Relation of Regularization Parameter and Scale in Total Variation Based Image Denoising *

David M. Strong † Tony F. Chan ‡

June 5, 1996

Abstract

In image processing, regularization is generally done to remove or smooth out noise, as well as to enhance features of the image. Often a parameter is chosen which determines the relationship between the amount of regularization done and the goodness of fit to the measured noisy data. In multi-scale and adaptive regularization, a nonconstant parameter is chosen which varies in the domain of the image, and may depend on the image itself. Unfortunately this parameter, whether constant or varying, is generally chosen without a precise relationship between the regularization parameter and the effect of the regularization on features of specific scales. This choice is generally based only on noise, or on some other information which does not depend on the features in the image. In this paper we present a precise relationship between the regularization parameter in total variation (TV) based regularization and the *spatial scale* of the image. This theory is for one-, two- and three-dimensional images. This regularization-scale theory can be used to choose the regularization parameter based on the scale of the image. This is especially useful when doing adaptive regularization. We present a scheme for determining an adaptive regularization parameter based on both local scale and noise in the image. Since the regularization parameter may be chosen in a more precise and useful way, the user will have more control in extracting the information from the noisy image that is most useful. The regularization-scale relationship defined in this paper, and the scheme we suggest for applying it, are used to remove noise from several test images, including brain scans.

1 Introduction and Motivation

A typical problem in image processing is to determine the unknown image u given the noisy measured data z. The corresponding equation is

$$Ku + \epsilon = z,$$

*This work is supported by the NSF under grant ASC 92-01266, and by the ONR under contract ONR-N00014-92-J-1890.

†Department of Mathematics, University of California, Los Angeles, 90024; dstrong@math.ucla.edu

‡Department of Mathematics, University of California, Los Angeles, 90024; chan@math.ucla.edu

where K is a known "blurring" operater, and ϵ represents the error, essentially the noise, in the measured data. In this paper, we will deal with the case of non-blurred images, where $K = I$. (Our results can be extended to the case of blurred images.) Typically we wish the remove the noise (or otherwise enhance the image) by choosing u to be near the measured image z, while minimizing the functional $R(u)$, which is chosen to control the irregularity or oscillatory behavior of u. One formulation of this problem is the unconstrained problem

$$\min_u \ \tfrac{1}{2}\|u - z\|_\Omega^2 + \alpha\, R(u), \tag{1}$$

where α is a positive parameter that determines the trade-off between goodness of fit to the measured data, and the amount of regularization done to the measured image z. The other common formulation of this problem is the constrained problem

$$\min_u R(u) \quad \text{subject to} \quad \|u - z\|_\Omega^2 = \sigma^2, \tag{2}$$

where the error (noise) level σ is assumed to be known. For both problems, $\|u - z\|_\Omega^2 = \int_\Omega (u-z)^2\, dx$. We note that solving (1) is equivalent to solving (2) when $\alpha = \frac{1}{2\lambda}$, where λ is the Lagrange multiplier found in solving (2).

1.1 Total Variation Based Regularization

Often the functional $R(u)$ is taken to have the form $R(u) = \|Qu\|^2$, where Q is a *linear* differential operator. For example, $Q = I$, $Q = \Delta$ (the Laplacian operator) and $Q = \nabla$ (the gradient operator) are often used (cf. [6]). The following (non-linear) total variation (TV) norm[1] regularization functional has recently been proposed for use in image regularization by Osher, Rudin and Fatemi[12]:

$$R(u) = TV(u) \equiv \int_\Omega |\nabla u(\vec{x})|\, d\vec{x} \ \equiv \int_\Omega \sqrt{u_x^2 + u_y^2}\, dx\, dy \tag{3}$$

The functional $TV(u)$ simply measures the total variation of u.[2] The main advantage of TV regularization is that it does not penalize discontinuities (i.e. edges) in u; in fact, it has been reported in [12, 14] that this approach is especially good in preserving *sharp* edges. On the other hand, TV has the advantage that it does not penalize *smooth* images. For a more intuitive explanation of this property of TV, consider the following Lemma:

Lemma 1 *Given the set of functions defined on the interval [0,1], the argument of*

$$\min_u TV(u) \quad \textit{subject to} \quad u(0) = a \textit{ and } u(1) = b \tag{4}$$

is any monotone, not necessarily continuous, function u(x) satisfying $u(0) = a$ *and* $u(1) = b$.

[1] $TV(u)$ is actually a semi-norm, since $TV(u) = 0$ if u is constant.

[2] This definition of $TV(u)$ is only valid for differentiable functions u. We must interpret $\nabla u(\vec{x})$ as the distribution derivative in order for (3) to be valid for non-differentiable functions.

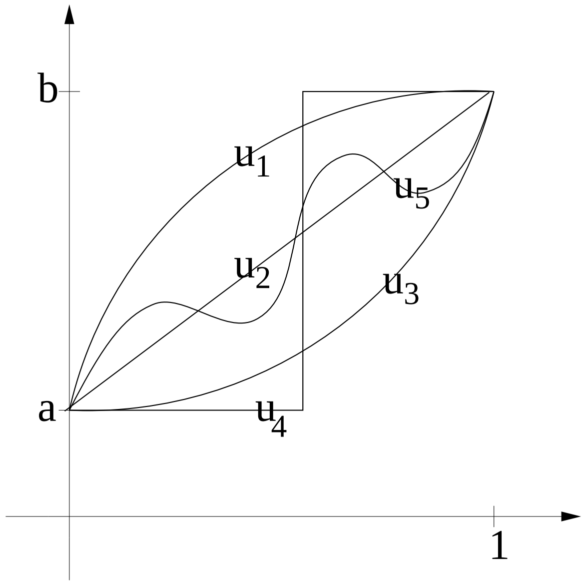

Figure 1: *Functions u_1, u_2, u_3 and u_4 each have total variation of $b-a$. The total variation of u_5 is greater than $b-a$ due to oscillations.*

In Figure 1 we see that u_1, u_2, u_3 and u_4 would all be acceptable functions to minimize (4), since $TV(u_1) = TV(u_2) = TV(u_3) = TV(u_4) = b - a = \min_u TV(u)$. Since u_5 is not monotone (in particular, it oscillates), $TV(u_5) > b - a$, and it is consequently not a solution to (4).

Thus the TV-regularized image restoration approach looks for an approximation to the observed image which has the least *total variation*. Note that there is no particular bias towards a discontinuous or smooth solution — the measured data z and the regularization parameter α (when solving (1)) or the noise level σ^2 (when solving (2)) will determine the sharpness or smoothness of the restored image.

In [12], a time marching scheme is proposed to solve (1) by marching the following PDE to steady state:

$$u_t = \alpha \nabla \cdot (\frac{\nabla u}{|\nabla u|}) - (u - z), \tag{5}$$

with initial condition $u_{t=0} = z$. The first term on the right-hand-side of (5) can be viewed as a nonlinear diffusion operator which diffuses less where the gradient is large (e.g. near edges). As such, the TV model is a special form of *anisotropic diffusion* models first proposed by Perona and Malik [10] and subsequently further developed by various authors (see references in [9, 11]).

More generally, one can consider a general function $f(|\nabla u|)$ in the integrand of $R(u)$ in (3), with f chosen not to penalize discontinuous functions u. We emphasize that one must be careful not to choose a function f which penalizes *smooth* functions u either, as it can possibly lead to "artificial edges" in images. The choice of $f(|\nabla u|) = |\nabla u|$, as in the TV norm, achieves these two goals.

