department of Mathematics, University of Michigan, Ann Arbor, MI 48109 (esedoglu@umich.edu). The research of this author was supported in part by NSF award DMS-0410085.

§Centre de Mathematiques et de Leurs Applications, ENS de Cachan, 61 av. du President Wilson, 94235 Cachan Cedex, France (nikolova@cmla.ens-cachan.fr).

makes the initial guess for gradient descent–based algorithms sometimes critically important for obtaining satisfactory results.

In this paper we propose algorithms which are guaranteed to find global minimizers of certain denoising and segmentation models that are known to have local minima. As a common feature, the models we consider involve minimizing functionals over characteristic functions of sets, which is a nonconvex collection; this feature is responsible for the presence of local minima. Our approach, which is based on observations of Strang in [23, 24], is to extend the functionals and their minimization to all functions in such a way that the minimizers of the extended functionals can be subsequently transformed into minimizers for the original models by simple thresholding. This allows, among other things, computing global minimizers for the original nonconvex variational models by carrying out standard convex minimization schemes.

Our first example is binary image denoising, which we briefly discuss as a precursor to and a motivation for the more general segmentation problem that we subsequently consider. Here the given noisy image for the ROF model is taken to be binary, and the solution is also sought among binary images. This problem has many applications where smoothing of geometric shapes is relevant. Some examples are the denoising of fax documents, and the fairing of surfaces in computer graphics. Because the space of binary functions is nonconvex, the minimization problem involved is actually harder than minimizing the original ROF model. In section 2, we recall results from [6] that show how this model can be written as a convex optimization problem, and we exhibit its use as a numerical algorithm. In this section, we also analytically verify that the energy concerned possesses local minimizers that are not global minimizers, which can trap standard minimization procedures.

In section 3, we extend some of the results of [6] to the more important problem of *image segmentation*. In particular, we consider the two-phase, piecewise constant Mumford–Shah segmentation functional proposed by Chan and Vese in [7] and show that part of the minimization involved can be given a convex formulation. This model has become a very popular tool in segmentation and related image processing tasks (also see [26] for its multiphase version). The convex formulation we obtain turns out to be closely related to the algorithm of Chan and Vese presented in [7]. Our observations indicate why this algorithm is successful in finding interior contours and other hard-to-get features in images.

2. Previous work. The results and the approach of this paper follow very closely the observations of Strang in [23, 24]. In those papers, optimization problems of the following form, among others, are studied:

$$\inf_{\{u:\int f u\,dx=1\}} \int |\nabla u|, \tag{1}$$

where $f(x)$ is a given function. It is shown in particular that the minimizers of (1) turn out to be characteristic functions of sets. The main idea involved is to express the functional to be minimized and the constraint in terms of the super level sets of the functions $u(x)$ and $f(x)$. The coarea formula of Fleming, Rishel, and Rishel [11] is the primary tool.

In this paper, the idea of expressing functionals in terms of level sets is applied to some simple image processing models. For instance, in section 4, where we study the piecewise constant Mumford–Shah energy, we show that the relevant energy, which is originally formulated in terms of *sets*, can be reformulated as an optimization problem over *functions* in such a way that the resulting *convex* energy turns out to be almost

the same as (1). After this reformulation, following Strang's work, we are also able to express the resulting convex variational problem in terms of super level sets of the unknown functions. That in turn allows us to extract a minimizer of the original nonconvex model from a minimizer of the convex functional by simple thresholding.

We should point out that our emphasis in this paper is in some sense opposite that of [23, 24]. Indeed, in those works the main point is that some energies of interest that need to be minimized over all functions turn out to have minimizers that take only two values. In our case, we start with a variational problem that is to be minimized over only functions that take two values (i.e., characteristic functions of sets) but show that we may instead minimize over all functions (that are allowed to take intermediate vales), i.e., we may ignore the nonconvex constraint. This allows us to end up with a convex formulation of the original nonconvex problem.

3. The ROF model. Rudin, Osher, and Fatemi's total variation image denoising model [22] is one of the best known and successful of PDE-based image denoising models. Indeed, being convex it is one of the simplest denoising techniques that has the all-important edge preserving property.

Let $D \subset \mathbf{R}^N$ denote the image domain. In practice, D is simply a rectangle, modeling the computer screen. Therefore, mathematically, it is natural to assume that D is a bounded domain with Lipschitz boundary. However, for the convenience of not dealing with boundaries, in this section we will take D to be the entire space $\mathbf{R}^N$. This simplifying assumption has no bearing on the essential ideas discussed below.

Let $f(x) : \mathbf{R}^N \to [0,1]$ denote the given (grayscale) possibly corrupted (noisy) image. The energy to be minimized in the standard ROF model is then given by

$$E_2(u,\lambda) = \int_{\mathbf{R}^N} |\nabla u| + \lambda \int_{\mathbf{R}^N} \Big(u(x) - f(x)\Big)^2 \, dx. \tag{2}$$

The appropriate value of the parameter $\lambda > 0$ in the model (2) can be determined if the noise level is known; an algorithm for doing so is given in [22]. If information about noise level is not available, then λ needs to be chosen by the user. This choice can be facilitated by the observation that λ acts as a scale parameter [25]: Its value determines in some sense the smallest image feature that will be maintained in the reconstructed image. Energy (2) is often minimized via gradient descent; however, see [5, 9] for an alternative approach in the $\lambda = 0$ case.

An interesting application of the ROF model described above is to *binary* image denoising. This situation arises when the given image $f(x)$ is binary (i.e., $f(x) \in \{0,1\}$ for all $x \in \mathbf{R}^N$) and is known to be the corrupted version of another binary image $u : \mathbf{R}^N \to \{0,1\}$ that needs to be estimated. Naturally, $f(x)$ can then be expressed as

$$f(x) = \mathbf{1}_\Omega(x),$$

where Ω is an arbitrary bounded measurable subset of $\mathbf{R}^N$. In this case, the noise is in the *geometry*; for example, the boundary $\partial\Omega$ of Ω might have spurious oscillations, or Ω might have small connected components (due to presence of noise) that need to be eliminated. The ROF model (2) can easily be specialized to this scenario by restricting the unknown $u(x)$ to have the form $u(x) = \mathbf{1}_\Sigma(x)$, where Σ is a subset of

$\mathbf{R}^N$. One then obtains the following optimization problem:

$$\min_{\substack{\Sigma\subset\mathbf{R}^N\\ u(x)=\mathbf{1}_\Sigma(x)}} \int_{\mathbf{R}^N} |\nabla u| + \lambda \int_{\mathbf{R}^N} \Big(u(x) - \mathbf{1}_\Omega(x)\Big)^2\, dx. \tag{3}$$

Problem (3) is nonconvex because the minimization is carried out over a nonconvex set of functions. Recalling that the total variation of the characteristic function of a set is its perimeter (see, e.g., [10, 12] for such basic facts), and noticing that the fidelity term in this case simplifies, we write (3) as the following geometry problem:

$$\min_{\Sigma\subset\mathbf{R}^N} \mathrm{Per}(\Sigma) + \lambda|\Sigma \triangle \Omega|, \tag{4}$$

where $\mathrm{Per}(\cdot)$ denotes the perimeter, $|\cdot|$ is the N-dimensional Lebesgue measure, and $S_1 \triangle S_2$ denotes the symmetric difference between the two sets S_1 and S_2.

