Aspects of Total Variation Regularized L^1 Function Approximation[1]

Tony F. Chan and Selim Esedoḡlu
UCLA Mathematics Department

February 24, 2004

Abstract

The total variation based image denoising model of Rudin, Osher, and Fatemi has been generalized and modified in many ways in the literature; one of these modifications is to use the L^1 norm as the fidelity term. We study the interesting consequences of this modification, especially from the point of view of geometric properties of its solutions. It turns out to have interesting new implications for data driven scale selection and multiscale image decomposition.

AMS Subject Classification: 94A08, 65K10.
Keywords: total variation, denoising, scale space.

1 Introduction

Variational models for image reconstruction have had great success. One of the best known and influential examples is the total variation based model of Rudin, Osher, and Fatemi (ROF)[17]. This model and its variants have been a very active research topic. The idea behind the model is to exhibit the reconstructed image as the minimizer of the following energy:

$$\int_D |\nabla u| + \lambda \int_D (f-u)^2 \, dx \tag{1}$$

The functional is to be minimized over all $u \in L^2(D)$. Here D is a domain in $\mathbf{R}^N$ with Lipschitz boundary; it represents, for example, the computer screen. In this

[1]This work was supported in part by NSF contract DMS-9973341, NSF contract ACI-0072112, ONR contract N00014-03-1-0888, and NIH contract P20 MH65166.

paper, we will work with $D = \mathbf{R}^N$ for convenience. The function $f(x)$ represents the observed and possibly degraded image, and is taken to be in $L^2(D)$. The second integral in the functional is the *fidelity* term; it encourages the solution $u(x)$ that is being sought to approximate the *observed image* $f(x)$. The first integral in the functional is the *regularization* term; it is the essential novelty of the ROF model, as it allows for the reconstruction of images with discontinuities across hypersurfaces. Nevertheless, it disfavors oscillations and is responsible for the elimination of noise in applications to noisy images.

The standard ROF model (1) is well known to have certain limitations. One important issue is the loss of contrast in solutions even for noise free observed images. For example, Strong and Chan studied in [18] the case when the observed image $f(x)$ is a disk, and showed that the solution to (1, for any given λ, is of the form $cf(x)$, where $c \in [0, 1)$ is a constant. We never get $c = 1$, no matter how large the constant λ is chosen. More generally, given any observed image $f(x)$ and $\lambda > (2\|f\|_*)^{-1}$, it can be shown [10] for the corresponding solution $u(x)$ that $\|f - u\|_* = \frac{1}{2\lambda}$. Here, $\|\cdot\|_*$ denotes the dual norm of total variation. (See [10] for definition of the dual norm, and proofs of the statements just mentioned). It is in general desirable for image denoising algorithms to have a large class of "noise free" images that they leave invariant. For the standard ROF model, as these results show, that class consists of only the trivial image $f(x) := 0$.

Recently, work of Y. Meyer inspired research into understanding the role of the fidelity term better. It highlighted the fact that the choice of a suitable fidelity term can have far reaching consequences. For example, following up on Meyer's ideas Vese and Osher [20], and then Osher, Sole, and Vese [16] came up with variants of the original model that replace the fidelity term with weaker norms. It is shown in these works that this modification allows for much better separation of the high frequency component of images, such as noise and texture, from the piecewise smooth, or "cartoon", part.

In this paper, we ask related but rather different questions. We study a version of the ROF model that uses the L^1-norm as a measure of fidelity between the observed and denoised images. Given an observed image $f(x) \in L^1(\mathbf{R}^N)$, this model is based on the following variational problem

$$\inf_{u(x)\in BV(\mathbf{R}^N)} \int_{\mathbf{R}^N} |\nabla u| + \lambda \int_{\mathbf{R}^N} |u(x) - f(x)|\, dx \tag{2}$$

Our goal in this paper is to explore the consequences of this modest modification on the standard ROF model. In particular, we shall obtain some results that allow us to contrast the modified model (2) with the standard one (1). Also, the new understanding we develop about the nature of the scale space, lack of uniqueness of solutions, and lack of continuous dependence on data, will suggest applications

beyond mere removal of noise for the modified model: We will argue that some of these ordinarily undesirable characteristics can be real assets. Indeed, it turns out that the L^1 fidelity based model has many desirable, and some unexpected, consequences in applications such as multiscale image decomposition, and data driven parameter selection.

Some distinctions of the modified model (2) from the standard ROF model (1) are immediate:

- The way the fidelity and regularization terms scale with respect to each other in the modified and standard models is different. In particular, unlike the standard model, the modified model is contrast invariant in the following sense: If $u(x)$ is a solution of the modified model for the observed image $f(x)$, then $cu(x)$ is a solution of the modified model for the observed image $cf(x)$.

- The original model is strictly convex, and therefore its solution (the minimizer of the functional) is unique. The modified model is not strictly convex, leading to non-uniqueness of minimizers. This makes the scale space generated by the modified model qualitatively very different – and, as explained in Sections 6 and 7, for certain purposes more suitable – than that of the standard ROF model.

We concentrate especially on the scale space and geometric features of the decomposition technique derived from this model. The analytical and numerical results presented in this paper suggest the following major advantages of the L^1 fidelity based model over the standard one:

- The regularization imposed on solutions by the L^1 model is more geometric. This means that the regularization process has less dependence on the contrast of image features than on their geometry. Indeed, as some of our analytical results show, the L^1 model almost decouples the level sets of the given image from each other and treats them independently of their associated level (grayscale value).

- As distinct from the standard model, small features in the image maintain their contrast even as the fidelity parameter λ is lowered, maintaining good contrast until they suddenly disappear.

- An unexpected consequence of the modification is that it suggests a data driven scale selection technique: it seems possible to identify certain critical values of the parameter λ at which features at the corresponding scale go through a discontinuous change.

The ROF model with L^1 fidelity was introduced and studied in the context of image denoising and deblurring by many previous authors, for example by Alliney and Nikolova in [1, 2, 3, 11, 12, 13]. Alliney's previous work involves the variational model (2) in only one space dimension; moreover his results are restricted to the discrete versions of the energy. Nevertheless, many of his observations are directly relevant to our results (see, for instance, Proposition 1 that we quote from his work), and some of our results (for instance part of Theorem 1) can be thought of as continuum analogues of his results in arbitrary dimensions. In [11] Nikolova shows that for certain types of noise the total variation regularization with L^1 fidelity outperforms the standard model. And [12] contains many impressive numerical results that clearly demonstrate the advantages of using the L^1 norm for fidelity term in some applications. In fact, the analysis presented in [11] applies more generally to fidelity terms that are, like the L^1 fidelity term and unlike the L^2 fidelity term, non-differentiable at the origin. The techniques of Nikolova also allow her to study certain typical properties of minimizers to the ROF model and its variants with different types of fidelity terms. For example, among the results is a characterization of the staircasing effect. Moreover, she calls attention to the fact that with L^1 type fidelity terms, the solution reconstructs the given image exactly at some pixels; this relates to the contrast preserving property we touched on above. However, unlike the focus of this paper, results in [11, 12] mostly concern discrete versions of the denoising energies and depend on the discretization size; continuum analogues are not treated. Our focus in this paper is squarely on the continuum energies so that we can study geometric properties of their minimizers independently of the discretization.

We conclude the introduction with an outline of the remaining sections. Section 2 introduces the notation that is used throughout the paper. Section 3 works out the solution to minimization problems (1) and (2) in the simple case when the observed image $f(x)$ is the characteristic function of a disk in two dimensions. This illustrates some of the results obtained in subsequent sections for more general types of images. Section 4 consists of a collection of simple but useful facts that follow immediately from the definitions of Section 2; these get used in the following sections of the paper. Section 5 deals with properties of minimizers of energy (2). In particular, it considers the case where the observed image is the characteristic function of a bounded set. It recalls the known results for standard ROF model in this case, and uses them for comparison. Section 6 elaborates on the differences between the scale spaces generated by the two models given by (1) and (2); it shows that the model based on L^1 fidelity makes it possible to determine special values of the parameter λ completely from the given observed image. Finally, Section 7 presents numerical experiments and gives some implementation details. The numerical results corroborate the overall picture suggested by the analytical

results of the previous sections.

2 Notation

In this section we introduce notation that will be used throughout the paper to compare the original ROF model 1 with the modified one 2 that uses L^1-fidelity term. First, we recall the standard definitions of total variation of a function and the perimeter of a set [8, 9]. The total variation of a function $u(x) \in L^1_{loc}(\mathbf{R}^N)$ is defined to be

$$\int_{\mathbf{R}^N} |\nabla u(x)| := \sup_{\substack{\phi \in C_c^1(\mathbf{R}^N;\mathbf{R}^N) \\ |\phi(x)| \leq 1 \forall x \in \mathbf{R}^N}} - \int_{\mathbf{R}^N} u(x) \operatorname{div} \phi(x)\, dx.$$

The perimeter of a set $\Sigma \subset \mathbf{R}^N$ is defined in terms of the above definition to be

$$\text{Per}(\Sigma) := \int_{\mathbf{R}^N} |\nabla \mathbf{1}_\Sigma(x)|.$$

For a given possibly noisy image $f(x) \in L^1(\mathbf{R}^N)$, we will denote the energy of the total variation model with L^1 fidelity $E_1(u, \lambda)$:

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

It will be compared, for $f \in L^1(\mathbf{R}^N) \cap L^2(\mathbf{R}^N)$, with the energy of the standard ROF model, which we denote $E_2(u, \lambda)$:

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

Of particular interest are the minimum values of these energies as a function of the parameter λ:

$$\mathcal{E}_1(\lambda) := \min_{u \in L^1(\mathbf{R}^N)} E_1(u, \lambda),$$
$$\mathcal{E}_2(\lambda) := \min_{u \in L^2(\mathbf{R}^N)} E_2(u, \lambda).$$

Minimizers of the standard ROF energy $E_2(\cdot, \lambda)$ for a fixed λ are unique; this is a consequence of its strict convexity. Minimizers of the modified energy $E_1(\cdot, \lambda)$

need not be unique in general. We therefore introduce the following notation to denote the set of minimizers of $E_1(\cdot,\lambda)$ at a given $\lambda \geq 0$:

$$M(\lambda) := \Big\{ u \in L^1(\mathbf{R}^N) : E_1(u,\lambda) = \mathcal{E}_1(\lambda) \Big\}.$$