There are several filtering or noise removal techniques which may be used to denoise, sharpen or otherwise enhance an image. Many of these are effective when the image is smooth, but tend to perform poorly when u has steep gradients or, in particular, discontinuities. There are also regularization techniques which may perform well in sharpening edges, but that perform poorly when dealing with gradual or smooth edges by introducing artificial sharpness in the image. Figure 2 gives a comparison of TV regularization to some standard noise removal techniques, which include using the FFT, H1 regularization, and wavelet decomposition (Haar, Daubechies4). Total variation based techniques are especially well suited for removing noise while preserving sharp edges, but they have the advantage that they do not penalize or try to over-sharpen smooth edges. Total variation based regularization techniques have been shown to be very effective in numerical studies (see [12], [14]) as well as in theoretical analysis (see [4]).

In [5] and [8], it is shown that a unique solution exists under certain conditions. This paper will not discuss numerical methods for solving (1) or (2). For some suggested methods see [12] and [14].

1.2 Spatially Adaptive Regularization in Image Restoration

In image restoration, there is generally a trade-off between noise removal and detail preservation. Images are typically comprised of multiple objects of different spatial scales. Generally, less smoothing is desired for features of smaller scale, while more smoothing is appropriate for features of larger scale. If the same amount of smoothing is done over the whole domain of the image, smaller-scaled features may be over-smoothed, resulting in loss of detail, while larger-scaled features may be under-smoothed, resulting in a noisier picture. See Figure 3 for an example of varying amounts of regularization.

Total variation methods are excellent for recovering "blocky" images but tend to somewhat smooth out "rough" boundaries of image segments [4]. Also, in removing noise from an image, the contrast is often reduced and can be lost completely. Both edges and contrast are essential in the usefulness of an image. A natural approach to partially alleviate this problem is to use spatial adaptivity. The basic idea is to apply more restoration (and less regularization) in regions of high intensity change and less restoration (and more regularization) in regions of low intensity change. Spatial adaptivity has been studied extensively in image restoration literature; see for example [2, 3]. In the context of TV restoration, one basic way of achieving spatial adaptivity would be by modifying equation (1), using $R(u) = TV(u)$, to

$$\min_u \ \tfrac{1}{2}\|u - z\|_\Omega^2 + \alpha(\vec{x})\, TV(u). \tag{6}$$

We could also solve a modified form of the constrained problem (2),

$$\min_u \int_\Omega \alpha(\vec{x})|\nabla u(\vec{x})|\, d\vec{x} \quad \text{subject to} \quad \|u - z\|_\Omega^2 = \sigma^2, \tag{7}$$

In both problems, $\alpha(\vec{x})$ can be chosen to vary based on the local features in the image as well as on the desired information to extract from the image. In order to realize the model in (6), we need to study the proper adaptive selection of α (as well as the development of efficient and accurate algorithms for solving the minimization problem).

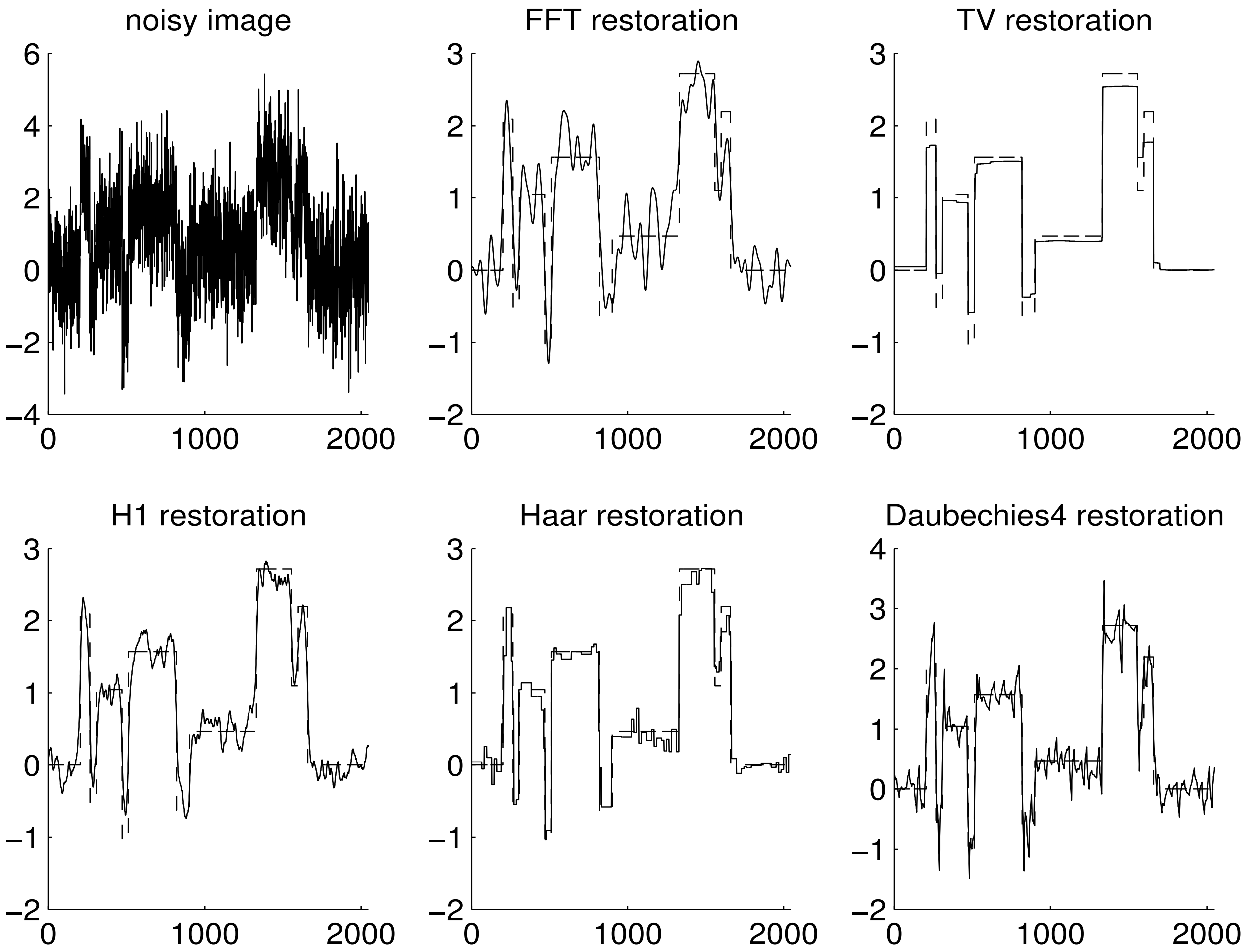

Figure 2: A comparison of several standard techniques in denoising a one-dimensional image. The dashed line represents the true image in each figure, and the solid line represents the regularized or denoised image.

1.3 Outline of Paper

The remainder of this paper is as follows. In Section 2, we will derive and discuss the relationship between the scale of an object, and the amount of regularization done to it as determined by the value of the regularization parameter α. In Section 3, we will outline a scheme for choosing $\alpha(\vec{x})$ in an adaptive way, depending on both the noise level and the local features in the image. This scheme is used to denoise several simple test images, as well as to a brain cross section image.