Usual techniques for approximating the solution. A very successful method of solving problems of the type (4) has been via some *curve evolution* process, sometimes referred to as *active contours*. Indeed, the unknown set Σ can be described by its boundary $\partial\Sigma$. The boundary $\partial\Sigma$ is then updated iteratively, usually according to gradient flow for the energy involved.

Numerically, there are several ways of representing $\partial\Sigma$. For the applications mentioned above, explicit curve representations as in Kass, Witkin, and Terzopoulos [15] are not appropriate, since such methods do not allow changes in curve topology (and have a number of other drawbacks). Instead, the most successful algorithms are those based on either the level set method of Osher and Sethian [21, 20] or on the variational approximation approach known as Gamma convergence theory [8].

In the level set formulation, the unknown boundary $\partial\Sigma$ is represented as the 0-level set of a (Lipschitz) function $\phi : \mathbf{R}^N \to \mathbf{R}$:

$$\Sigma = \big\{x \in \mathbf{R}^N : \phi(x) > 0\big\},$$

so that $\partial\Sigma = \{x \in \mathbf{R}^N : \phi(x) = 0\}$. The functional to be minimized in (3), which we called $E_2(\cdot, \lambda)$, can then be expressed in terms of the function $\phi(x)$ as follows:

$$\int_{\mathbf{R}^N} |\nabla H(\phi(x))|\, dx + \lambda \int_{\mathbf{R}^N} \Big(H(\phi(x)) - \mathbf{1}_\Omega(x)\Big)^2\, dx. \tag{5}$$

Here, the function $H(x) : \mathbf{R} \to \mathbf{R}$ is the Heaviside function:

$$H(\xi) = \begin{cases} 0 & \text{if } x < 0,\\ 1 & \text{if } x \geq 0. \end{cases}$$

In practice, one takes a smooth (or at least Lipschitz) approximation to $H(x)$, which we shall call $H_\varepsilon(\xi)$, where $H_\varepsilon(\xi) \to H(\xi)$ in some manner as $\varepsilon \to 0$.

The Euler–Lagrange equation for (5) is easy to obtain. It leads to the following gradient flow:

$$\phi_t(x,t) = H'_\varepsilon(\phi)\left\{\mathrm{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right) + 2\lambda\Big(\mathbf{1}_\Omega(x) - H_\varepsilon(\phi)\Big)\right\}. \tag{6}$$

When (6) is simulated using reinitialization for the level set function $\phi(x)$ and a compactly supported approximation $H_\varepsilon(x)$ to $H(x)$, it is observed to define a continuous

evolution (with respect to, say, the L^1-norm) for the unknown function $u(x) = \mathbf{1}_\Sigma(x)$ and decreases the objective energy (3) through binary images. It is analogous to the gradient descent equation in [7], which is natural since (3) is the restriction to binary images of also the energy considered in that work, namely, the two-phase, piecewise constant Mumford–Shah segmentation energy. In section 4, we will consider this energy for general (not necessarily binary) images.

Another representation technique for the unknown set Σ in (4) is, as we mentioned, based on the Gamma convergence ideas. Here, the given energy is replaced by a sequence of approximate energies with a small parameter $\varepsilon > 0$ in them. The sequence converges to the original energy as $\varepsilon \to 0$. The approximations have the form

$$E_\varepsilon(u, \lambda) = \int_{\mathbf{R}^N} \varepsilon|\nabla u|^2 + \frac{1}{\varepsilon}W(u) + \lambda\Big\{u^2\big(c_1 - f\big)^2 + (1-u)^2\big(c_2 - f\big)^2\Big\}\, dx.$$

In this energy, $W(\xi)$ is a double-well potential with equidepth wells at 0 and 1; for instance, a simple choice is $W(\xi) = \xi^2(1-\xi)^2$. The term $\frac{1}{\varepsilon}W(u)$ can be thought of as a penalty term that forces the function u to look like the characteristic function of a set: u is forced to be approximately 0 or 1 on most of $\mathbf{R}^N$. The term $\varepsilon|\nabla u|^2$, on the other hand, puts a penalty on the transitions of u between 0 and 1. Taken together, it turns out that these terms both impose the constraint that u should be a characteristic function and approximate its total variation. Precise versions of these statements have been proved in [16]. The remaining terms in E_ε are simply the fidelity term written in terms of u. This approach was extended to the full Mumford–Shah functional in [4].

We now argue, with the help of a very simple example, that these techniques will get stuck in local minima in general, possibly leading to resultant images with the wrong level of detail. This fact is already quite familiar to researchers working with these techniques from practical numerical experience.

Example. Consider the two-dimensional case, where the observed binary image $f(x)$ to be denoised is the characteristic function of a ball $B_R(0)$ of radius R, which is centered at the origin. In other words, we take $\Omega = B_R(0)$. Implementing the gradient descent algorithm defined by (6) requires the choice of an initial guess for the interface $\phi(x)$ (or, equivalently, an initial guess for the set Σ that is represented by $\phi(x)$). A common choice in practical applications is to take the observed image itself as the initial guess. In our case, that means initially we set $\Sigma = B_R(0)$.

Now, one can see without much trouble that the evolution defined by (6) will maintain radial symmetry of $\phi(x)$. That means, at any given time $t \geq 0$, the set (i.e., the candidate for minimization) represented by $\phi(x)$ is of the form

$$\Big\{x \in \mathbf{R}^2 : \phi(x) > 0\Big\} = B_r(0)$$

for some choice of the radius $r \geq 0$. We can write the energy of $u(x) = \mathbf{1}_{B_r(0)}(x)$ in terms of r, as follows:

$$E(r) := E_2(\mathbf{1}_{B_r(0)}(x), \lambda) = 2\pi r + \lambda\pi|R^2 - r^2|.$$

A simple calculation shows that if $\lambda < \frac{2}{R}$, then the minimum of this function is at $r = 0$. Hence, if we fix $\lambda > 0$, then the denoising model prefers to remove disks of radius smaller than the critical value $\frac{2}{R}$.