For any given $f(x) \in L^1(\mathbf{R}^N)$ and $\lambda \geq 0$, the set $M(\lambda)$ is non-empty: a standard argument shows the existence of minimizers. Because of non-uniqueness, $M(\lambda)$ can have several elements. Different elements of $M(\lambda)$ can stand at different distances to the observed image $f(x)$. This motivates the following notation:

$$\mu^+(\lambda) := \sup \Big\{ \|f-u\|_{L^1(\mathbf{R}^N)} : u \in M(\lambda) \Big\},$$
$$\mu^-(\lambda) := \inf \Big\{ \|f-u\|_{L^1(\mathbf{R}^N)} : u \in M(\lambda) \Big\}.$$

The values of the parameter λ at which $M(\lambda)$ contains elements whose distances to the given image $f(x)$ are different turn out to be special. We therefore adopt the following notation to denote this set of special λ values:

$$S(f) := \Big\{ \lambda \in \mathbf{R}^+ : \mu^-(\lambda) \neq \mu^+(\lambda) \Big\}.$$

To emphasize the dependence of $E_i(\cdot,\lambda)$, $\mathcal{E}_i(\lambda)$,$M(\lambda)$, and $\mu^{\pm}(\lambda)$ on the observed image $f(x)$ in addition to λ, we will write $E_i(\cdot,\lambda,f), \mathcal{E}_i(\lambda,f)$, $M(\lambda,f)$, and $\mu^{\pm}(\lambda,f)$ whenever necessary.

3 An example

In this section we consider a very simple but illustrative example. Namely, we work out explicitly the solution to the problem of minimizing the two dimensional version of $E_1(\cdot,\lambda)$ in case when the observed image $f(x)$ is given by the characteristic function $\mathbf{1}_{B_r(0)}(x)$ of a disk $B_r(0)$ that is centered at the origin and with radius r. It is important to compare the result with the one for the standard ROF model, which – as we noted in the introduction – was calculated in [18].

We start by recalling the calculation of [18]. For $\lambda \geq 0$ and the observed image given by $f(x) = \mathbf{1}_{B_r(0)}(x)$, the unique minimizer $u_\lambda(x)$ of $E_2(\cdot,\lambda)$ is given by:

$$u_\lambda(x) \equiv \begin{cases} 0 & \text{if } 0 \leq \lambda \leq \dfrac{1}{r}, \\ \left(1 - \dfrac{1}{\lambda r}\right) \mathbf{1}_{B_r(0)}(x) & \text{if } \lambda > \dfrac{1}{r}. \end{cases}$$

Turning now to the case of $E_1(\cdot,\lambda)$, one can reason (for example with the help of some of the results presented in Sections 5 and 6 of this paper) that for each $\lambda \geq 0$,

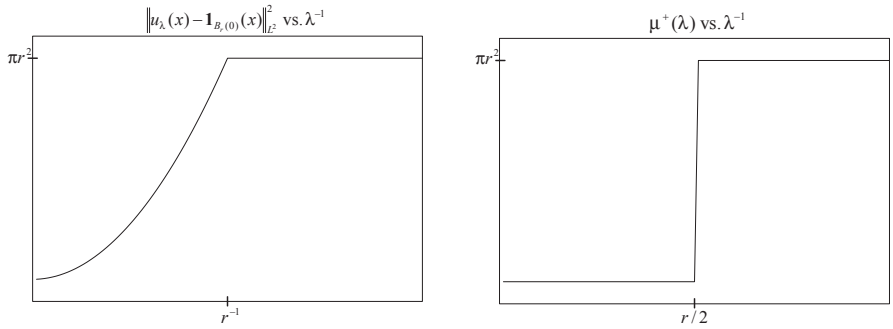

Figure 1: **Left:** Plot of $\|u_\lambda(x) - f(x)\|^2_{L^2}$ vs λ^{-1} for the example of Section 3, where $u_\lambda(x)$ denotes the unique minimizer of $E_2(\cdot, \lambda)$. **Right:** Plot of $\mu^+(\lambda)$ vs. λ^{-1} for the ROF model with L^1 fidelity, using the example of Section 3.

every minimizer has to be of the form $c\mathbf{1}_{B_r(0)}(x)$ for some constant $c \in [0, 1]$. We therefore need to minimize the function

$$E_1(c\mathbf{1}_{B_r(0)}(x), \lambda) = 2\pi rc + \lambda\pi r^2|1 - c|$$

over $c \in [0, 1]$. We get

$$M(\lambda) = \begin{cases} \{0\} & \text{if } 0 \le \lambda < \dfrac{2}{r}, \\ \left\{c\mathbf{1}_{B_r(0)}(x) : c \in [0,1]\right\} & \text{if } \lambda = \dfrac{2}{r}, \\ \left\{\mathbf{1}_{B_r(0)}(x)\right\} & \text{if } \lambda \ge \dfrac{2}{r}. \end{cases}$$

Thus, we see that the solution is unique for all except one special value of the parameter λ. The special value is related to radius of the disk; for more general images we would expect such special values of the parameter λ to be related to the geometric scale of distinct objects contained in the scene.

The difference between scale spaces generated by the standard and modified ROF models is made abundantly clear by this simple example. When L^1 fidelity is used, unlike in the standard ROF model, the scale space is mostly constant; it only makes a sudden transition at a special value of the scale parameter. This difference can also be manifested by plotting the "fidelity of minimizer" as a function of the parameter λ for each model and comparing the qualitative properties. Figure 1 shows the plots obtained based on the minimizers calculated above.

This example brings out another elementary aspect of using an L^1 fidelity term with total variation regularization. Fix a $\lambda > 0$. Then, the unique minimizer of

$E_1(\cdot, \lambda)$ with the observed image $f(x) = \mathbf{1}_{B_r(0)}(x)$ is identically 0 if $r < \frac{2}{\lambda}$, but $\mathbf{1}_{B_r(0)}(x)$ if $r > \frac{2}{\lambda}$. Thus the dependence of the solution to the L^1 model on the observed image is not continuous with respect to, say, the L^1 norm. This is clearly related to the lack of uniqueness in solutions to the model, and is a price to pay for having solutions in which features of interest maintain good contrast until they are completely eliminated. However, Sections 6 and 7 explain some applications for which such a discontinuity can be actually desirable, and Proposition 5 shows that certain important features of the *scale space* are continuous as a function of observed signal.

4 Basic facts

In this section, we collect a number of elementary facts that follow immediately from the definitions introduced in the previous section. These results will be useful in the subsequent sections.

The following claim shows that the minimum energies $\mathcal{E}_i(\lambda)$ are well-behaved functions of the parameter λ:

Claim 1 *For any given observed image $f(x) \in L^1(\mathbf{R}^N)$ the function $\mathcal{E}_1(\lambda)$, and for any given observed image $f(x) \in L^2(\mathbf{R}^N)$ the function $\mathcal{E}_2(\lambda)$ satisfy the following properties:*

1. *$\mathcal{E}_i(\lambda)$ for $i = 1, 2$ are increasing and concave.*
2. *$\mathcal{E}_i(0) = 0$ for $i = 1, 2$.*
3. *$0 \leq \mathcal{E}_1(\lambda) \leq \|f\|_{L^1}\lambda$ and $0 \leq \mathcal{E}_2(\lambda) \leq \|f\|_{L^2}^2\lambda$ for all $\lambda \in [0, \infty)$.*
4. *$\mathcal{E}_i(\lambda)$ are Lipschitz continuous for $i = 1, 2$.*

Proof: $\mathcal{E}_i(\lambda)$ are defined as pointwise infema of a collection of linear functions that are increasing in λ; this makes them increasing and concave. Statements *2* and *3* follow from the trivial fact that $\mathcal{E}_i(\lambda) \leq E_i(0, \lambda)$ for $i = 1, 2$. Statement *4* now follows from the first three. □

Claim 2 *The set $M(\lambda)$ is closed and convex.*

Proof: This follows from convexity of the energy $E_{\mathbf{1}}$. □

The following claim shows that the fidelity of the minimizer to the original ROF model varies continuously as a function of λ. This should be contrasted with the results for the L^1 model that are obtained in the subsequent sections.

Claim 3 *Given $f(x) \in L^2(\mathbf{R}^N)$, for each $\lambda \geq 0$ let $u_\lambda(x)$ denote the unique minimizer of $E_2(\cdot, \lambda)$. Then the function $\lambda \to \|f - u_\lambda\|_{L^2}$ is continuous.*

Proof: Fix $\lambda_* \geq 0$ and let $u_{\lambda_*}(x)$ be the unique minimizer of $E_2(\cdot, \lambda_*)$. Let $\{\lambda_j\}_j^\infty \subset \mathbf{R}^+$ converge to λ_*. Consider the sequence of corresponding minimizers $\{u_{\lambda_j}\}$. The obvious relation $E_2(u_{\lambda_j}, \lambda_j) \leq E_2(0, \lambda_j) = \lambda_j \|f\|_{L^2}^2$ implies that the sequence has uniformly bounded total variation and L^2-norm. It also implies that $\|u_\lambda - f\|_{L^2} \leq \|f\|_{L^2}$ for every $\lambda \geq 0$. We may therefore pass to a subsequence, also denoted $\{u_{\lambda_j}\}$, and find a $v \in L^2(\mathbf{R}^N)$ such that $u_{\lambda_j}(x) \to v(x)$ pointwise a.e. Note that by Fatou $\|v - f\|_{L^2} \leq \liminf_{j\to\infty} \|u_{\lambda_j} - f\|_{L^2}$. Furthermore, by the standard lower semi-continuity result we have $\int |\nabla v| \leq \liminf_{j\to\infty} \int |\nabla u_{\lambda_j}|$. Hence we get that $E_2(v, \lambda_*) \leq \liminf_{j\to\infty} E_2(u_{\lambda_j}, \lambda_j)$.