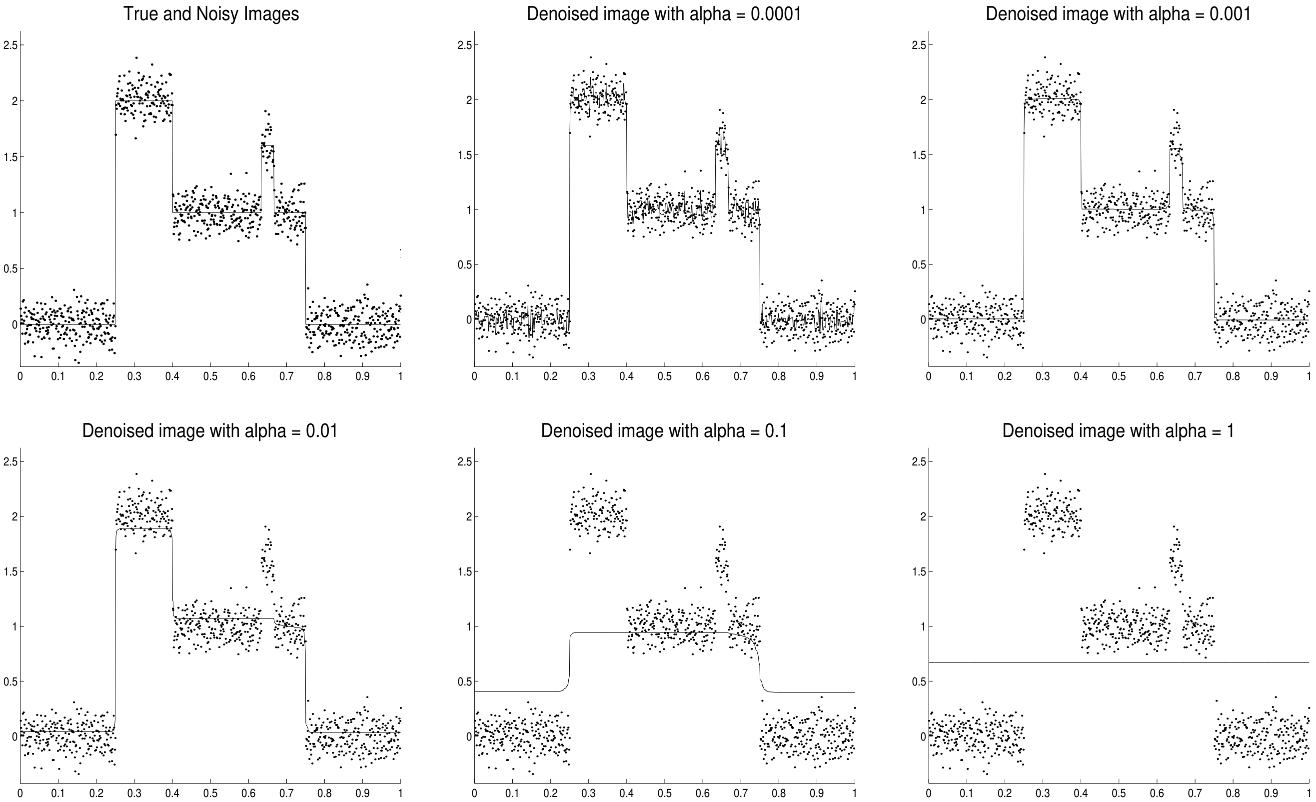

Figure 3: *A comparison of the results of using different values of the regularization parameter α in doing TV regularization on a one-dimensional image with multiple local scales. The dotted line is the noisy image, and the solid line is the regularized or denoised image. As α becomes larger, the smaller-scaled objects are lost.*

2 Exact Solution of the TV Image Restoration Problem

We first consider the unconstrained problem (1). Solving (1) gives us more obvious control over the balance between noise removal and goodness of fit to the measured image z. As we will see later, we can also use our analysis of the solution of problem (1) in adaptivity solving the constrained problem (2) to locally control the balance between fit and noise removal.

One benefit in using TV based regularization is that, as we will show in this paper, we *can* find a precise relationship between the regularization parameter and the amount of smoothing that occurs to objects of any scale. Often the regularization parameter is based on the noise only, or on some other scheme which uses no information about the scale or detail in the image in determining the parameter (see [7] for a general discussion of choosing regularization parameters). It can be very effective to base our choice of regularization parameter, especially when it is locally adaptive, on the local scale of the image. We can define a precise relationship between the amount of smoothing done to an object of a particular scale and the regularization parameter. By using this relationship, the user will now have more control in extracting the information from the noisy image that is most useful.

In this section we define and a precise relationship between the regularization parameter and the scale (or detail), as well as the noise in the image. It is natural then to briefly discuss the problems that may occur if α is not chosen properly. The goal of using TV based noise removal is essentially to smooth out the highly oscillatory noise, while preserving the edges and contrast of the true image. The larger α is, the more smoothing that occurs due to the TV minimization constraint. Generally, the more noise that is present in the measured data, the larger α is chosen to be, in order to smooth out the greater amount of noise. However, as α becomes larger, more detail is compromised or lost, due to the smoothing. There must be a balance between the amount of smoothing to occur and the amount of detail to be preserved.

Consider the problem of removing noise from the one-dimensional image shown in Figure 3. We must find u, the minimizer of (1), which will depend on z and our choice of α. For α too small, not enough smoothing is done to remove the noise, while for α too large, we lose too much of the detail. As α becomes larger, features with smaller scale are lost due to the condition of trying to lower the total variation of the image. (In this sense, TV regularization is a type of multi-scale regularization, with the scale determined by α.) The key is to find an appropriately varying value of α, either based on the amount of detail we wish to preserve in the image, or based on the amount of noise (whether known apriori or estimated given the noisy image), or based on both factors.

It is necessary to establish a relationship between the regularization parameter α and the amount of smoothing which occurs. In [4], Dobson and Santosa suggest that the effectiveness of TV based noise removel will depend on the "mass" (essentially the area, and the height or intensity) of the image compared to the total variation of the image. It could be very useful to define more precise relationship between the amount of smoothing which occurs to an object of given "mass" or area, and of a given scale (which corresponds to its total variation). With a more precise relationship defined, it becomes possible to make a more useful choice of the regularization parameter α. The results of this paper are especially useful when incorporated into an adaptive or multi-scale regularization scheme. When solving the modified unconstrained problem (6), the relationship we describe in this paper between local scale and regularization can be used in a very straightforward way to define a spatially varying value of α. When solving the modified constrained problem (7), we can use the regularization-scale relationship to determine more precisely the relative effect of the regularization on different local features, and we can consequently determine in a locally adaptive way the desired balance between fitting error and total variation of u.

In this section we present and discuss a precise relationship between the regularization parameter and the effect it has on an object of a given scale, particularly the the effect of regularization on the reduction in contrast. We do this first for one-dimensional images, then simultaneously for two- and three-dimensional images, including the special case of radially symmetric images in two and three dimensions. We look at how closely the ideal case (i.e. with certain assumptions) resembles the actual discrete case (without the assumptions). We also suggest a heuristic for estimating the scale of an object.

2.1 One-dimensional Piecewise Constant Images

To examine the relationship between the regularization parameter α and the amount of smoothing that occurs to an image, we will consider a simple one-dimensional example where no noise is present. This is useful in determining how the contrast if affected, as well as how a single noise pixel is affected, by TV regularization. The domain of the image for all examples in one dimension will be the interval $[0, 1]$. (The results that we present are valid, of course, for any domain.) For discrete images, the grid is divided into n intervals, so each pixel will have a width of $\frac{1}{n}$.