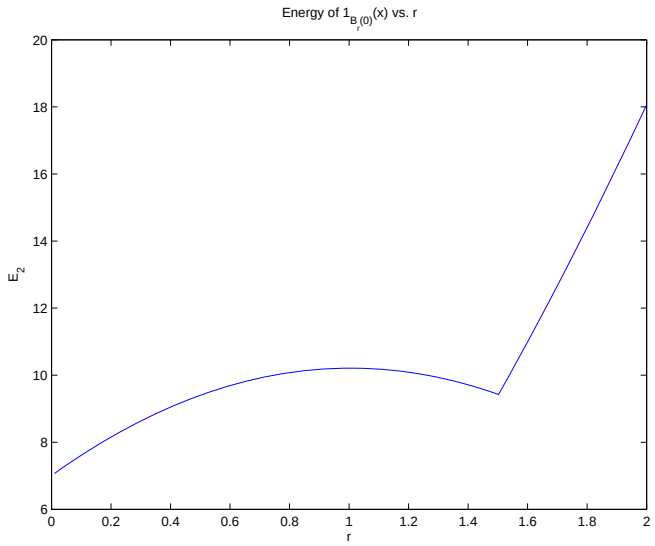

FIG. 1. *Energy* (3) *of* $u(x) = \mathbf{1}_{B_r(0)}(x)$ *as a function of* $r \in [0, 2]$ *when the observed image is given by* $f(x) = \mathbf{1}_{B_R(0)}(x)$. *Here,* $R = \frac{3}{2}$ *and the parameter* λ *was chosen to be* $\lambda = 1$. *There is clearly a local minimum, corresponding to* $r = R = \frac{3}{2}$.

But now, once again an easy calculation shows that if $R > \frac{1}{\lambda}$, then $E(r)$ has a local maximum at $r_{max}(\lambda) = \frac{1}{\lambda}$. See Figure 1 for the plot of $E(r)$ in such a case. Thus the energy minimization procedure described by (6) cannot shrink disks of radius $R \in (\frac{1}{\lambda}, \frac{2}{\lambda})$ to a point, even though the global minimum of the energy for an original image given by such a disk is at $u(x) \equiv 0$.

We can easily say a bit more: There exists $\delta > 0$ such that if $\Sigma \subset \mathbf{R}^N$ satisfies $|\Sigma \triangle B_R(0)| < \delta$, then $E_2(\mathbf{1}_\Sigma(x), \lambda) > E_2(\mathbf{1}_{B_R(0)}(x), \lambda)$. In other words, all binary images close to, but not identical with, the observed image $\mathbf{1}_{B_R(0)}(x)$ have strictly higher energy. This can be seen simply by noting that the energy of any region that is *not* a disk is strictly larger than the energy of the disk having the same area as the given region and its center at the origin.

To summarize: If $f(x) = \mathbf{1}_{B_R(0)}(x)$ with $R \in (\frac{1}{\lambda}, \frac{2}{\lambda})$, and if the initial guess for the continuous curve evolution–based minimization procedure (6) is taken to be the observed image $f(x)$ itself, then the procedure gets stuck in the local minimizer $u(x) = f(x)$. The unique global minimizer is actually $u(x) \equiv 0$.

Our example highlights the following caveat of using continuous curve evolution–based gradient descent algorithms in practice: There are many situations in which the user should be able to choose the value of λ that appears in the model in such a way that all image features smaller than the one implied by this choice of parameter are eliminated from the final result. (Such a need might arise, for instance, in the denoising of printed text, where the noise can consist of small ink blots.) With continuous curve evolution techniques, whether this goal will be achieved depends on the initial guess (our example above exhibits an unfortunate initial guess). It is clearly of interest to find an algorithm that does not have this dependence on initial conditions.

Proposed method for finding the global minimum. We now turn to an alternative way of carrying out the constrained, nonconvex minimization problem (3) that is guaranteed to yield a global minimum.

The crux of our approach is to consider minimization of the following convex energy, defined for any given observed image $f(x) \in L^1(\mathbf{R}^N)$ and $\lambda \geq 0$:

$$E_1(u(x), \lambda) := \int_{\mathbf{R}^N} |\nabla u| + \lambda \int_{\mathbf{R}^N} |u(x) - f(x)|\, dx. \tag{7}$$

This energy differs from the standard ROF model only in the fidelity term: The L^2-norm square of the original model is replaced by the L^1-norm as a measure of fidelity. It was previously introduced and studied in signal and image processing applications in [1, 2, 3, 18, 19, 6]. This variant of the ROF model has many interesting properties and uses; the point we'd like to make in this section is that it also turns out to solve our geometry denoising problem (4).

First, let us state the obvious fact that energies (2) and (7) agree on binary images (i.e., when both u and f are characteristic functions of sets). On the other hand, energy (7) is convex, but unlike energy (2), it is not strictly so. Accordingly, its global minimizers are not unique in general. Nevertheless, being convex, it does not have any local minima that are not global minima, unlike the constrained minimization (3). We therefore adopt the following notation: For any $\lambda \geq 0$, we let $M(\lambda)$ denote the set of all minimizers of $E_1(\cdot, \lambda)$. It is easy to show that for each $\lambda \geq 0$ the set $M(\lambda)$ is nonempty, closed, and convex.

The relevance of energy (7) for our purposes is established in Theorem 5.2 of [6], where additional geometric properties of it are noted. The proof is based on the following proposition, taken from [6], that expresses energy (7) in terms of the super level sets of u and f.

PROPOSITION 1. *The energy $E_1(u, \lambda)$ can be rewritten as follows:*

$$E_1(u, \lambda) = \int_{-\infty}^{\infty} \mathrm{Per}\big(\{x : u(x) > \mu\}\big) + \lambda \Big| \{x : u(x) > \mu\} \triangle \{x : f(x) > \mu\} \Big|\, d\mu. \tag{8}$$

Proof. The proof can be found in [6] (Proposition 5.1). □

We now recall also Theorem 5.2 of [6].