On the other hand, $E_2(u_{\lambda_*}, \lambda_*) \geq \limsup_{j\to\infty} E_2(u_{\lambda_j}, \lambda_j)$. To see this, suppose not. Then there is $\varepsilon > 0$ and arbitrarily large j such that $E_2(u_{\lambda_*}, \lambda_*) \leq E_2(u_{\lambda_j}, \lambda_j) - \varepsilon$. But also, $\lim_{j\to\infty} E_2(u_{\lambda_*}, \lambda_j) = E_2(u_{\lambda_*}, \lambda_*)$. These two statements mean $E_2(u_{\lambda_*}, \lambda_j) < E_2(u_{\lambda_j}, \lambda_j)$ for some large j, which is a contradiction since u_{λ_j} are supposed to be minimizers of $E_2(\cdot, \lambda_j)$. This, along with the remarks of the previous paragraph, adds up to the following conclusion:

$$\limsup_{j\to\infty} E_2(u_{\lambda_j}, \lambda_j) \leq E_2(u_{\lambda_*}, \lambda_*) \leq E_2(v, \lambda_*) \leq \liminf_{j\to\infty} E_2(u_{\lambda_j}, \lambda_j).$$

We thus see that v is a minimizer of $E_2(\cdot, \lambda_*)$; by uniqueness of minimizers of $E_2(\cdot, \lambda_*)$, we get that $v = u_{\lambda_*}$.

If $\lambda_* = 0$, then $u_{\lambda_*} = 0$ and so $\|u_{\lambda_*} - f\|_{L^2} = \|f\|_{L^2}$. Recalling from above that $\|u_\lambda - f\|_{L^2} \leq \|f\|_{L^2}$ for all λ, we see that in this case

$$\limsup_{j\to\infty} \|u_{\lambda_j} - f\|_{L^2} \leq \|u_{\lambda_*} - f\|_{L^2} \leq \liminf_{j\to\infty} \|u_{\lambda_j} - f\|_{L^2}$$

which establishes continuity of the map in question at $\lambda = 0$.

If $\lambda_* > 0$, we reason as follows: We must once again have $\limsup_{j\to\infty} \|u_{\lambda_j} - f\|_{L^2} \leq \|u_{\lambda_*} - f\|_{L^2}$, which immediately leads to the conclusion of the claim. To see this, we suppose it's false and proceed as we did in the previous paragraphs. There is then arbitrarily large j and an $\varepsilon > 0$ such that $\|u_{\lambda_*} - f\|_{L^2} \leq \|u_{\lambda_j} - f\|_{L^2} - \varepsilon$. But then

$$E_2(u_{\lambda_*}, \lambda_*) \leq \liminf_{j\to\infty} E_2(u_{\lambda_j}, \lambda_j) - \varepsilon\lambda_j.$$

Also, $E_2(u_{\lambda_*}, \lambda_j) \to E_2(u_{\lambda_*}, \lambda_*)$ as $j \to \infty$. These last two statements lead as before to the contradictory statement that $E_2(u_{\lambda_*}, \lambda_j) < E_2(u_{\lambda_j}, \lambda_j)$. □

We will see whether the analogue of Claim 3 holds for E_1. In that regard, we first make the following basic observation:

Claim 4 *Let $\lambda_2 > \lambda_1 \geq 0$, and assume that u_{λ_1} and u_{λ_2} are any two minimizers of $E_1(\cdot, \lambda_1)$ and $E_1(\cdot, \lambda_2)$ respectively. Then:*

$$\|u_{\lambda_1} - f\|_{L^1(\mathbf{R}^N)} \geq \|u_{\lambda_2} - f\|_{L^1(\mathbf{R}^N)}.$$

Proof: Suppose $\|u_{\lambda_2} - f\|_{L^1} > \|u_{\lambda_1} - f\|_{L^1}$. Then, since $u_{\lambda_1} \in M(\lambda_1)$, we have $E_1(u_{\lambda_1}, \lambda_1) \leq E_1(u_{\lambda_2}, \lambda_1)$. We then have

$$\begin{aligned} E_1(u_{\lambda_1}, \lambda_2) &= E_1(u_{\lambda_1}, \lambda_1) + (\lambda_2 - \lambda_1)\|u_{\lambda_1} - f\|_{L^1} \\ &\leq E_1(u_{\lambda_2}, \lambda_1) + (\lambda_2 - \lambda_1)\|u_{\lambda_1} - f\|_{L^1} \\ &< E_1(u_{\lambda_2}, \lambda_1) + (\lambda_2 - \lambda_1)\|u_{\lambda_2} - f\|_{L^1} \\ &= E_1(u_{\lambda_2}, \lambda_2). \end{aligned}$$

which is a contradiction, since $u_{\lambda_2} \in M(\lambda_2)$ by hypothesis. □

Corollary 1 *The functions $\mu^{\pm}(\lambda)$ are decreasing. In fact,*

$$\mu^-(\lambda_1) \leq \mu^+(\lambda_1) \leq \mu^-(\lambda_2) \leq \mu^+(\lambda_2)$$

whenever $\lambda_1 > \lambda_2 \geq 0$.

The functions $\mu^{\pm}(\lambda)$ are the analogue for E_1 of $\|u_\lambda - f\|_{L^2}$ in Claim 3. These functions in general can be discontinuous; in fact their set of discontinuity is precisely $S(f)$ according to our notation. The Corollary above allows us to make the following simple statement about the discontinuities of these functions:

Claim 5 *For any given $f \in L^1(\mathbf{R}^N)$, the set $S(f)$ is at most countable.*

Proof: If $\lambda \in S(f)$, then $\mu^-(\lambda) < \mu^+(\lambda)$. By the corollary above, at such a λ both μ^- and μ^+ have a jump discontinuity. The set of discontinuities of a monotone function are at most countable. □

Finally, for completeness let us state the following rather obvious fact about the asymptotic value of the functions $\mu^{\pm}(\lambda)$ as $\lambda \to \infty$:

Claim 6 *Given $f(x) \in L^1(\mathbf{R}^N)$, we have $\lim_{\lambda\to\infty} \mu^{\pm}(\lambda) = 0$.*

Proof: Given $\varepsilon > 0$ we can find $f_\varepsilon(x) \in BV(\mathbf{R}^N)$ such that $\|f_\varepsilon - f\|_{L^1} \leq \frac{\varepsilon}{2}$. If $u_\lambda(x) \in M(\lambda)$ with $\mu^+(\lambda) = \|u_\lambda - f\|_{L^1}$, then

$$\mu^-(\lambda) \leq \mu^+(\lambda) \leq \frac{1}{\lambda}E_1(u_\lambda, \lambda) \leq \frac{1}{\lambda}E_1(f_\varepsilon, \lambda) \leq \frac{1}{\lambda}\int |\nabla f_\varepsilon| + \frac{\varepsilon}{2}.$$

Hence, for all large enough λ we have $\mu^{\pm}(\lambda) \leq \varepsilon$. $\square$

The following fact is taken directly from [2]. It says that any image $u_*(x)$ which arises as the solution to model (2) for *some* observed image $f(x)$ is in fact also the solution to model (2) with observed image $f(x)$ taken to be $u_*(x)$ itself provided that the parameter λ is taken large enough. We include it as a good way to emphasize the difference of model (2) from (1) in regard to the loss of contrast in solutions.

Proposition 1 *Let $\lambda_* \geq 0$, $f(x) \in L^1(\mathbf{R}^N)$, and $u_*(x) \in M(\lambda_*, f)$. Then for every $\lambda \geq \lambda_*$ we have $u_*(x) \in M(\lambda, u_*)$.*

Proof: See [2].

5 Minimizers of E_1

In this section, we study the behavior of the ROF model with L^1 fidelity on simple images. Our motivation is twofold. First, studying the behavior of image denoising models on simple images is a first step towards understanding the type of images they can successfully process. Second, this type of question allows us to compare different models. And in fact, we will stress the difference of these results from the analogous ones obtained for the standard ROF model by previous authors. In particular, our results will bolster the intuitive observation that the L^1 fidelity term leads to more geometric regularizations.

The following proposition constitutes our starting point. It shows that the ROF model with L^1 fidelity term almost decouples the level sets of the given image from each other; it almost becomes a geometry problem for each level set, independent of the level.

Proposition 2 *The energy $E_1(u, \lambda)$ can be rewritten as follows:*

$$E_1(u,\lambda) = \int_{-\infty}^{\infty} \mathit{Per}\left(\{x : u(x) > \mu\}\right) + \Big|\{x : u(x) > \mu\} \triangle \{x : f(x) > \mu\}\Big| \, d\mu \quad (3)$$

Proof: Recall the coarea formula for functions of bounded variation (see [9] or [8]):

$$\int_{\mathbf{R}^N} |\nabla u| = \int_{-\infty}^{\infty} \mathrm{Per}\big(\{x : u(x) > \mu\}\big)\, d\mu \quad (4)$$

Also, there is the following "layer cake" formula:

$$\begin{aligned}\int_{\mathbf{R}^N} |u-f|\,dx &= \int_{\{u>f\}} |u-f|\,dx + \int_{\{f>u\}} |u-f|\,dx \\ &= \int_{\{u>f\}} \int_{f(x)}^{u(x)} d\mu\,dx + \int_{\{f>u\}} \int_{u(x)}^{f(x)} d\mu\,dx \\ &= \int_{\mathbf{R}^N} \int_{\mathbf{R}} \mathbf{1}_{\{u>f\}}(x)\mathbf{1}_{[f(x),u(x))}(\mu) + \mathbf{1}_{\{f>u\}}(x)\mathbf{1}_{[u(x),f(x))}(\mu)\,d\mu\,dx \\ &= \int_{\mathbf{R}} \int_{\mathbf{R}^N} \mathbf{1}_{\{u>f\}}(x)\mathbf{1}_{[f(x),u(x))}(\mu) + \mathbf{1}_{\{f>u\}}(x)\mathbf{1}_{[u(x),f(x))}(\mu)\,dx\,d\mu\end{aligned}$$

where we simply changed the order of integration in the last step. But now we have:

$$\mathbf{1}_{\{u>f\}}(x)\mathbf{1}_{[f(x),u(x))}(\mu) = 1 \text{ iff } x \in \{u>f\} \cap \{u>\mu\} \cap \{f>\mu\}^c$$

and 0 otherwise, and

$$\mathbf{1}_{\{f>u\}}(x)\mathbf{1}_{[u(x),f(x))}(\mu) = 1 \text{ iff } x \in \{f>u\} \cap \{u>\mu\}^c \cap \{f>\mu\}$$

and 0 otherwise. That means

$$\mathbf{1}_{\{u>f\}}(x)\mathbf{1}_{[f(x),u(x))}(\mu) + \mathbf{1}_{\{f>u\}}(x)\mathbf{1}_{[u(x),f(x))}(\mu) = \mathbf{1}_{\{u>\mu\}\triangle\{f>\mu\}}(x)$$

Therefore

$$\int_{\mathbf{R}^N} |u-f|\,dx = \int_{-\infty}^{\infty} |\{x : u(x)>\mu\} \triangle \{x : f(x)>\mu\}|\,d\mu$$

Putting these formulae together gives the one in the statement of the claim. □

We now explore some consequences of Proposition 2. First, we consider what happens when the observed image is binary. In other words, we assume that $f(x)$ is the characteristic function of a domain. We assume that the domain is bounded, but for now make no assumptions about the boundary of the domain.