For our first example, z is defined as

$$z(x) = \begin{cases} 1 & \text{if} \quad u \,\epsilon\, \Omega_1 \\ 0 & \text{if} \quad u \,\epsilon\, \Omega_2 \end{cases}$$

as shown in Figure 4. We assume that α is not large enough to flatten out any edges.

In order to solve this problem analytically, we begin by using a bit of intuition. (We also rely on our past experience in the numerical results we have seen in applying TV regularization to various images.) For this example, the total variation $TV(z)$ of z is simply equal to 1, the absolute sum of the jump in z. Thus the only way to decrease the total variation of z is to decrease the size of the jump. The problem is to find the image u which is close to the measured image z, but which has a smaller total variation than z. It would seem natural to expect u to assume the same shape as z, in order to keep the fitting error as small as possible, but with a smaller jump, so as to decrease the total variation. This would be accomplished by a u of the form

$$u(x) = \begin{cases} 1 - \delta_1 & \text{if} \quad x \,\epsilon\, \Omega_1 \\ 0 + \delta_2 & \text{if} \quad x \,\epsilon\, \Omega_2 \end{cases}$$

Figure 4 displays the relationship between u and z which we have just described.

If u has this form, then solving for u in (1) is equivalent to solving for $\{\delta_i\}$ in the problem

$$\min_{\delta_1, \delta_2} \tfrac{1}{2}(|\Omega_1|\delta_1^2 + |\Omega_2|\delta_2^2) + \alpha(H - \delta_1 - \delta_2).$$

To solve for each δ_i, we differentiate with respect to δ_i for $i = 1$ and 2, and find the resulting equations and solutions δ_i

$$|\Omega_i|\, \delta_i - \alpha = 0 \qquad \Rightarrow \qquad \delta_i = \frac{\alpha}{|\Omega_i|} \tag{8}$$

This result can be generalized (specifically to the case where noise is present) and proven more rigorously in the following Theorem:

Theorem 1 *Suppose the measured image $z(x)$ is defined so that*

$$\begin{aligned} mean_{\Omega_1} z &= 1 \\ mean_{\Omega_2} z &= 0 \end{aligned}$$

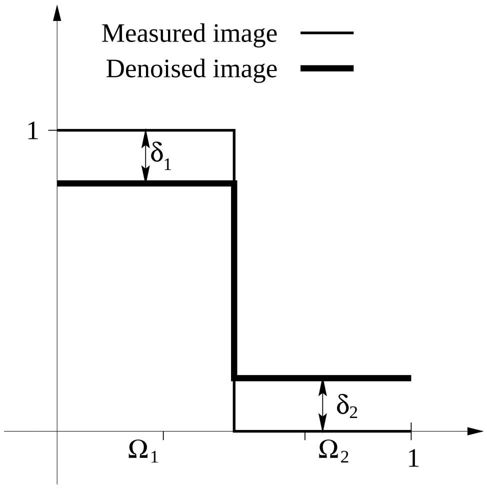

Figure 4: *A piecewise constant image, and the effect that TV regularization has on the image.*

such that $\max_{\Omega_2} z \le \min_{\Omega_1} z$.

Assume that α is chosen (sufficiently large, as well as sufficiently small) so that

$$\max_{\Omega_2} z \le \frac{\alpha}{|\Omega_2|} \le 1 - \frac{\alpha}{|\Omega_1|} \le 1 - \min_{\Omega_1} z$$

Then the unique argument to the minimization problem

$$\min_u \ \tfrac{1}{2}\|u - z\|_\Omega^2 + \alpha\, TV(u),$$

is given by

$$u(x) = \begin{cases} 1 - \frac{\alpha}{|\Omega_1|} & \textit{if } \ x \,\epsilon\, \Omega_1 \\ 0 + \frac{\alpha}{|\Omega_2|} & \textit{if } \ x \,\epsilon\, \Omega_2 \end{cases} \tag{9}$$

Proof. Define $f(v) = \frac{1}{2}\|v - z\|_\Omega^2 + \alpha\, TV(v)$. We will show that $f(u+\epsilon) > f(u)$ for any $\epsilon = \epsilon(x)$ (unless $\epsilon \equiv 0$), where u is as defined in (9). First we note that the total variation of $u + \epsilon$ can be written in terms of the total variation of u, the total variation of ϵ, and $j(\epsilon)$, the jump (if any) of ϵ at the boundary point between Ω_1 and Ω_2:

$$TV(u+\epsilon) = TV(u) + TV(\epsilon) - 2J(\epsilon)$$

where

$$J(\epsilon) = \begin{cases} 0 & \text{if } \ \text{jump}(\epsilon) \le 0 \\ jump(\epsilon) & \text{if } \ \text{jump}(\epsilon) > 0 \end{cases}$$

Then we have

$$\begin{aligned} f(u+\epsilon) &= \tfrac{1}{2}\int_\Omega (u + \epsilon - z)^2\, dx + \alpha\, TV(u+\epsilon) \\ &= \tfrac{1}{2}\int_\Omega (u - z)^2\, dx + \int_\Omega \epsilon(u - z)\, dx + \tfrac{1}{2}\int_\Omega \epsilon^2\, dx + \alpha\,(TV(u) + TV(\epsilon) - 2J(\epsilon)) \\ &= f(u) + g(u, \epsilon) \end{aligned}$$

where

$$g(u,\epsilon) = \int_\Omega \epsilon(u-z)\,dx + \tfrac{1}{2}\int_\Omega \epsilon^2\,dx \ + \alpha\,(TV(\epsilon) - 2J(\epsilon)).$$

We next see that

$$\begin{aligned}
\int_\Omega \epsilon(u-z)\,dx &= \int_{\Omega_1} \epsilon(u-z)\,dx + \int_{\Omega_2} \epsilon(u-z)\,dx \\
&= -\int_{\Omega_1} \epsilon(z-u)\,dx + \int_{\Omega_2} \epsilon(u-z)\,dx \\
&\geq -\int_{\Omega_1} \max_{\Omega_1} \epsilon\,(z-u)\,dx + \int_{\Omega_2} \min_{\Omega_2} \epsilon\,(u-z)\,dx \\
&= -\max_{\Omega_1} \epsilon \int_{\Omega_1} (z-u)\,dx + \min_{\Omega_2} \epsilon \int_{\Omega_2} (u-z)\,dx \\
&= -\max_{\Omega_1} \epsilon\,(mean_{\Omega_1}(z-u))|\Omega_1| + \min_{\Omega_2} \epsilon\,(mean_{\Omega_2}(u-z))|\Omega_2| \\
&= -\max_{\Omega_1} \epsilon\,(mean_{\Omega_1}(z) - mean_{\Omega_1}(u))|\Omega_1| + \min_{\Omega_2} \epsilon\,(mean_{\Omega_2}(u) - mean_{\Omega_2}(z))|\Omega_2| \\
&= -\max_{\Omega_1} \epsilon(1 - (1 - \frac{\alpha}{|\Omega_1|})|\Omega_1| + \min_{\Omega_2} \epsilon(0 + \frac{\alpha}{|\Omega_2|} - 0)|\Omega_2| \\
&= -\alpha(\max_{\Omega_1} \epsilon - \min_{\Omega_2} \epsilon).
\end{aligned}$$

Finally, to remove the $J(\epsilon)$ term from the derivations, define

$$\tilde{\epsilon} = \begin{cases} \epsilon & \text{if} \quad x \,\epsilon\, \Omega_1 \\ \epsilon - J(\epsilon) & \text{if} \quad x \,\epsilon\, \Omega_2 \end{cases}$$