THEOREM 1. *If the observed image $f(x)$ is the characteristic function of a bounded domain $\Omega \subset \mathbf{R}^N$, then for any $\lambda \geq 0$ there is a minimizer of $E_1(\cdot, \lambda)$ that is also the characteristic function of a (possibly different) domain. In other words, when the observed image is binary, then for each $\lambda \geq 0$ there is at least one $u(x) \in M(\lambda)$ which is also binary.*

In fact, if $u_\lambda(x) \in M(\lambda)$ is any minimizer of $E_1(\cdot, \lambda)$, then for almost every $\gamma \in [0, 1]$ we have that the binary function

$$\mathbf{1}_{\{x : u_\lambda > \gamma\}}(x)$$

is also a minimizer of $E_1(\cdot, \lambda)$.

Proof. The proof can be found in [6] (Theorem 5.2). □

The proposition and its consequence, the theorem cited above from [6] (which are related to observations in [23, 24]), lead to a guaranteed algorithm for solving the binary image denoising problem (3), which we now state.

ALGORITHM 1. *To find a solution (i.e., a global minimizer) $u(x)$ of the nonconvex variational problem (3), it is sufficient to carry out the following three steps:*

1. *Find any minimizer of the* **convex** *energy (7); call it $v(x)$.*
2. *Let $\Sigma = \{x \in \mathbf{R}^N : v(x) > \mu\}$ for some $\mu \in (0, 1)$.*
3. *Set $u(x) = \mathbf{1}_\Sigma(x)$.*

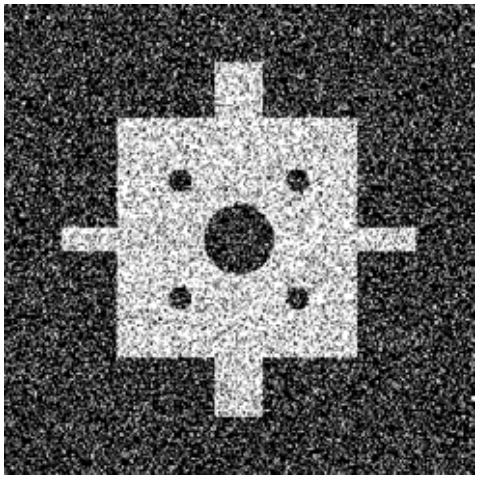

FIG. 2. *Original noisy binary image used in the numerical experiment of section* 3.

Then $u(x)$ *is a global minimizer of* (3) *for almost every choice of* μ.

Algorithm 1 reduces the *shape optimization* problem (4) to the *image denoising* problem (7). In section 4, we will obtain the analogue of Proposition 1 for the piecewise constant Mumford–Shah segmentation model of Chan and Vese, which will lead to a guaranteed algorithm for finding global minimizers of the more general shape optimization problem involved in that model, just like Algorithm 1 did for model (3); this is the content of Theorem 2 in that section. Our convex formulation will once again reduce the Chan–Vese shape optimization to a variant of the image denoising model (7).

The most involved step in the solution procedure described in Algorithm 1 is finding a minimizer of (7). One can approach this problem in many ways; for instance, one possibility is to simply carry out gradient descent.

Numerical example. The synthetic image of Figure 2 represents the given binary image $f(x)$, which is a simple geometric shape covered with random (binary) noise. The initial guess was an image composed of all 1's (an all white image). In the computation, the parameter λ was chosen to be quite moderate, so that in particular the small circular holes in the shape should be removed while the larger one should be kept. The result of the minimization is shown in Figure 3; in this case the minimizer is automatically very close to being binary, and hence the thresholding step of Algorithm 1 is almost unnecessary.

Figure 4 shows the histograms of intermediate steps during the gradient descent based minimization. As can be seen, the intermediate steps themselves are very far from being binary. The histogram in the lower right-hand corner belongs to the final result shown in Figure 3. Thus the gradient flow goes through nonbinary images, but in the end reaches another binary one. Although this is not implied by Proposition 1, Theorem 1, or Algorithm 1, it seems to hold in practice.

4. Piecewise constant segmentation. In this section, we extend the discussion of section 3 to the two-phase, piecewise constant Mumford–Shah segmentation model [17] of Chan and Vese [7]. Unlike in the previous section, this time we let the corrupted image $f(x)$ be nonbinary: it is merely assumed to be some measurable

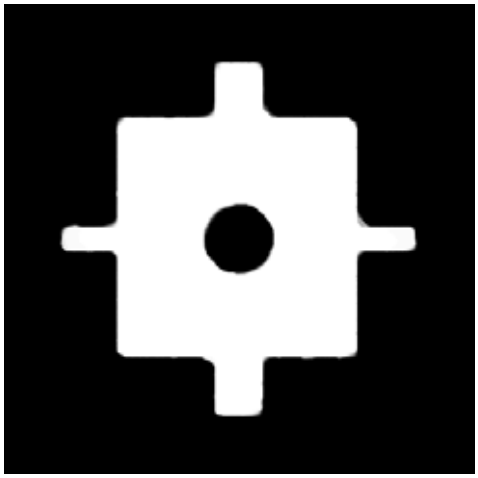

FIG. 3. *Final result found using the algorithm proposed in section* 3*, by minimizing* (7)*. Algorithm* 1 *says that global minimizers of the binary image denoising problem can be obtained by simply thresholding this result. In this case, the minimizer of energy* (7) *turns out to be very close to being binary itself, so there is no need to threshold. In the experiment, the value of* λ *was chosen small enough so that small holes in the original shape should get filled in, but also large enough so that the large hole in the middle should be maintained.*