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 $\mu \in [0,1]$ we have that the binary function

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

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

Proof: Let $f(x) := \mathbf{1}_\Omega(x)$, where Ω is a bounded domain in $\mathbf{R}^N$. It can be easily seen that any minimizer $u(x)$ of E_1 satisfies $u(x) \in [0,1]$ for a.e. $x \in \mathbf{R}^N$. Formula (3) of Proposition 2 above becomes in this case:

$$E_1(u,\lambda) = \int_0^1 \text{Per}\left(\{x : u(x) > \mu\}\right) + \lambda \left| \{x : u(x) > \mu\} \triangle \Omega \right| d\mu$$

This suggests we consider for each level set of $u(x)$ the following geometry problem:

$$\min_{\Sigma \subset \mathbf{R}^N} \Big(\text{Per}(\Sigma) + \lambda \big| \Sigma \triangle \Omega \big| \Big). \tag{5}$$

Standard compactness and lower semi-continuity facts show the existence of minimizers; let $\Sigma_* \subset \mathbf{R}^N$ be one of them. Let $u_\lambda(x)$ be any minimizer of $E_1(\cdot,\lambda)$, i.e. $u_\lambda(x) \in M(\lambda)$. Set

$$\Sigma(\mu) := \big\{ x : u(x) > \mu \big\}.$$

Then,

$$\text{Per}\big(\Sigma(\mu)\big) + \lambda \big| \Sigma(\mu) \triangle \Omega \big| \geq \text{Per}(\Sigma_*) + \lambda \big| \Sigma_* \triangle \Omega \big| \tag{6}$$

for a.e. $\mu \in [0,\infty)$. This now immediately implies that

$$E_1(u_\lambda(x),\lambda) \geq E_1(\mathbf{1}_{\Sigma_*}(x),\lambda)$$

which means that $\mathbf{1}_{\Sigma_*}(x)$ is also a minimizer of $E(\cdot,\lambda)$.

Furthermore, since $u_\lambda(x)$ is a minimizer, the inequality of (6) is in fact an equality for a.e. $\mu \in [0,1]$. Thus, $\Sigma(\mu)$ is a minimizer of the geometry problem (5) and $1_{\Sigma(\mu)}(x)$ is a minimizer of $E_1(\cdot,\lambda)$ for a.e. μ. □

Remark: A version of the first statement of Theorem 1 was obtained for the discrete analogue of model (2) in one space dimension by Alliney in [3]. □

Remark: The claim leaves open the possibility that for a given $\lambda \geq 0$ there might be $u \in M(\lambda)$ that takes more than two values. □

Remark: The conclusion of Theorem 1 is interesting because it establishes the equivalence of a non-convex problem (the geometry problem of minimizing over only binary images, which is encountered in many applications such as improving the appearance of fax documents) to a convex problem (minimizing over all images). Indeed, it follows form the corollary that to obtain a solution to (5), one can first minimize $E_1(\cdot,\lambda)$ taking $f(x) = 1_\Omega(x)$ as the observed image, and then look at a level set of the solution obtained. Whether this observation can be turned into a useful computational tool needs to be explored, but this question will not be pursued any further here. □

The previous two claims highlight an important *qualitative difference* of the L^1 model from the standard ROF model. In contrast to the content of these claims, it is

easy to show that for certain types of binary images (even with smooth edge sets) the minimizer of the standard ROF model takes more than two values for every large enough choice of the parameter λ.

We do not know if the following comparison principle holds for the geometry problem (5): If $\Omega_1 \subset \Omega_2$ and Σ_1, Σ_2 are minimizers of (5) with $\Omega = \Omega_1$ and $\Omega = \Omega_2$ respectively, then do we necessarily have $\Sigma_1 \subset \Sigma_2$? If true, this would imply, in particular, uniqueness for solutions of (5). In any case, we can make the following statement:

Corollary 2 *If the observed image $f(x)$ is the characteristic function of a bounded, convex domain $\Omega \subset \mathbf{R}^N$, then for almost every $\lambda \geq 0$ the minimizer of $E_1(\cdot, \lambda)$ is unique and is the characteristic function of a set contained in Ω.*

Proof: Let $\lambda \in [0, \infty) \setminus S(f)$. And let $u_\lambda(x) \in M(\lambda)$. We recall from the remark that follows Theorem 1 that, using the same notation as in that remark, the set $\Sigma(\gamma)$ minimizes the geometry problem (5) for almost every $\gamma \in [0, 1]$. Let $1 \geq \gamma_1 > \gamma_2 \geq 0$, and assume that $\Sigma(\gamma_1) \neq \Sigma(\gamma_2)$ both minimize the geometry problem. By definition, we have $\Sigma(\gamma_1) \subset \Sigma(\gamma_2)$. Furthermore, convexity of Ω implies that

$$\text{Per}(\Sigma(\gamma_i) \cap \Omega) \leq \text{Per}(\Sigma(\gamma_i)) \text{ for } i = 1, 2.$$

Since $\mathbf{1}_{\Sigma(\gamma_1)}(x)$ and $\mathbf{1}_{\Sigma(\gamma_2)}(x)$ are minimizers, it follows that $\Sigma(\gamma_1) \subset \Sigma(\gamma_2) \subseteq \Omega$. Hence, $|\Sigma(\gamma_1) \triangle \Omega| \neq |\Sigma(\gamma_2) \triangle \Omega|$. But then $\lambda \in S(f)$, which is a contradiction. We thus reached the conclusion that if $\lambda \in [0, \infty) \setminus S(f)$, then any minimizer of $E_1(\cdot, \lambda)$ is necessarily binary (i.e. the characteristic function of a set). Now suppose that $u_1(x)$ and $u_2(x)$ are two binary minimizers of $E_1(\cdot, \lambda)$. By convexity of $E_1(\cdot, \lambda)$, we then have that $\frac{1}{2}(u_1(x) + u_2(x))$ is also a minimizer, and thus binary. But the average of two binary functions is binary only if the two functions are identical.

Thus, whenever $\lambda \in [0, \infty) \setminus S(f)$, the minimizer of $E_1(\cdot, \lambda)$ is unique, and is binary: it is of the form $\mathbf{1}_\Sigma(x)$ for some set Σ. The argument above shows that $\Sigma \subseteq \Omega$. And Claim 5 says that $S(f)$ is at most countable, and thus negligible. That proves the claim. $\square$

As an aside, we note the following result about problem (5) that follows immediately from the previous corollary (perhaps it can be obtained also in a less roundabout way):

Corollary 3 *Let Ω be a bounded, convex domain in $\mathbf{R}^N$. Then, for almost every $\lambda \geq 0$ the solution of problem (5) is unique.*

Proof: If Σ_1 and Σ_2 are solutions to (5), then $\mathbf{1}_{\Sigma_1}(x)$ and $\mathbf{1}_{\Sigma_2}(x)$ are minimizers of $E_1(\cdot, \lambda)$ with the observed image given by $f(x) = \mathbf{1}_\Omega(x)$. Conditions on Ω imply that Corollary 2 applies so that $\Sigma_1 = \Sigma_2$. That proves the claim. $\square$

We will next consider some simple images $f(x)$ for which the minimizer of $E_1(\cdot, \lambda)$ turns out to be precisely the image $f(x)$ itself for every large enough λ. In Section 1, we recalled a result from Meyer's lecture notes [10] which says that for the standard ROF model given by $E_2(\cdot, \lambda)$ the only such image is $f(x) := 0$. For E_1, however, there are many such images, as shown by Proposition 1 that we quoted in Section 2 from [3]. The following Lemma will be instrumental in establishing whether certain simple observed images $f(x)$ have this property.

Lemma 1 *Given an observed image $f(x) \in BV(\mathbf{R}^N)$, assume that there is a vector field $\phi(x)$ with the following properties:*

1. $\phi(x) \in C_c^1(\mathbf{R}^N; \mathbf{R}^N)$,
2. $|\phi(x)| \leq 1$ *for all* $x \in \mathbf{R}^N$,
3. $\displaystyle\int_{\mathbf{R}^N} f(x) \operatorname{div} \phi(x)\, dx = \int_{\mathbf{R}^N} |\nabla f|$.

Then there exists a threshold $\lambda_* \geq 0$ such that $M(\lambda) = \{f(x)\}$ for all $\lambda > \lambda_*$. In other words, the unique minimizer of $E_1(\cdot, \lambda)$ is given by the observed image $f(x)$.