Because of how $J(\epsilon)$ is defined, we will have $TV(\tilde{\epsilon}) = TV(\epsilon) - J(\epsilon)$. Also note that we then have

$$\begin{aligned}
\max_{\Omega_1} \tilde{\epsilon} &= \max_{\Omega_1} \epsilon \\
\min_{\Omega_2} \tilde{\epsilon} &= \min_{\Omega_2} \epsilon \ - J(\epsilon).
\end{aligned}$$

So we have

$$\begin{aligned}
g(u,\epsilon) &= \int_\Omega \epsilon(u-z)\,dx + \tfrac{1}{2}\int_\Omega \epsilon^2\,dx + \alpha(TV(\epsilon) - 2J(\epsilon)) \\
&\geq -\alpha(\max_{\Omega_1} \epsilon - \min_{\Omega_2} \epsilon) + \alpha(TV(\epsilon) - 2J(\epsilon)) + \tfrac{1}{2}\int_\Omega \epsilon^2\,dx \\
&= -\alpha(\max_{\Omega_1} \tilde{\epsilon} - (\min_{\Omega_2} \tilde{\epsilon} + J(\epsilon)) + \alpha(TV(\tilde{\epsilon}) + J(\epsilon) - 2J(\epsilon)) + \tfrac{1}{2}\int_\Omega \epsilon^2\,dx \\
&= -\alpha(\max_{\Omega_1} \tilde{\epsilon} - \min_{\Omega_2} \tilde{\epsilon}) + \alpha TV(\tilde{\epsilon}) + \tfrac{1}{2}\int_\Omega \epsilon^2\,dx \\
&\geq \tfrac{1}{2}\int_\Omega \epsilon^2\,dx \\
&> 0 \qquad (\text{unless} \ \ \epsilon \equiv 0).
\end{aligned}$$

So we've shown that $f(u+\epsilon) = f(u) + g(u,\epsilon)$, where $g(u,\epsilon) > 0$ unless $\epsilon \equiv 0$. Therefore u is the unique minimizer of this problem.

$\square$

This simple result can be extended to more complicated piecewise continuous images, as follows:

Theorem 2 *Suppose the measured image $z(x)$ is defined (as shown in Figure 5) as*

$$z(x) = \begin{cases} \frac{1}{2} & if \;\; x \in \Omega_1 \\ 1 & if \;\; x \in \Omega_1 \\ \frac{1}{2} & if \;\; x \in \Omega_1 \\ 0 & if \;\; x \in \Omega_1 \\ 1 & if \;\; x \in \Omega_1 \\ 0 & if \;\; x \in \Omega_1 \end{cases}$$

Assume that α is chosen sufficiently small (so as not to remove edges). Then the unique argument to the minimization problem

$$\min_u \;\; \frac{1}{2}\|u - z\|_\Omega^2 + \alpha\, TV(u),$$

is given by

$$u(x) = \begin{cases} \frac{1}{2} + \frac{\alpha}{|\Omega_1|}\,\alpha & if \;\; x \in \Omega_1 \\ 1 - \frac{2}{|\Omega_2|}\,\alpha & if \;\; x \in \Omega_2 \\ \frac{1}{2} & if \;\; x \in \Omega_3 \\ 0 + \frac{2}{|\Omega_4|}\,\alpha & if \;\; x \in \Omega_4 \\ 1 - \frac{2}{|\Omega_5|}\,\alpha & if \;\; x \in \Omega_5 \\ 0 + \frac{1}{|\Omega_6|}\,\alpha & if \;\; x \in \Omega_6 \end{cases}$$

The results from Theorem 1 can be used to prove the results claimed in Theorem 2. Also, this theorem could be generalized from the case "$z(x) \equiv z_i$ in region Ω_i" to the case where "$mean(z(x)) = z_i$ in region Ω_i" (where noise is present), with the same resulting solution u, under certain conditions (except in region Ω_3, where we would have $u(x) \approx \frac{1}{2}$ for $x \in \Omega_3$, rather than $u(x) = \frac{1}{2}$ for $x \in \Omega_3$). Finally, Theorem 2 and its extension to noisy images can be extended to any piecewise constant function in one dimension. See Figure 6 for a comparison of a noise-free image to a noisy image.

In reviewing the results of Theorems 1 and 2, we see that there are several important things to notice about the solution u in relation to the measured image z. First, the δ_i (change in intensity in each region) background, is changed) are mutually independent. This is true as long as α is not large enough to flatten out any boundaries. Also, for larger α, the intensity in each region is changed more, so that the contrast between the various objects in the image, including the background, is reduced. We can also deduce how TV regularization would affect a single noise element, which in the discrete case would be an object of pixel

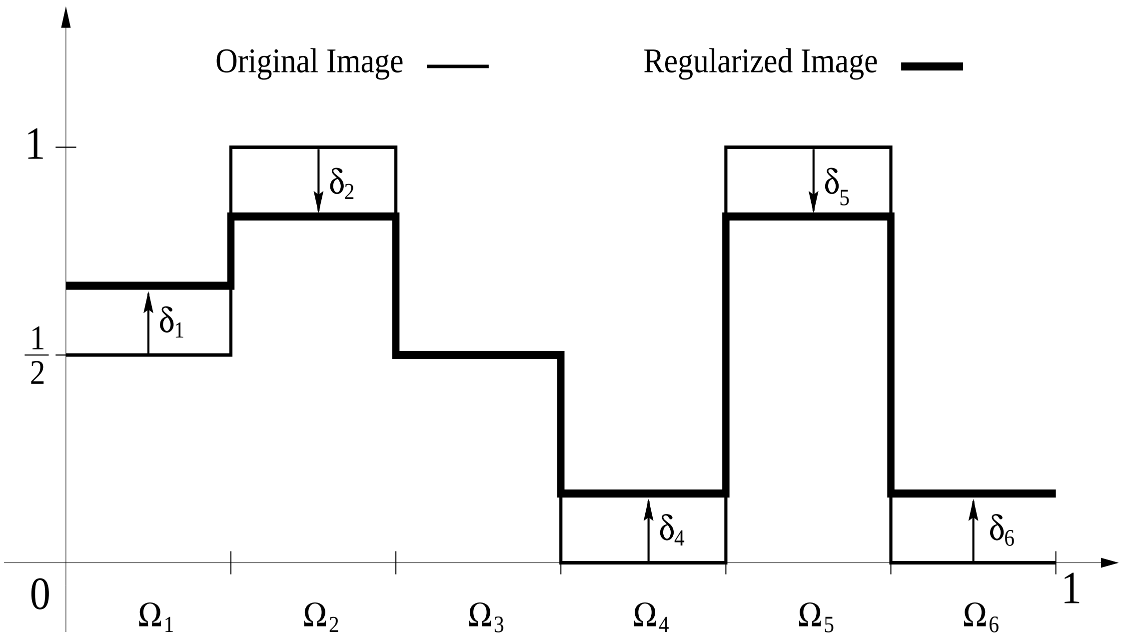

Figure 5: *A piecewise continuous one-dimensional image, before and after regularization.*

width one. Notice that each δ_i depends only on the scale or width $|\Omega_i|$ of the object, and on α, but each δ_i is independent of I_i, the original intensity in region Ω_i. We note that if α is large enough, then the object is completely flattened out, as illustrated previously in Figure 3. We define α_0 as the smallest value of α which results in a completely flat or constant-valued u. Of course, if $\alpha > \alpha_0$, then the result is still the same, since the completely flattened out object cannot be flattened any further. If $\alpha \geq \alpha_0$, then the regularized image is constant-valued at the mean of the noisy image over the whole domain,

$$u \equiv \frac{\int_\Omega z}{|\Omega|}.$$

This formula is also true for two- and three-dimensional images for $\alpha \geq \alpha_0$.