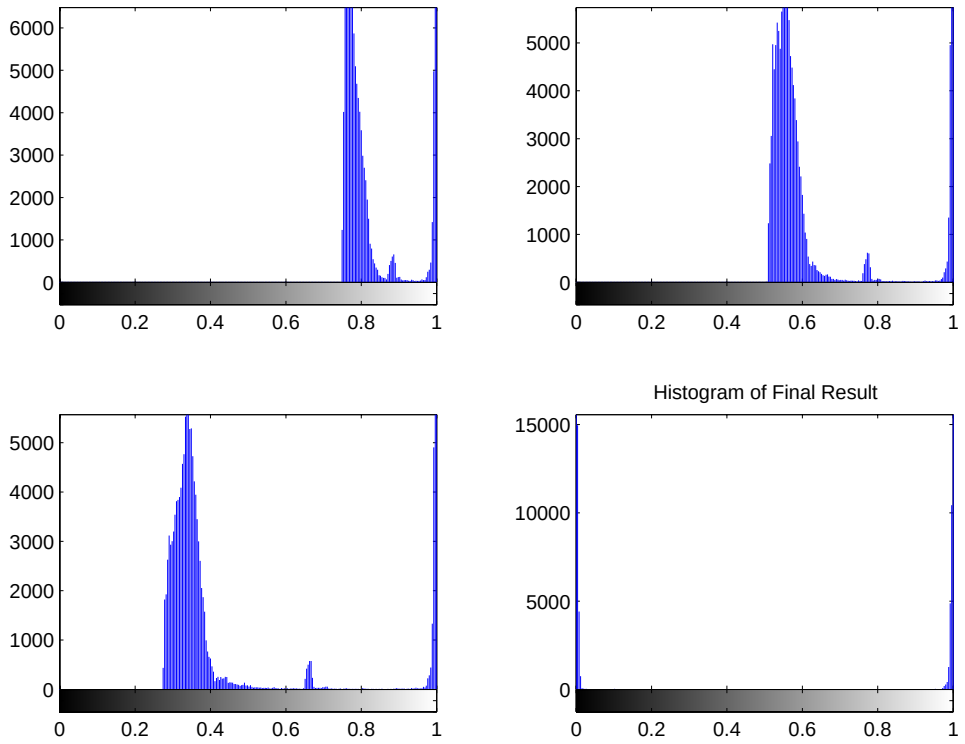

FIG. 4. *Histograms for intermediate images as the gradient descent proceeds. As can be seen, the intermediate images themselves are not binary; however, by the time the evolution reaches steady state, we are back to a binary image.*

function that takes its values in the unit interval. Thus, the discussion of this section supersedes that of the previous. Also, from now on we will assume that the image domain D is a bounded subset of $\mathbf{R}^N$ with Lipschitz boundary. The segmentation energy, which we will call MS, can then be written as

$$MS(\Sigma, c_1, c_2) := \mathrm{Per}(\Sigma; D) + \lambda \int_{\Sigma} (c_1 - f(x))^2 \, dx + \lambda \int_{D \setminus \Sigma} (c_2 - f(x))^2 \, dx. \tag{9}$$

The model says we should solve

$$\min_{\substack{c_1, c_2 \in \mathbf{R} \\ \Sigma \subset D}} MS(\Sigma, c_1, c_2). \tag{10}$$

This optimization problem can be interpreted to be looking for the best approximation in the L^2 sense to the given image $f(x)$ among all functions that take only two values. These values, denoted c_1, c_2, and where each is taken, namely, Σ and $D \setminus \Sigma$, are unknowns of the problem. As before, there is a penalty on the geometric complexity of the interface $\partial\Sigma$ that separates the regions where the two values c_1 and c_2 are taken. Functional (9) is nonconvex and can have more than one minimizer. Existence of at least one minimizer follows easily from standard arguments. Notice that if Σ is fixed, the values of c_1 and c_2 that minimize $MS(\Sigma, \cdot, \cdot)$ read

$$c_1 = \frac{1}{|\Sigma|} \int_{\Sigma} f(x) \, dx \quad \text{and} \quad c_2 = \frac{1}{|D \setminus \Sigma|} \int_{D \setminus \Sigma} f(x) \, dx. \tag{11}$$

A natural way to approximate the solution is a two-step scheme where in the first step one computes c_1 and c_2 according to these formulae, and in the second step updates the shape Σ. Even the minimization of $MS(\cdot, c_1, c_2)$ is a difficult problem since this functional is nonconvex. In what follows we focus on the minimization of $MS(\cdot, c_1, c_2)$.

We point out that if the two constants c_1 and c_2 are fixed to be 1 and 0, respectively, and if the given image $f(x)$ in (9) is taken to be the characteristic function $\mathbf{1}_{\Omega}(x)$ of a set Ω, then the minimization problem (10) reduces to the geometry problem (4); it is in this sense that this section's problem is a generalization of the previous section's.

Chan–Vese algorithm. In [7] Chan and Vese proposed a level set–based algorithm for solving the optimization problem (10). The idea is to represent the boundary $\partial\Sigma$ with the 0-level set of the function $\phi : D \to \mathbf{R}^N$. Energy (9) can then be written in terms of the level set function ϕ; it turns out to be

$$\begin{aligned} CV(\phi, c_1, c_2) = &\int_D |\nabla H_\varepsilon(\phi)| \\ &+ \lambda \int_D H_\varepsilon(\phi)(c_1 - f(x))^2 + (1 - H_\varepsilon(\phi))(c_2 - f(x))^2 \, dx. \end{aligned} \tag{12}$$

The function H_ε is, as before, a regularization of the Heaviside function. *The precise choice of the regularization H_ε of H is a crucial ingredient of the Chan–Vese algorithm.* We will return to this topic.

Variations of energy (12) with respect to the level set function ϕ lead to the following gradient descent scheme:

$$\phi_t = H'_\varepsilon(\phi) \left\{ \mathrm{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) - \lambda \left((c_1 - f(x))^2 - (c_2 - f(x))^2 \right) \right\}.$$

The optimal choice for the constants c_1, c_2 is easily determined in terms of the function ϕ.

The proposed algorithm. The Chan–Vese algorithm chooses a noncompactly supported, smooth approximation H_ε for H. As a result, the gradient descent equation given above and the following one have the same stationary solutions:

$$\phi_t = \operatorname{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right) - \lambda\left((c_1 - f(x))^2 - (c_2 - f(x))^2\right),$$

where we simply omitted the approximate Heaviside function altogether. This equation, in turn, is gradient descent for the following energy:

$$\int_D |\nabla\phi| + \lambda \int_D \left((c_1 - f(x))^2 - (c_2 - f(x))^2\right)\phi\, dx. \tag{13}$$

This energy is homogeneous of degree 1 in ϕ. As a result, it does not have a minimizer in general. In other words, the gradient descent written above does not have a stationary state: If the evolution is carried out for a long time, the level set function ϕ would tend to $+\infty$ wherever it is positive, and to $-\infty$ wherever it is negative. This issue is related to the nonuniqueness of representation with level sets and is easy to fix: one can simply restrict minimization to ϕ such that $0 \le \phi(x) \le 1$ for all $x \in D$. With this fix, and following [23, 24], we arrive at the statement below.