Proof: Set $\lambda_* := \max_{x \in \mathbf{R}^N} |\operatorname{div} \phi(x)|$. Take any $\lambda > \lambda_*$. Then, given any $u(x) \in BV(\mathbf{R}^N)$ we have:

$$\begin{aligned} E_1(u, \lambda) &= \int |\nabla u| + \lambda \int |u - f|\, dx \\ &\geq \int u \operatorname{div} \phi\, dx + \lambda \int |u - f|\, dx \\ &= \int f \operatorname{div} \phi\, dx + \lambda \int |u - f|\, dx + \int (u - f) \operatorname{div} \phi\, dx \\ &\geq E_1(f, \lambda) + \Big(\lambda - \max_{x \in \mathbf{R}^N} |\operatorname{div} \phi(x)|\Big) \int |u - f|\, dx. \end{aligned}$$

Since $\lambda > \lambda_* := \max |\operatorname{div} \phi(x)|$, the last inequality shows that $E_1(u, \lambda) > E_1(f, \lambda)$ unless $u \equiv f$. Since u is a minimizer, it must in fact be the case that $u \equiv f$. $\square$

Lemma 1 can now be applied, for example, to binary images to obtain an important class of exact solutions. This requires making some smoothness assumption about the interface between then two values of the binary function:

Theorem 2 *Let $\Omega \subset \mathbf{R}^N$ be a bounded domain with C^2 boundary. Let the observed image $f(x)$ be given by $f(x) = \mathbf{1}_\Omega(x)$. Then there exists a threshold $\lambda_* \geq 0$ such that whenever $\lambda > \lambda_*$ the unique minimizer of $E_1(\cdot, \lambda)$ is the observed image $f(x) = \mathbf{1}_\Omega(x)$ itself.*

Proof: Since the boundary $\partial\Omega$ of the bounded domain Ω is assumed to be C^2, the outward unit normal vector field $n(x) : \partial\Omega \to \mathbf{S}^{N-1}$ of $\partial\Omega$ can be extended in a C^1 manner to a tubular neighborhood of $\partial\Omega$, so that one gets a vector field $\phi(x) \in C_c^1(\mathbf{R}^N; \mathbf{R}^N)$ such that $\phi(x)\big|_{x\in\partial\Omega} = n(x)$, and $|\phi(x)| \leq 1$ for all $x \in \mathbf{R}^N$. But then

$$\begin{aligned}\int_{\mathbf{R}^N} f \operatorname{div} \phi \, dx &= \int_\Omega \operatorname{div} \phi(x) \, dx = \int_{\partial\Omega} \phi(x) \cdot n(x) \, d\sigma \\ &= \text{Per}(\partial\Omega) = \int_{\mathbf{R}^N} |\nabla f| \, dx.\end{aligned}$$

Hence, the vector field $\phi(x)$ satisfies all the requirements of Lemma 1, from which the conclusion of the present claim follows. □

At this point it is worth recalling the behavior of the standard ROF model on binary images of the form $f(x) = 1_\Omega(x)$. As we noted above, simple considerations show that the minimizer of the standard ROF model almost never turns out to be $u(x) = f(x) = 1_\Omega(x)$. A related question is whether the solution $u(x)$ has at least the correct "set of edges". In case Ω is a ball, one can calculate the minimizer explicitly [18]; it turns out to be $u(x) = c\mathbf{1}_\Omega(x)$, where $c = 1 - \frac{\text{Per}(\Omega)}{2\lambda|\Omega|}$. In particular, $u(x)$ has the same set of edges as $f(x)$. The results of [4] generalize the results of [18], but also show that the class of binary images that have this weaker property (i.e. images for which the solution to the standard ROF model turns out to be a constant multiple of the observed image) is still rather limited; for example, there are smooth but non-convex shapes that lack this property.

Remark: Theorem 2 can be easily extended to images of a more general form. Indeed, if the level sets $\{x : f(x) = \gamma\}$ of the given image $f(x)$ are smooth and vary smoothly with respect to γ, the same conclusion holds. We also see, among other things, that such an image $f(x)$ cannot have strict local extrema, for at a strict local extrema the level sets shrink to a point. Moreover, there are also binary images that lack this property (i.e. which are not exactly recovered for any $\lambda \geq 0$, no matter how large). In fact, a repetition of some of the arguments of Meyer given in his lecture notes [10] on the standard ROF model show that the characteristic function of, say, a square cannot arise as the solution to the ROF model with L^1 fidelity, either, no matter what the observed image $f(x) \in L^1$ is, and no matter how large the parameter λ is chosen. □

The last few claims dealt with the behavior of the L^1 fidelity based model for large values of the parameter λ. Next, we consider what happens when $\lambda \geq 0$ is small enough. The following claim is a very simple application of the isoperimetric inequality:

Proposition 3 *Let $R > 0$. Then, there exists a threshold $\lambda_* = \lambda_*(R, N)$ such that if $f \in L^1(\mathbf{R}^N)$ with supp$(f) \subset B_R(0)$, then $M(\lambda) = \{0\}$ for any $\lambda < \lambda_*$. In other words, the unique minimizer of $E_1(\cdot, \lambda)$ is given by $u(x) \equiv 0$.*

Proof: Let $C = C(N)$ be the isoperimetric constant:

$$\int_{\mathbf{R}^N} |\nabla u| \geq C(N) \|u\|_{L^{\frac{N}{N-1}}(\mathbf{R}^N)} \text{ for all } u \in BV(\mathbf{R}^N).$$

Then we set

$$\lambda_*(R, N) := \frac{C(N)}{R\omega_N^{\frac{1}{N}}}$$

where ω_N is the volume of the unit ball in $\mathbf{R}^N$. Take a $\lambda > \lambda_*$ and let $u(x) \in M(\lambda)$. Then $E_1(u, \lambda) \leq E_1(0, \lambda)$. By the isoperimetric inequality, that means

$$C(N)\|u\|_{L^{\frac{N}{N-1}}(\mathbf{R}^N)} + \lambda\|u - f\|_{L^1(\mathbf{R}^N)} \leq \lambda\|f\|_{L^1(\mathbf{R}^N)} = \lambda\|f\|_{L^1(B_R(0))}.$$

We apply Holder's inequality to the first term on the left hand side after splitting it into integrations over $B_R(0)$ and $B_R^c(0)$. That gives

$$\frac{C(N)}{R\omega_N^{\frac{1}{N}}} \|u\|_{L^1(B_R(0))} + \lambda\|u - f\|_{L^1(B_R(0))} + C(N)\|u\|_{L^{\frac{N}{N-1}}(B_R^c(0))} \leq \lambda\|f\|_{L^1(B_R(0))}$$

which shows that if $\lambda < \frac{C(N)}{R\omega_N^{\frac{1}{N}}} = \lambda_*$, then

$$\|u\|_{L^1(B_R(0))} = \|u\|_{L^{\frac{N}{N-1}}(B_R^c(0))} = 0.$$

In other words, $u \equiv 0$. $\square$

Remark: This behavior of the L^1 model is to be expected, based on its contrast invariance, as we have already noted in the introduction. It differs from the behavior at small λ values of the standard ROF model which, according to [10], entails not just the support of a given compactly supported image $f(x)$ but its $\|\cdot\|_*$-norm. $\square$

6 Scale space and the set $S(f)$

The set $S(f)$ of discontinuities of the functions $\mu^{\pm}$ play a distinguished role in the scale space generated by varying the parameter λ in the L^1 model. As the value of λ is gradually decreased, minimizers of the image models become coarser as small scale objects in the image merge to form larger scale structures. Intuitively, for the L^1 model we can expect the values of $\lambda \in S(f)$ to correspond to scales of distinct objects that make up the image. These are the values of λ at which the scale space makes a rapid and drastic transition.

We would first like to prove that the set $S(f)$ is non-empty for the kind of images we have been considering in the previous sections, namely images of the form $f(x) = \mathbf{1}_{\Omega}(x)$ where Ω is a bounded domain. Our arguments are based on verifying this claim for the special case where the given image is the characteristic function of a ball, and then generalizing the result to $f(x) = \mathbf{1}_{\Omega}(x)$ by comparing Ω with a ball that is contained in Ω.

Lemma 2 *Let Ω be a bounded domain in $\mathbf{R}^2$, and assume that $B_R(p) \subset \Omega$. Consider the observed image given by $f(x) = \mathbf{1}_{\Omega}(x)$. Then for any $\lambda \geq 0$ and $r \in (0, R)$ we have*

$$E_1(\mathbf{1}_{B_r(p)}(x), \lambda) > \min\Big\{E_1(0, \lambda), E_1(\mathbf{1}_{B_R(p)}(x), \lambda)\Big\}.$$

Proof: Since $B_r(p) \subset B_R(p) \subset \Omega$ for each $r \in (0, R)$, we have

$$\|\mathbf{1}_{\Omega}(x) - \mathbf{1}_{B_r(p)}(x)\|_{L^1(\mathbf{R}^2)} = |\Omega| - \pi r^2.$$

That means

$$E_1(\mathbf{1}_{B_r(p)}(x), \lambda) = \lambda\Big(|\Omega| - \pi r^2\Big) + 2\pi r.$$

Considering $E_1(\mathbf{1}_{B_r(p)}(x), \lambda)$ as a function of r, we see that it achieves its minimum on $[0, R]$ strictly at the end points of the interval. $\square$

In order to show that $\mu^{\pm}(\lambda)$ is a discontinuous function, we will show that its range omits a full interval of values, but does include certain values on either side of that interval. The next claim exhibits such an omitted interval:

Proposition 4 *Let Ω be a bounded domain in $\mathbf{R}^2$, and let $B_R(0) \subset \Omega$. Consider the observed image given by $f(x) = \mathbf{1}_{\Omega}(x)$. There is no $\lambda \in \mathbf{R}^+$ such that*

$$|\Omega| - \pi R^2 < \mu^+(\lambda) < |\Omega|.$$

Proof: Suppose there is a $\lambda \geq 0$ such that $|\Omega| - \pi R^2 < \mu^+(\lambda) < |\Omega|$. There exists $u(x)$ such that $u(x) \in M(\lambda)$ and $\|u - f\|_{L^1(\mathbf{R}^N)} = \mu^+(\lambda)$. As before, let

$\Sigma(\gamma) := \{x : u(x) > \gamma\}$. By Proposition 2, we have $\mathbf{1}_{\Sigma(\gamma)}(x) \in M(\lambda)$ for a.e. $\gamma \in (0,1)$. Therefore, for a.e. γ we have

$$\|\mathbf{1}_{\Sigma(\gamma)}(x) - f\|_{L^1(\mathbf{R}^2)} < |\Omega|$$

(otherwise $\mu^+(\lambda) \geq |\Omega|$). It also *cannot be the case* that $|\Sigma(\gamma) \triangle \Omega| \leq |\Omega| - \pi R^2$ for a.e. $\gamma \in (0,1)$ since we know that

$$\int_0^1 |\Sigma(\gamma) \triangle \Omega| \, d\gamma = \|u - f\|_{L^1(\mathbf{R}^2)} = \mu^+(\lambda) > |\Omega| - \pi R^2.$$

Thus, there exists $\gamma_* \in (0,1)$ such that

$$\mathbf{1}_{\Sigma(\gamma_*)}(x) \in M(\lambda) \text{ and } |\Omega| - \pi R^2 < \|\mathbf{1}_{\Sigma(\gamma_*)}(x) - f(x)\|_{L^1(\mathbf{R}^2)} < |\Omega|.$$

Case 1: $|\Sigma(\gamma_*)| \geq \pi R^2$. But then $\text{Per}(B_R(0)) \leq \text{Per}(\Sigma(\gamma_*))$, and

$$|\Omega \triangle B_R(0)| = |\Omega| - \pi R^2 < \|\mathbf{1}_{\Sigma(\gamma_*)}(x) - f(x)\|_{L^1}.$$

Hence, $E_1(\mathbf{1}_{B_R(0)}(x), \lambda) < E_1(\mathbf{1}_{\Sigma(\gamma_*)}(x), \lambda)$. This is a contradiction, since $\mathbf{1}_{\Sigma(\gamma)}(x)$ was supposed to be a minimizer.