Quite often we are interested in the effect of the regularization on a particular object or local feature of an image. We define α_f as the α which causes an object of constant intensity I and width $|\Omega|$, such as shown in Figure 5, to be reduced to the fraction f of its original intensity (that is $\delta = (1-f)I$). By using the results of (8), we have

$$\alpha_f = \frac{(1-f)I|\Omega|}{2}. \tag{10}$$

2.2 Two-dimensional and Three-dimensional Images

We can extend the results developed in one dimension to the two- and three-dimensional cases. First we define

$$\partial\Omega_{i,j} \equiv \partial\Omega_i \cap \partial\Omega_j,$$

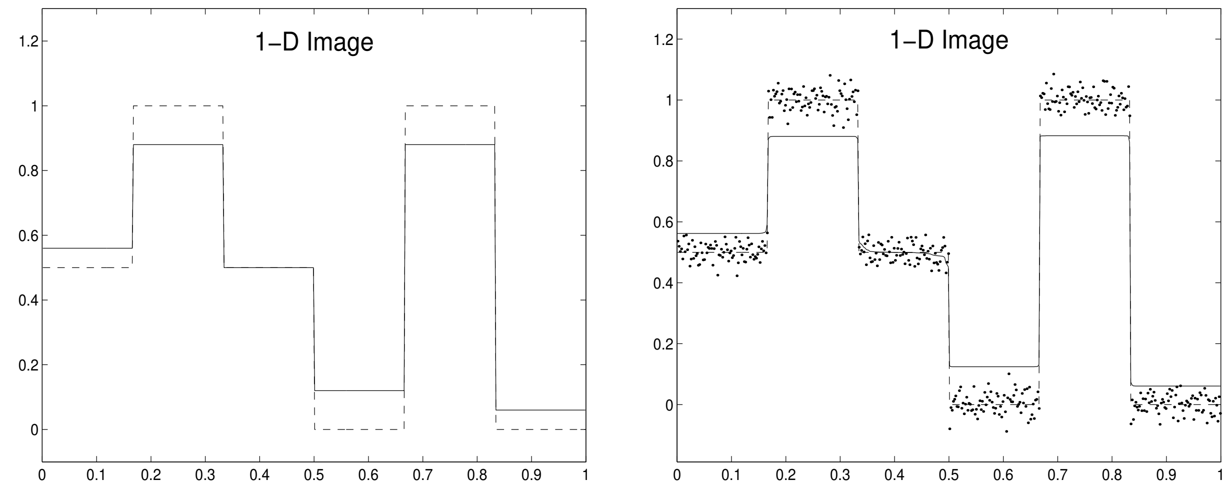

Figure 6: *A noise-free image and the same image with noise added, before and after regularization. In both images, the dashed line is the true image, the solid line is the regularized or denoised image, and in the second image, the dotted line is the noisy image.*

Next, we introduce a Lemma which we will use in determining the total variation of u in two- or three-dimensional images. The Lemma simply says that the total variation of u is the product of the jump $|A-B|$ and the size of the boundary of the jump $|\partial\Omega_{1,2}|$.

Lemma 2 *Let Ω be a compact domain in R^2 or R^3. Let u be defined in Ω as*

$$u = \begin{cases} A & \text{if } u \in \Omega_1 \\ B & \text{if } u \in \Omega_2 \;\; (= \Omega \backslash \Omega_1) \end{cases}$$

(an example of this is shown in Figure 7).

Then

$$TV(u) = |A-B|\,|\partial\Omega_{1,2}|.$$

We note that in two dimensions, $|\partial\Omega_{1,2}|$ is the length of the jump between regions Ω_1 and Ω_2, and in three dimensions, $|\partial\Omega_{1,2}|$ is the surface area of the jump between regions Ω_1 and Ω_2.

We now consider the example shown in Figure 8. Here z is defined as

$$z = \begin{cases} 1 & \text{if } z \in \Omega_1 \\ 0 & \text{if } z \in \Omega_2 \end{cases}$$

We find u in essentially the same way as we found u in the one-dimensional case in section 2.1. To do this, we assume that u will assume the same shape as z, that is, that the boundary $\partial\Omega_1$ will not be deformed. We also assume that α is not large enought to completely flatten

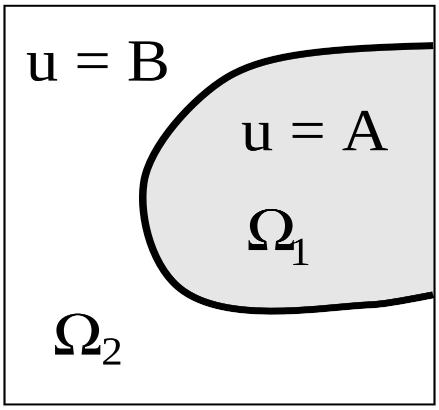

Figure 7: *$u = A$ in Ω_1, and $u = B$ in $\Omega \backslash \Omega_1$. The total variation of u is $TV(u) = |A - B|\,|\partial\Omega_{1,2}|$.*

out the image, which (as we will see) means more precisely that $0 + \frac{|\partial\Omega_{1,2}|}{|\Omega_2|} \leq 1 - \frac{|\partial\Omega_{1,2}|}{|\Omega_1|}$. Let δ_1 be the amount by which the intensity of z is reduced in region Ω_1, and let δ_2 be the amount by which z is increased in region Ω_2, to form the solution u. The fitting error between u and z will be

$$\begin{aligned} \|u - z\|_\Omega^2 &= \int_\Omega (u - z)^2 \, dx \\ &= \int_{\Omega_1} \delta_1^2 \, dx + \int_{\Omega_2} \delta_2^2 \, dx \\ &= |\Omega_1| \, \delta_1^2 + |\Omega_2| \, \delta_2^2 \end{aligned}$$

and the according to Lemma 2, the total variation of u will be $TV(u) = (1 - \delta_1 - \delta_2)\,|\partial\Omega_{1,2}|$. To solve for u, we can solve for δ_1 and δ_2 in

$$\min_{\delta_1, \delta_2} \tfrac{1}{2}|\Omega_1| \, \delta_1^2 + |\Omega_2| \, \delta_2^2 + \alpha \, (1 - \delta_1 - \delta_2) \, |\partial\Omega_{1,2}| \tag{11}$$

To solve for each δ_i, we differentiate with respect to δ_i, and find

$$|\Omega_i| \, \delta_i - \alpha \, |\partial\Omega_{1,2}| = 0 \qquad \Rightarrow \qquad \delta_i = \frac{|\partial\Omega_{1,2}|}{|\Omega_i|} \, \alpha$$

We state these results (which have just proved) in the following Theorem:

Theorem 3 *Given a compact domain Ω in R^2, let z be defined as*

$$z = \begin{cases} 1 & if \;\; u \, \epsilon \, \Omega_1 \\ 0 & if \;\; u \, \epsilon \, \Omega_2 \;\; (= \Omega \backslash \Omega_1) \end{cases}$$