THEOREM 2. *For any given fixed $c_1, c_2 \in \mathbf{R}$, a global minimizer for $MS(\cdot, c_1, c_2)$ can be found by carrying out the following convex minimization:*

$$\min_{0\le u\le 1} \underbrace{\int_D |\nabla u| + \lambda \int_D \left\{(c_1 - f(x))^2 - (c_2 - f(x))^2\right\} u(x)\, dx}_{:=\tilde{E}(u,c_1,c_2)}$$

and then setting $\Sigma = \{x : u(x) \ge \mu\}$ for a.e. $\mu \in [0,1]$.

Proof. We once again rely on the coarea formula; since u takes its values in $[0,1]$, we have

$$\int_D |\nabla u| = \int_0^1 \operatorname{Per}(\{x : u(x) > \mu\}; D)\, d\mu.$$

For the other terms that constitute the fidelity term, we proceed as follows:

$$\begin{aligned}\int_D (c_1 - f(x))^2 u(x)\, dx &= \int_D (c_1 - f(x))^2 \int_0^1 \mathbf{1}_{[0,u(x)]}(\mu)\, d\mu\, dx \\ &= \int_0^1 \int_D (c_1 - f(x))^2 \mathbf{1}_{[0,u(x)]}(\mu)\, dx\, d\mu \\ &= \int_0^1 \int_{D\cap\{x:u(x)>\mu\}} (c_1 - f(x))^2\, dx\, d\mu.\end{aligned}$$

Also, we have

$$\begin{aligned}\int_D (c_2 - f(x))^2 u(x)\, dx &= \int_0^1 \int_{D\cap\{x:u(x)>\mu\}} (c_2 - f(x))^2\, dx\, d\mu \\ &= C - \int_0^1 \int_{D\cap\{x:u(x)>\mu\}^c} (c_2 - f(x))^2\, dx\, d\mu,\end{aligned}$$

where $C = \int_D (c_2 - f)^2\,dx$ is independent of u. Putting it all together, and setting $\Sigma(\mu) := \{x : u(x) > \mu\}$, we get the following formula that is valid for any $u(x) \in L^2(D)$ such that $0 \le u(x) \le 1$ for a.e. $x \in D$:

$$\begin{aligned}\tilde{E}(u, c_1, c_2) &= \int_0^1 \left\{ \mathrm{Per}(\Sigma(\mu); D) + \lambda \int_{\Sigma(\mu)} (c_1 - f(x))^2\,dx \right. \\ &\qquad \left. + \lambda \int_{D\setminus\Sigma(\mu)} (c_2 - f(x))^2\,dx \right\} d\mu - C \\ &= \int_0^1 MS\big(\Sigma(\mu), c_1, c_2\big)\,d\mu - C.\end{aligned}$$

It follows that if $u(x)$ is a minimizer of the convex problem, then for a.e. $\mu \in [0, 1]$ the set $\Sigma(\mu)$ has to be a minimizer of the original functional $MS(\cdot, c_1, c_2)$. ∎

Remark. The optimization problem that forms the content of Theorem 2 can be interpreted as follows: The level set formulation of the two-phase model depends on the level set function ϕ only through the term $H(\phi)$. The term $H(\phi)$ represents a parametrization of binary functions (since, for any given function ϕ, the function $H(\phi)$ is binary). So the minimization of (12) is thus a minimization over binary functions. Minimization of (13), on the other hand, corresponds to removing the nonconvex constraint of being binary; instead we minimize over functions that are allowed to take intermediate values. The content of the theorem above is that the minimizers (essentially) automatically satisfy the more stringent constraint.

We now turn to the question of how to minimize the convex problem stated in the theorem. In that connection, we have the following claim.

Claim 1. *Let $s(x) \in L^\infty(D)$. Then the convex, constrained minimization problem*

$$\min_{0\le u\le 1} \int_D |\nabla u| + \lambda \int_D s(x)u\,dx$$

has the same set of minimizers as the following convex, unconstrained *minimization problem:*

$$\min_u \int_D |\nabla u| + \int_D \alpha\nu(u) + \lambda s(x)u\,dx,$$

where $\nu(\xi) := \max\{0, 2|\xi - \frac{1}{2}| - 1\}$, provided that $\alpha > \frac{\lambda}{2}\|s(x)\|_{L^\infty(D)}$.

Proof. The term $\alpha\nu(u)$ that appears in the second, unconstrained minimization problem given in the claim is an *exact penalty* term [13, 14]; see Figure 5 for a plot of its graph. Indeed, the two energies agree for $\{u \in L^\infty(D) : 0 \le u(x) \le 1\ \forall x\}$. So we only need to show that any minimizer of the unconstrained problem automatically satisfies the constraint $0 \le u \le 1$. This is immediate: If $\alpha > \frac{\lambda}{2}\|s(x)\|_{L^\infty}$, then

$$|\lambda s(x)| \max\{|u(x)|, |u(x) - 1|\} < \alpha\nu(u(x)) \text{ whenever } u(x) \in [0, 1]^c,$$

which means that the transformation $u \to \min\{\max\{0, u\}, 1\}$ always decreases the energy of the unconstrained problem (strictly if $u(x) \in [0, 1]^c$ on a set of positive measure). That leads to the desired conclusion. ∎

Numerical examples. Here we detail how we obtained the numerical results pertaining to the two-phase piecewise constant segmentation models that are presented in this paper. Given c_1, c_2, the "exact penalty" formulation of the equivalent

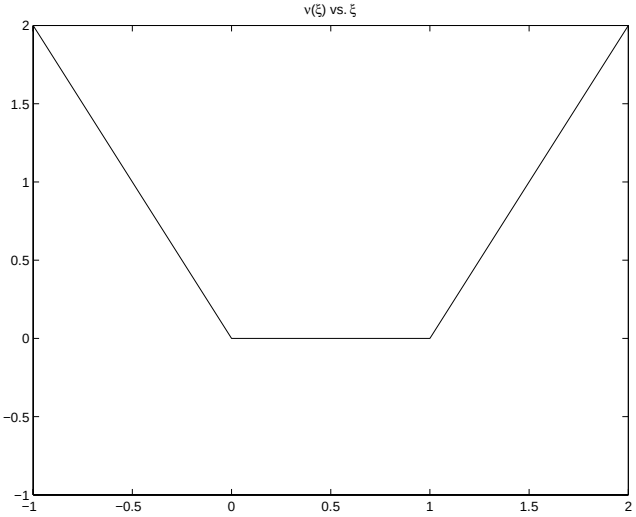

FIG. 5. *The function $\nu(\xi)$ is used for exact penalization as a method to impose the constraint $0 \le u \le 1$ in the minimization of Claim* 1.