Case 2: $|\Sigma(\gamma_*)| < \pi R^2$. In this case, take $r = \frac{1}{\sqrt{\pi}} |\Sigma(\gamma_*)|^{\frac{1}{2}}$. Since $r \in (0, R)$, we have that $B_r(0) \subset \Omega$. This implies

$$\|\mathbf{1}_{B_r(0)}(x) - f(x)\|_{L^1(\mathbf{R}^2)} \leq \|\mathbf{1}_{\Sigma(\gamma_*)}(x) - f(x)\|_{L^1(\mathbf{R}^2)}.$$

Moreover, as before, $\text{Per}(B_R(0)) \leq \text{Per}(\Sigma(\gamma_*))$. Therefore,

$$E_1(\mathbf{1}_{B_r(0)}(x), \lambda) \leq E_1(\mathbf{1}_{\Sigma(\gamma_*)}(x), \lambda) = E_1(u(x), \lambda).$$

On the other hand, by Lemma 2 we have

$$E_1(\mathbf{1}_{B_r(0)}(x), \lambda) > \min\Big\{E_1(0, \lambda), E_1(\mathbf{1}_{B_R(0)}(x), \lambda)\Big\}.$$

This is a contradiction, since $u(x) \in M(\lambda)$. $\square$

Theorem 3 *Let Ω be a non-empty, bounded domain in $\mathbf{R}^2$. Consider the observed image given by $f(x) = \mathbf{1}_\Omega(x)$. Then the functions $\mu^\pm(\lambda)$ are discontinuous.*

Proof: By Proposition 3, we have that $\mu^+(\lambda) = \|f\|_{L^1} = |\Omega|$ for all small enough λ. On the other hand, by Claim 6 we have that $\mu^\pm(\lambda) \to 0$ as $\lambda \to \infty$. Yet by Proposition 4, there is a range of values near $|\Omega|$ that the function μ^+ cannot take.

It therefore has to be discontinuous. Discontinuity of μ^- follows form that of μ^+ via Claim 4. $\square$

Remark: This should be contrasted with the situation for the standard total variation model (with L^2 fidelity), which is explained in Claim 3. $\square$

We thus see that the scale spaces generated by the two models, the standard ROF model and the one with L^1 fidelity, are very different. With the standard ROF model, pronounced objects of distinct scale with sharp edges in the image gradually lose their contrast and merge with their neighbors as the parameter λ is lowered. With the L^1 model, such objects maintain their contrast with respect to their neighbors – however their boundaries might be gradually smoothed out. This goes on until a critical value of λ is reached – one that belongs to the set $S(f)$, at which point the object suddenly merges with a neighboring one.

At this point, it is also worth comparing the scale space generated by the L^1 model with that generated by anisotropic diffusion via motion by mean curvature of level sets. The two are drastically different. This can be seen most easily in the case when $f(x)$ is the characteristic function of a disk. The scale space generated by motion by curvature consists of a family of concentric disks shrinking gradually to a point. Hence the same feature, i.e. the original disk, appears at many intermediate scales, albeit in different sizes. On the other hand, the scale space generated by the total variation model with L^1 fidelity term consists of either the original disk or the constant background at any given scale.

Finally, we return to the topic of continuous dependence on the observed signal for the L^1 model. Despite our remarks in Section 3, we show in the next claim that the fidelity of minimizer versus λ graph depends on the observed image continuously.

Proposition 5 *Let $\{f_j(x)\}_{j=1}^{\infty}$ be a sequence in $L^1(\mathbf{R}^N)$ that converges to $f(x)$ in L^1. Then, for all $\lambda \in [0,\infty) \setminus S(\lambda, f)$, the sequence $\mu^{\pm}(\lambda, f_j)$ converges to $\mu^{\pm}(\lambda, f)$ as $j \to \infty$.*

Proof: Fix $\lambda \in [0,\infty) \setminus S(f)$. For each j, we can find a $u_j \in M(\lambda, f_j)$ such that $\mu^+(\lambda, f_j) = \|u_j - f_j\|_{L^1}$. This sequence is bounded in total variation and L^1; we may therefore pass to a subsequence, also denoted u_j, and find a $u \in L^1(\mathbf{R}^N)$ such that $u_j \to u$ pointwise a.e. Proceeding exactly as in the proof of Claim 3, we can now show, based on lower semi-continuity property, that $u \in M(\lambda, f)$ and

$$\mu^{\pm}(\lambda, f) = \|u - f\|_{L^1} \leq \liminf_{j\to\infty} \|u_j - f_j\|_{L^1} = \liminf_{j\to\infty} \mu^+(\lambda, f_j).$$

Exactly as in that proof, we can also show, based on the fact that $E_1(\cdot, \lambda, f_j)$ are a

continuous perturbation in L^1 from $E_1(\cdot,\lambda,f)$, that actually

$$\mu^{\pm}(\lambda,f)=\|u-f\|_{L^1}\geq\limsup_{j\to\infty}\|u_j-f_j\|_{L^1}=\limsup_{j\to\infty}\mu^{+}(\lambda,f_j)$$

That proves the claim. □

7 Computation

In this section, we show numerical examples that bring out unique features of the total variation based denoising model with L^1 fidelity term. We also give some details on the numerical schemes used to obtain these results.

Our computations are based on gradient descent schemes for decreasing the energies involved. The non-differentiability of the terms involved in the energies call for some sort of regularization. The regularized versions of energies $E_1(\cdot,\lambda)$ and $E_2(\cdot,\lambda)$ used in our numerical experiments are the following:

$$E_1^{\varepsilon,\delta}(u,\lambda):=\int_{\mathbf{R}^N}\sqrt{|\nabla u|^2+\varepsilon}+\lambda\int_{\mathbf{R}^N}\sqrt{(f-u)^2+\delta}\,dx,$$
$$E_2^{\varepsilon}(u,\lambda):=\int_{\mathbf{R}^N}\sqrt{|\nabla u|^2+\varepsilon}+\lambda\int_{\mathbf{R}^N}(f-u)^2\,dx.$$

This type of approximation to total variation based models is very standard. The discrete versions of these energies lead to the following equally standard explicit gradient descent schemes in two space dimensions:

$$\frac{u_{i,j}^{n+1}-u_{i,j}^{n}}{\delta t}=D_x^-\left(\frac{D_x^+u_{i,j}^n}{\sqrt{(D_x^+u_{i,j}^n)^2+(D_y^+u_{i,j}^n)^2+\varepsilon}}\right)$$
$$+D_y^-\left(\frac{D_y^+u_{i,j}^n}{\sqrt{(D_x^+u_{i,j}^n)^2+(D_y^+u_{i,j}^n)^2+\varepsilon}}\right)+\lambda\frac{(f-u_{i,j}^n)}{\left((f-u_{i,j}^n)^2+\delta\right)^{\alpha}}.$$

where $\alpha=\frac{1}{2}$ for $E_1^{\varepsilon,\delta}$ and $\alpha=0$ for E_2^{ε}. Here, D^+ and D^- denote forward and backward difference quotients, respectively, in the direction of their subscript. We note that efficient numerical minimization of energies considered in this work is a topic onto itself; no doubt there are better ways to do it then the gradient descent approach taken and the specific choice of scheme made above.

An important point we need to clarify is the following. Although as we already noted several times the energy $E_1(\cdot,\lambda)$ is not strictly convex and its minimizers in general lack uniqueness, for any given $\delta>0$ the approximate energy $E_1^{\varepsilon,\delta}(\cdot,\lambda)$ *is*

strictly convex so that its minimizers enjoy uniqueness. It is these minimizers that we have computed. Moreover, it is a very routine matter to verify that a sequence of minimizers of $E_1^{\varepsilon,\delta}(\cdot,\lambda)$ converges to the set of minimizers $M(\lambda)$ of $E_1(\cdot,\lambda)$ as $\varepsilon,\delta \to 0^+$. The analogous convergence statement is of course true also for a sequence of minimizers of $E_2^{\varepsilon}(\cdot,\lambda)$.

Figures 2 and 3 compare the scale spaces generated by the standard total variation model and the one with L^1 fidelity on a synthetic image. This experiment makes the more geometric nature of the L^1 model abundantly clear. The observed image consists of squares of various sizes and gray levels. In the scale space generated by the standard total variation model, the squares gradually lose their contrast (while at the same time their geometries get regularized) and gradually disappear. Moreover, some large squares with low contrast against the background – namely the square near the upper left corner – disappear before some smaller squares that have higher contrast against the background – namely the two intermediate sized squares along the diagonal. On the other hand, in the scale space generated by the model with L^1 fidelity, the squares get processed only in terms of their geometry: they preserve their contrast until all of a sudden they disappear. The contrast of the squares plays no role in determining the order in which they are removed; that order is determined completely in terms of the geometry of the features.

Figure 4 shows the graph of the fidelity of the minimizer vs. λ for the standard total variation model, and the model with L^1 fidelity. An important ambiguity we need to resolve is how nonuniqueness of minimizers of $E_1(\cdot,\lambda)$ affects the fidelity vs. λ plot for $E_1(\cdot,\lambda)$. To answer this question, recall that the fidelity of various minimizers of $E_1(\cdot,\lambda)$ differ from each other only at countably many values of λ. In particular, all ways of obtaining the second graph in Figure 3 yield plots that are identical up to a set of measure 0. Hence, there is no ambiguity in the results shown.

Discontinuities in the minimizer's fidelity versus λ graph for the L^1 model correspond to distinguished values of the parameter λ. As can be seen from the results, these are the values of λ at which a drastic change in the scale space takes place. Namely, at such values of λ one of the "features" (squares in this example) gets eliminated. There is no such distinguished value of λ in the plot for the standard ROF model at which the graph becomes discontinuous (as shown both by our theoretical results and by the numerical example shown). However, the graph in that case might have kinks, which are of course harder to detect than discontinuities. Thus, unlike the standard total variation model, the model with L^1 fidelity thus suggests a method for *data driven parameter selection*.

The special values of parameter λ obtained from the fidelity of minimizer graph via L^1 model can be used in many ways. For example, denoising models are sometimes used for generating multi-scale decomposition of images, for example as in

[19]. In such applications, it is necessary to select a *schedule* for the parameter λ *a priori*. In [19], this schedule is chosen in the form $\lambda = 2^j\lambda_0$ with $j = 1, 2, 3, \ldots$, and the initial value λ_0 is arbitrarily chosen by the user. The L^1 scale space suggests a more natural, data driven way to select these parameters using the discontinuities in the fidelity of minimizers graph. Moreover, even if one opts to use a λ-schedule of the form used in [19], the theoretical results and preliminary numerical examples of this paper suggest that one might obtain a much cleaner decomposition using the ROF model with L^1-fidelity in place of the standard ROF model. All these ideas pertaining to multiscale decomposition of images using the L^1 fidelity based model will be explored elsewhere.

Finally, Figures 5 and 6 illustrate the differences between the standard ROF model and the one with L^1 fidelity on a real medical image. In this example also, one can see that the small scale features in the observed image, such as the ones indicated by the arrow on the lower left hand side image of Figure 6, maintain their contrast much better in the L^1 fidelity model than in the standard ROF model, even as the parameter λ is gradually decreased to very low values.

8 Conclusion

We considered the total variation based image denoising model of Rudin, Osher, and Fatemi with the L^1 norm as the fidelity term. Our results highlight that this modification leads to many interesting qualitative differences in the behavior of the modified model from the standard one. These differences have important consequences for image denoising. They also suggest interesting new research directions into applications to data driven parameter selection, and multiscale image decomposition.

Acknowledgment: The authors would like to thank Antonin Chambolle, Mila Nikolova, Stanley Osher, and Luminita Vese for helpful discussions.

References

[1] Alliney, S. *Digital filters as absolute norm regularizers.* IEEE Transactions on Signal Processing. **40**:6 (1992), pp. 1548 – 1562.

[2] Alliney, S. *Recursive median filters of increasing order: a variational approach.* IEEE Transactions on Signal Processing. **44**:6 (1996), pp. 1346 – 1354.

[3] Alliney, S. *A property of the minimum vectors of a regularizing functional defined by means of the absolute norm.* IEEE Transactions on Signal Processing. **45**:4 (1997), pp. 913 – 917.

[4] Bellettini, G.; Caselles, V.; Novaga, M. *Total variation flow in* $\mathbf{R}^N$. Journal of Differential Equations. **184** (2002), pp. 475 – 525.

[5] Catte, F.; Dibos, F.; Koepfler, G. *A morphological scheme for mean curvature motion and applications to anisotropic diffusion and motion of level sets.* SIAM J. Numer. Anal. **32** (1995), no. 6, pp. 1895 – 1909.

[6] Dibos, F.; Koepfler G. *Global total variation minimization.* SIAM J. Numer. Anal. **37** (2000), no. 2, pp. 646 – 664.

[7] Dobson, D.; Santosa, F. *Recovery of blocky images from noisy and blurred data*. SIAM Journal on Applied Mathematics. **56** (1996), pp. 1181 – 1198.

[8] Evans, L. C.; Gariepy, R. F. *Measure theory and fine properties of functions.* Studies in Advanced Mathematics. CRC Press, Roca Baton, FL, 1992.

[9] Giusti, E. *Minimal surfaces and functions of bounded variation.* Monographs in Mathematics, 80. Birkhauser Verlag, Basel, 1984.

[10] Meyer, Y. *Oscillating patters in image processing and nonlinear evolution equations*. AMS University Lecture Series Vol. 22 (2002).

[11] Nikolova, M. *Minimizers of cost-functions involving nonsmooth data-fidelity terms*. SIAM Journal on Numerical Analysis. **40**:3 (2002), pp. 965 – 994.

[12] Nikolova, M. *A variational approach to remove outliers and impulse noise*. Journal of Mathematical Imaging and Vision.

[13] Nikolova, M. *Weakly constrained minimization. Application to the estimation of images and signals involving constant regions*. Journal of Mathematical Imaging and Vision.

[14] Osher, S.; Fedkiw, R. *Level set methods and dynamic implicit surfaces.* Applied Mathematical Sciences, 153. Springer-Verlag, New York, 2003.

[15] Osher, S.; Sethian, J. *Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations.* Journal of Computational Physics. **79**:1 (1988), pp. 12 – 49.

[16] Osher, S.; Sole, A.; Vese, L. *Image decomposition and restoration using total variation minimization and the H^{-1} norm*. SIAM Journal on Multiscale Modeling and Simulation. **1**:3 (2003), pp. 349 – 370.

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

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

[19] Tadmor, E.; Nezzar, S.; Vese, L. *A multiscale image representation using hierarchical (BV, L^2) decompositions.* UCLA CAM Report 03-32, July 2003.

[20] Vese, L.; Osher, S. *Modeling textures with total variation minimization and oscillating patterns in image processing*. Journal of Scientific Computing. **19**:1-3 (2003), pp. 553 – 572.

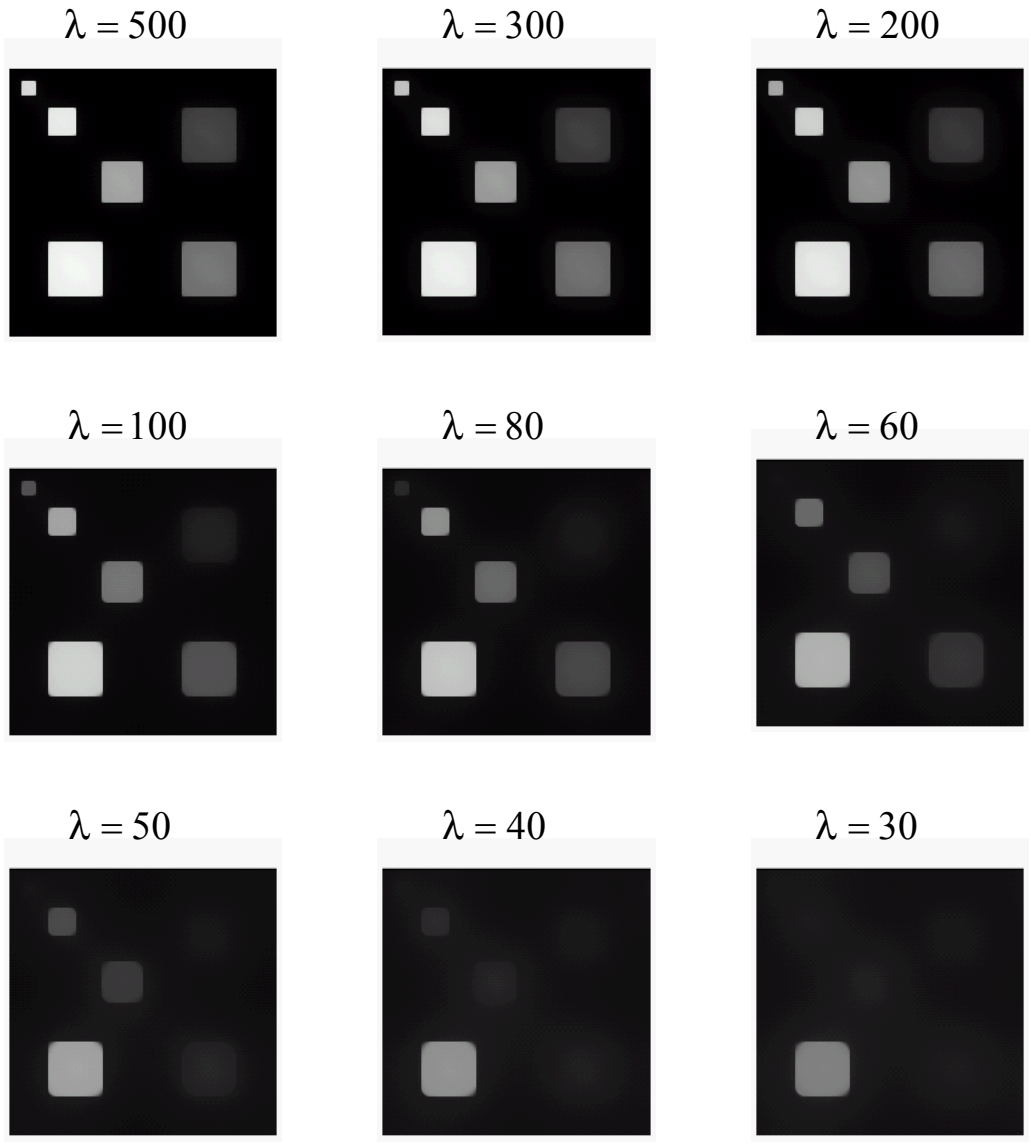

Figure 2: Example of scale space generated by the standard total variation model. Compare with the same example for the model with L^1 fidelity, shown in Figure 3.