minimization problem described above leads to the following Euler–Lagrange equation:

$$\operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) - \lambda s(x) - \alpha \nu'(u) = 0,$$

where $s(x) = (c_1 - f(x))^2 - (c_2 - f(x))^2$. The following explicit gradient descent scheme was used to solve the last equation:

$$\frac{u^{n+1} - u^n}{\delta t} = D_x^- \left(\frac{D_x^+ u^n}{\sqrt{(D_x^+ u^n)^2 + (D_y^+ u^n)^2 + \varepsilon_1}} \right) + D_y^- \left(\frac{D_y^+ u^n}{\sqrt{(D_x^+ u^n)^2 + (D_y^+ u^n)^2 + \varepsilon_1}} \right) - \lambda s(x) - \alpha \nu'_{\varepsilon_2}(u^n), \tag{14}$$

where $\varepsilon_1, \varepsilon_2 > 0$ are small constants, and $\nu_{\varepsilon_2}(\xi)$ is a regularized version of $\nu(\xi)$ that smooths the latter's kinks at 0 and 1.

The image shown in the Figure 6 is not piecewise constant with two regions; in fact it is not very well approximated by any image that takes only two values. This makes it a challenging test case for the two-phase segmentation problem (images that are already approximately two-valued are easily and very quickly segmented by these algorithms, and thus are easier examples).

Figure 7 shows the result found (i.e., the function u) using (14) to update the unknown function $u(x)$ that represents the two phases, when the given image $f(x)$ is the one shown in Figure 6. The two constants c_1 and c_2 were initially chosen to be 1 and 0, and updated occasionally according to (11); they eventually converged to 0.4666 and 0.0605, respectively. Although the considerations above (in particular, Theorem 2) do not imply that the minimizers of the objective functional turn out to be binary themselves (which would make the thresholding step in the algorithm

FIG. 6. *The given image $f(x)$ used in the two-phase segmentation numerical results discussed in section* 4.

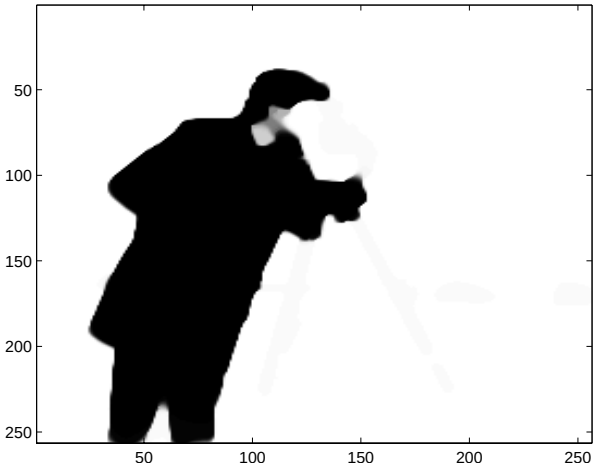

FIG. 7. *The solution $u(x)$ obtained by the numerical scheme of section* 4. *Although our claims do not imply the solution itself turns out to be binary, this seems to be always the case in practice. As can be seen, the computed solution is very close to being binary.*

of Theorem 2 unnecessary), in practice they seem to. Indeed, the image of Figure 7 is very close to being binary. Furthermore, it gets even closer to being binary if the computation is repeated using smaller values of the regularization parameters ε_1 and ε_2 that appear in scheme (14). On the other hand, it might be possible to cook up special given images f and special values λ for which there are nonbinary minimizers; for instance, in the case of the convex formulation (7) of the related geometry problem (4), the simple example of a disk as the given shape leads to nonbinary solutions for a specific choice of the parameter λ, as shown in [6].

Figure 8 displays the histograms of $u(x)$ at intermediate stages of the gradient descent computation. During the evolution, the function u certainly takes a continuum of values; however, as the steady state approaches, the values that u can take accumulate at the extreme ends of its allowed range. In this case, the extreme values

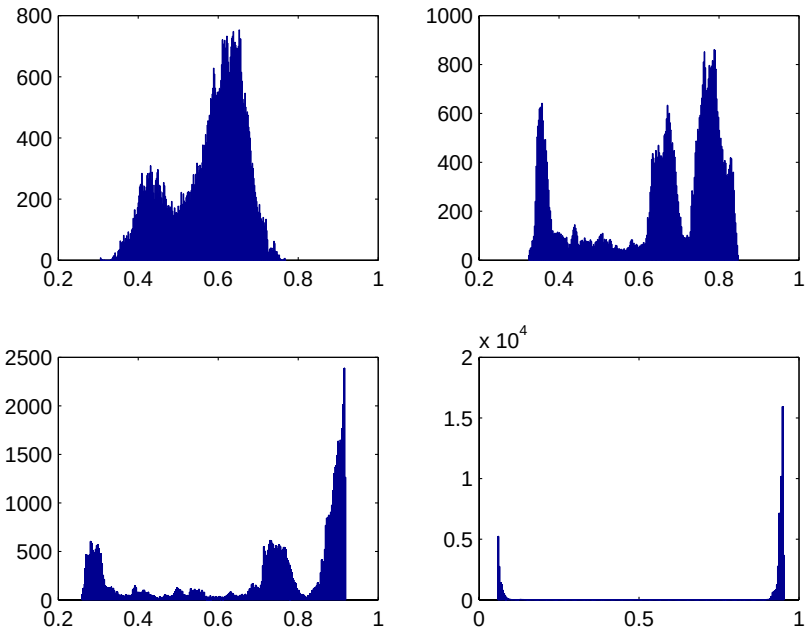

FIG. 8. *Histograms of intermediate solutions* $u^n(x)$ *of the flow in* (14)*, for the example of Figures* 6 *and* 7.

seem to be about 0.04 at the low end and 0.97 at the high end. They are not 0 and 1 because the exact penalty function ν that appears in (14) is regularized.

The theory above says that for fixed c_1 and c_2, all level sets of the function $u(x)$ are minimizers. The table below shows the value of energy (9) computed by taking $\Sigma = \{x : u(x) = \mu\}$ for several different values of μ, where $u(x)$ is the numerical example of Figures 6, 7, and 8, and the constants c_1 and c_2 have the values quoted above.

μ	Energy
0.2	17.7055
0.4	17.6458
0.5	17.6696
0.6	17.6655
0.8	17.6740

For comparison, we note that the energy of a disk centered at the middle of the image with radius a quarter of a side of the image domain has energy of about 112. Thus, even though there is some minor variation among different level sets of the function $u(x)$ (see Figure 9 for a plot of several level contours) and their corresponding energies (due to some of the approximations, such as in the penalty function μ, that were made to get a practical numerical algorithm), the difference in energy between them is quite small; they are all almost minimizers.

Acknowledgment. The authors would like to thank Prof. Robert V. Kohn, from whom they learned the observations of Strang in [23, 24].

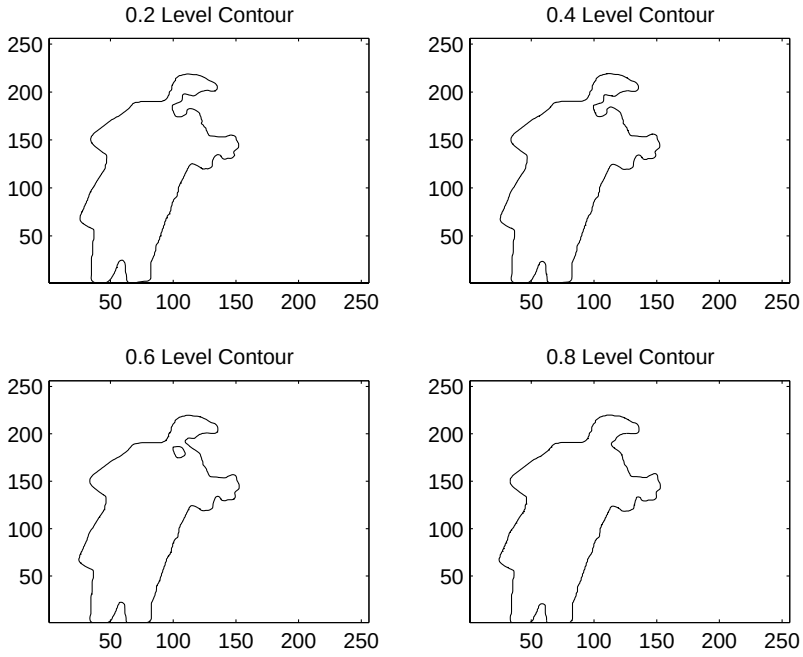

FIG. 9. *Plot of several level contours of the solution obtained. They are all very close to each other.*

REFERENCES

[1] S. ALLINEY, *Digital filters as absolute norm regularizers*, IEEE Trans. Signal Process., 40 (1992), pp. 1548–1562.

[2] S. ALLINEY, *Recursive median filters of increasing order: A variational approach*, IEEE Trans. Signal Process., 44 (1996), pp. 1346–1354.

[3] S. ALLINEY, *A property of the minimum vectors of a regularizing functional defined by means of the absolute norm*, IEEE Trans. Signal Process., 45 (1997), pp. 913–917.

[4] L. AMBROSIO AND V. M. TORTORELLI, *Approximation of functionals depending on jumps by elliptic functionals via Gamma convergence,* Comm. Pure Appl. Math., 43 (1990), pp. 999–1036.

[5] F. CATTÉ, F. DIBOS, AND G. KOEPFLER, *A morphological scheme for mean curvature motion and applications to anisotropic diffusion and motion of level sets*, SIAM J. Numer. Anal., 32 (1995), pp. 1895–1909.

[6] T. F. CHAN AND S. ESEDOḠLU, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math., 65 (2005), pp. 1817–1837.

[7] T. F. CHAN AND L. A. VESE, *Active contours without edges*, IEEE Trans. Image Process., 10 (2001), pp. 266–277.

[8] G. DAL MASO, *An Introduction to Γ-convergence*, Progr. Nonlinear Differential Equations Appl. 8, Birkhäuser Boston, Boston, MA, 1993.

[9] F. DIBOS AND G. KOEPFLER, *Global total variation minimization*, SIAM J. Numer. Anal., 37 (2000), pp. 646–664.

[10] L. C. EVANS AND R. F. GARIEPY, *Measure Theory and Fine Properties of Functions*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1992.

[11] W. FLEMING, W. RISHEL, AND R. RISHEL, *An integral formula for total gradient variation*, Arch. Math., 11 (1960), pp. 218–222.

[12] E. GIUSTI, *Minimal Surfaces and Functions of Bounded Variation*, Monogr. Math. 80, Birkhäuser Verlag, Basel, 1984.

[13] J.-B. HIRIART-URRUTY AND C. LEMARECHAL, *Convex Analysis and Minimization Algorithms. I. Fundamentals*, Grundlehren Math. Wiss. 305, Springer-Verlag, New York, 1993.

[14] J.-B. HIRIART-URRUTY AND C. LEMARECHAL, *Convex Analysis and Minimization Algorithms.*

II. *Advanced Theory and Bundle Methods*, Grundlehren Math. Wiss. 306, Springer-Verlag, New York, 1993.

[15] M. Kass, A. Witkin, and D. Terzopoulos, *Snakes: Active contour models*, Internat. J. Comput. Vision, 1 (1987), pp. 321–331.

[16] L. Modica and S. Mortola, *Un esempio di Gamma-convergenza*, Boll. Un. Mat. Ital. B, 14 (1977), pp. 285–299.

[17] D. Mumford and J. Shah, *Optimal approximations by piecewise smooth functions and associated variational problems*, Comm. Pure Appl. Math., 42 (1989), pp. 577–685.

[18] M. Nikolova, *A variational approach to remove outliers and impulse noise*, J. Math. Imaging and Vision, 20 (2004), pp. 99–120.

[19] M. Nikolova, *Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers*, SIAM J. Numer. Anal., 40 (2002), pp. 965–994.

[20] S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*, Appl. Math. Sci., 153, Springer-Verlag, New York, 2003.

[21] S. Osher and J. Sethian, *Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton–Jacobi formulations*, J. Comput. Phys., 79 (1988), pp. 12–49.

[22] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.

[23] G. Strang, *L^1 and L^∞ approximation of vector fields in the plane*, in Nonlinear Partial Differential Equations in Applied Science (Tokyo, 1982), North-Holland Math. Stud. 81, North-Holland, Amsterdam, 1983, pp. 273–288.

[24] G. Strang, *Maximal flow through a domain,* Math. Programming, 26 (1983), pp. 123–143.

[25] D. Strong and T. F. Chan, *Edge-preserving and scale-dependent properties of total variation regularization*, Inverse Problems, 19 (2003), pp. S165–S187.

[26] L. A. Vese and T. F. Chan, *A multiphase level set framework for image segmentation using the Mumford and Shah model*, Internat. J. Comput. Vision, 50 (2002), pp. 271–293.