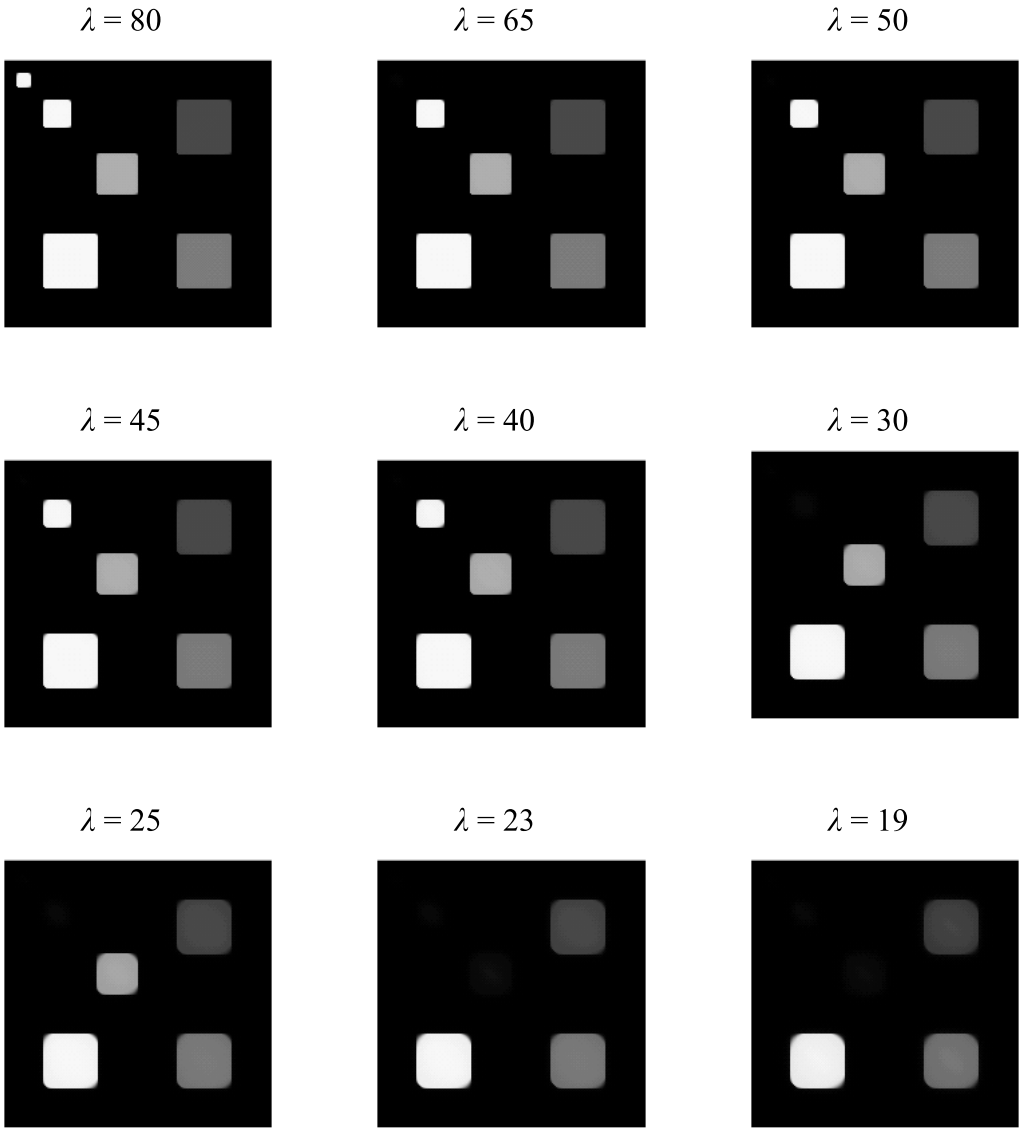

Figure 3: Example of scale space generated by the total variation model with L^1 fidelity. Compare with Figure 2.

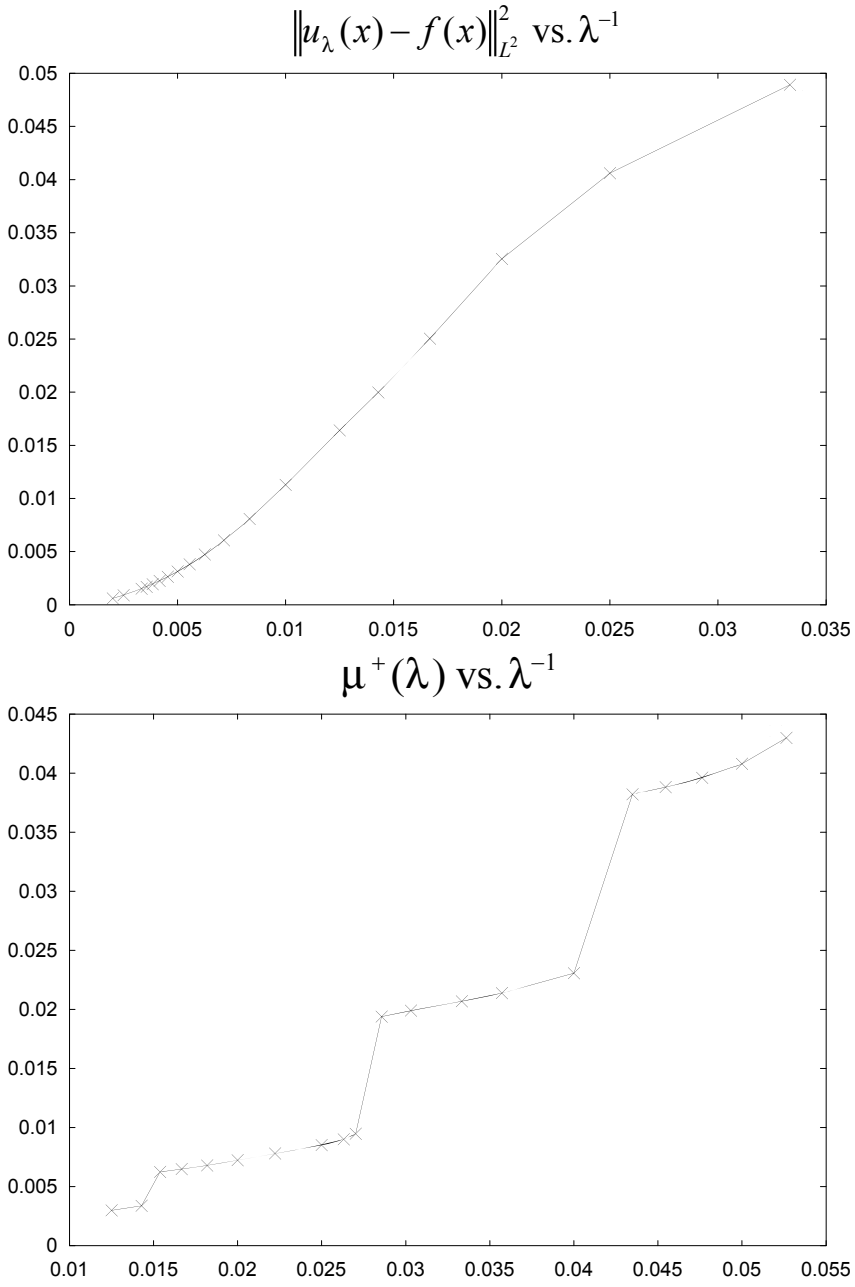

Figure 4: Plot of fidelity of minimizer (i.e. $\|u_\lambda(x) - f(x)\|_{L^2}^2$) vs. λ^{-1} for the standard ROF model (top graph) and the plot of fidelity of minimizer (i.e. $\|u_\lambda(x) - f(x)\|_{L^1}$) vs. λ^{-1} for the ROF model with L^1 fidelity (bottom graph).

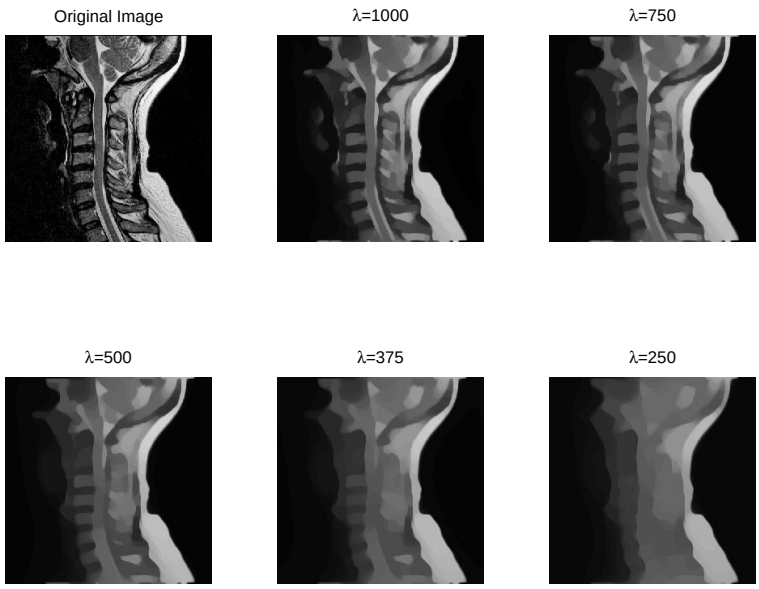

Figure 5: Scale space generated by the standard ROF model.

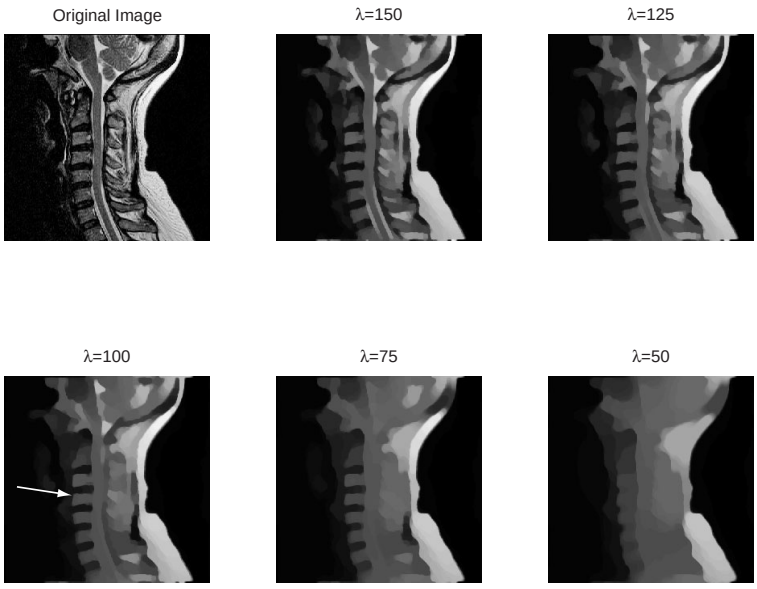

Figure 6: Scale space generated by the ROF model with L^1 fidelity term.