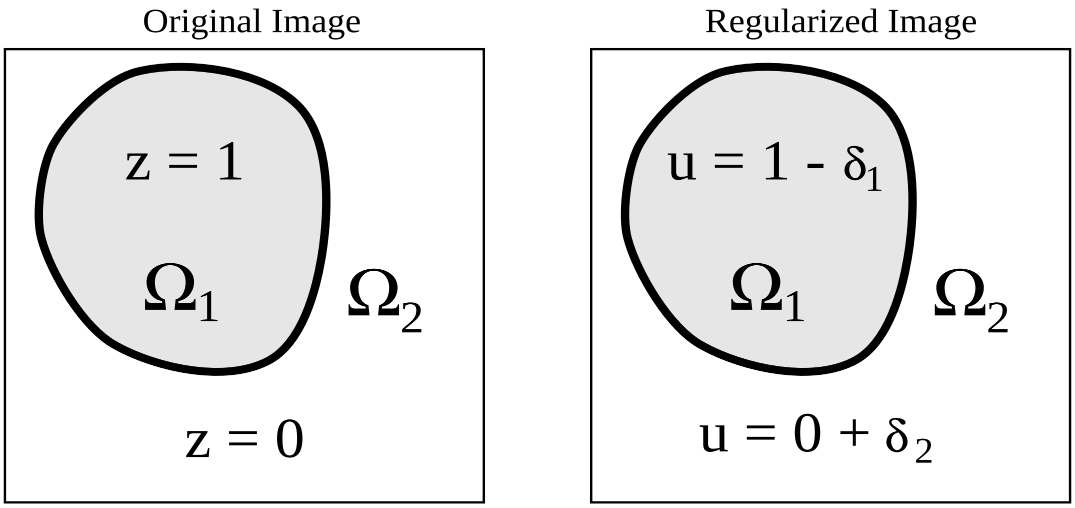

Figure 8: *2-D region-wise constant image, or cross-section of 3-D region-wise constant image*

as shown in Figure 8. Given z, let u be the argument to the minimization problem

$$\min_u \tfrac{1}{2}\|u - z\|^2 + \alpha\, TV(u). \tag{12}$$

If we assume that u will assume the same shape as z, that is,

$$u = \begin{cases} 1 - \delta_1 & if u \in \Omega_1 \\ 0 + \delta_2 & if u \in \Omega_2 \; (= \Omega\Omega_1) \end{cases}$$

and assume that α is not large enough to flatten out the image, then

$$\begin{aligned} \delta_1 &= \frac{|\partial\Omega_{1,2}|}{|\Omega_1|}\,\alpha, \\ \delta_2 &= \frac{|\partial\Omega_{1,2}|}{|\Omega_2|}\,\alpha. \end{aligned}$$

We note that in two dimensions, $|\Omega_i|$ is the area of Ω_i, and in three dimensions, $|\Omega_i|$ is the volume of region Ω_i.

There are two conditions which are quite often encountered in practice. First Ω_1 is often completely contained within Ω, so that $\partial\Omega_{1,2} = \partial\Omega_1$. Second, the area of the background region is often much larger than the area of the object of interest, and consequently the the background region is relatively unaffected by the regularization. That is, $|\Omega_2| \gg |\Omega_1| \Rightarrow \delta_2 \ll \delta_1$. When these two conditions are met, we are most interested in the regularization-scale relationship

$$\delta_1 = \frac{|\partial\Omega_1|}{|\Omega_1|}\alpha, \tag{13}$$

where Ω_1 is the region of the object or local feature of interest.

If we define the scale, s, of the object as the ratio of the boundary to the size of the object

$$s = \frac{|\partial\Omega|}{|\Omega|}, \tag{14}$$

then (13) becomes

$$\delta = \frac{\alpha}{s}. \tag{15}$$

So the smaller the scale of an object, the more the intensity is changed due to regularization, while the noise is removed. This is important in understanding how the regularization affects the contrast in the image.

In Theorem 3, we assumed that u would assume the same shape as z. This condition is not exactly met in practice. It will depend how smooth the boundary $\partial\Omega$ is in the measured image z. If the size of the boundary $|\partial\Omega|$ were to decrease, the total variation of u would decrease (remember that we are trying to find a u which is close to z but with a smaller total variation). In general it can happen that the shape of $\partial\Omega$ may be deformed. However, if $\partial\Omega$ were to be deformed, the fitting error would increase, usually by much more than the total variation is reduced. Consequently, the deformation of the boundary of the image $\partial\Omega$ will often be insignificant. See section 2.4 for a brief look at a numerical example which illustrates how the edges of an object may be deformed due to regularization.

Finally, (13) and (15) can be used to define (10) more generally so that

$$\alpha_f = \frac{(1-f)I|\Omega|}{|\partial\Omega|} = \frac{(1-f)I}{s}. \tag{16}$$

This definition will be of some use, as will be discussed later.

2.3 Radially Symmetric Objects in Two-dimensional Images

One special case of two- and three-dimensional images are those that are radially symmetric. We will shown that the boundaries of a two- or three-dimensional object are not deformed at all (only the contrast is affect), as in the one-dimensional case, for objects which are radially symmetric.

We can write $|\nabla u|$ in polar coordinates. We will treat the two-dimensional case for now. For this we have

$$\begin{aligned} x &= r\,cos\theta, \\ y &= r\,sin\theta, \\ r &= \sqrt{x^2+y^2}. \end{aligned}$$

so that

$$\begin{aligned} r_x &= \frac{x}{r}, \\ r_y &= \frac{y}{r}. \end{aligned}$$

Then since $u_\theta = 0$, we have

$$u_x = u_r r_x + u_\theta \theta_x = u_r r_x = \frac{x}{r} u_r.$$

Similarly, $u_y = \frac{y}{r} u_r$. So

$$|\nabla u| = \sqrt{u_x^2 + u_y^2} = \sqrt{(\frac{x}{r} u_r)^2 + (\frac{y}{r} u_r)^2} = \sqrt{\frac{x^2 + y^2}{r^2} u_r^2} = |u_r|$$

We consequently have

$$\int_{-1}^{1} \int_{-1}^{1} |\nabla u| \, dx \, dy = 2\pi \int_0^1 r \, |u_r| \, dr$$

If z is also rotationally invariant (if it weren't, we wouldn't expect u to be), then the fitting error (in the region of the unit circle) is given by

$$\int_{-1}^{1} \int_{-1}^{1} \frac{1}{2}(u(x,y) - z(x,y))^2 \, dx \, dy = 2\pi \int_0^1 r(u(r) - z(r))^2 \, dr$$

In one dimension, we recall that the problem we wish to solve is to find the minimizer u of

$$\int_0^1 \frac{1}{2}(u(x) - z(x))^2 + \alpha |u_x| \, dx,$$

In two dimensions, if u and z are rotationally invariant, then we wish to find the u which minimizes

$$2\pi \int_0^1 r \, [\frac{1}{2}(u(r) - z(r))^2 + \alpha |u_r|] \, dr$$

This is easily extended to three dimensions, in which the problem is to minimize

$$4\pi^2 \int_0^1 r^2 \, [\frac{1}{2}(u(r) - z(r))^2 + \alpha |u_r|] \, dr$$

The problems in the radially-symmetric two- and three-dimensional cases have reduced a problem that is almost exactly the same as the problem in the one-dimensional case (only differing by the weighting factor r or r^2 in the integrand). We might expect similar results as in the one-dimensional case, that is, preserved edges but with a change in intensity (which results in a reduction of contrast). We give the following theorem:

Theorem 4 *For a two- or three-dimensional image, suppose the measured image z (where $z = z(r)$ is a function only of radial position r) is defined as*

$$z(r) = \begin{cases} 1 & \textit{if } r \in \Omega_1 \\ 0 & \textit{if } r \in \Omega_2 \end{cases}$$

Assume that α is chosen sufficiently small (to not remove edges). Then the unique argument to the minimization problem

$$\min_u \ \tfrac{1}{2}\|u - z\|_\Omega^2 + \alpha \, TV(u),$$

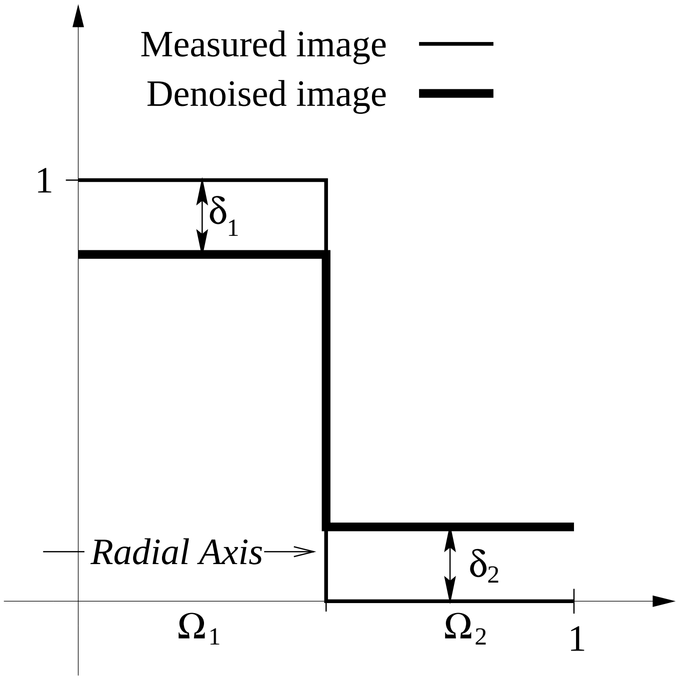

Figure 9: *A one-dimensional cross section of a regionwise continuous two- or three-dimensional image, before and after regularization.*

is given by

$$u(r) = \begin{cases} 1 - \frac{|\partial\Omega_{1,2}|}{|\Omega_1|} & if\;\; x \in \Omega_1 \\ 0 + \frac{|\partial\Omega_{1,2}|}{|\Omega_2|} & if\;\; x \in \Omega_2 \end{cases} \tag{17}$$

We give a proof only for the two-dimensional case, since the proof for the three-dimensional case is almost identical. The proof is quite similar to that given for Theorem 1 in the one-dimensional case.

Proof. Define $f(v) = \frac{1}{2}\|v - z\|_\Omega^2 + \alpha TV(v)$. We will show that $f(u+\epsilon) > f(u)$ for any $\epsilon = \epsilon(x)$ (unless $\epsilon \equiv 0$), where u is as defined in (17). Since z is radially symmetric, we assume that ϵ is also radially symmetric. First we note that the total variation of $u + \epsilon$ can be written in terms of the total variation of u, the total variation of ϵ, and the jump of ϵ at the boundary $\partial\Omega_{1,2}$ between Ω_1 and Ω_2:

$$TV(u+\epsilon) = TV(u) + TV(\epsilon) - 2J(\epsilon)$$

where

$$J(\epsilon) = \begin{cases} 0 & \text{if } \;\text{jump}(\epsilon) \le 0 \\ jump(\epsilon) & \text{if } \;\text{jump}(\epsilon) > 0 \end{cases}$$

Then we have

$$f(u+\epsilon) \;\; = \;\; \tfrac{1}{2} \cdot 2\pi \int_\Omega r(u+\epsilon-z)^2\, dr + \alpha TV(u+\epsilon)$$

$$
\begin{aligned}
&= \tfrac{1}{2}\cdot 2\pi \int_\Omega r(u-z)^2\,dr + 2\pi\int_\Omega r\epsilon(u-z)\,dr + \tfrac{1}{2}\cdot 2\pi\int_\Omega r\,\epsilon^2\,dr + \alpha\,(TV(u)+TV(\epsilon)-2J(\epsilon)) \\
&= f(u) + g(u,\epsilon)
\end{aligned}
$$

where

$$
g(u,\epsilon) = \int_\Omega r\epsilon(u-z)\,dr + \tfrac{1}{2}\int_\Omega r\epsilon^2\,dr + \alpha\,(TV(\epsilon)-2J(\epsilon)).
$$

We next see that

$$
\begin{aligned}
2\pi\int_\Omega r\,\epsilon\,(u-z)\,dr &= 2\pi\int_{\Omega_1} r\epsilon(u-z)\,dr + 2\pi\int_{\Omega_2} r\epsilon(u-z)\,dr \\
&= -2\pi\frac{|\partial\Omega_{1,2}|}{|\Omega_1|}\alpha\int_{\Omega_1} r\epsilon\,dr + 2\pi\frac{|\partial\Omega_{1,2}|}{|\Omega_2|}\alpha\int_{\Omega_2} r\,\epsilon\,dr \\
&\geq -2\pi\frac{|\partial\Omega_{1,2}|}{|\Omega_1|}\alpha\max_{\Omega_1}\epsilon\int_{\Omega_1} r\,dr + 2\pi\frac{|\partial\Omega_{1,2}|}{|\Omega_2|}\alpha\min_{\Omega_1}\epsilon\int_{\Omega_2} r\,dr \\
&= -|\partial\Omega_{1,2}|\alpha\max_{\Omega_1}\epsilon + |\partial\Omega_{1,2}|\alpha\min_{\Omega_1}\epsilon \\
&= -\alpha|\partial\Omega_{1,2}|(\max_{\Omega_1}\epsilon - \min_{\Omega_1}\epsilon)
\end{aligned}
$$

Similar to before, to remove the $J(\epsilon)$ term from the derivations, define

$$
\tilde{\epsilon} = \begin{cases} \epsilon & \text{if} \quad x \in \Omega_1 \\ \epsilon - J(\epsilon) & \text{if} \quad x \in \Omega_2 \end{cases}
$$

so that, because of how $J(\epsilon)$ is defined, we will have $TV(\tilde{\epsilon}) = TV(\epsilon) - J(\epsilon)$. Also note that we then have

$$
\begin{aligned}
\max_{\Omega_1}\tilde{\epsilon} &= \max_{\Omega_1}\epsilon \\
\min_{\Omega_2}\tilde{\epsilon} &= \min_{\Omega_2}\epsilon\ - J(\epsilon).
\end{aligned}
$$

So we have

$$
\begin{aligned}
g(u,\epsilon) &= \int_\Omega r\,\epsilon\,(u-z)\,dx + \tfrac{1}{2}\int_\Omega r\epsilon^2\,dx + \alpha(TV(\epsilon)-2J(\epsilon)) \\
&\geq -\alpha|\partial\Omega_{1,2}|(\max_{\Omega_1}\epsilon - \min_{\Omega_2}\epsilon) + -\alpha(TV(\epsilon)-2J(\epsilon)) + \tfrac{1}{2}\int_\Omega r\epsilon^2\,dx \\
&= -\alpha|\partial\Omega_{1,2}|(\max_{\Omega_1}\tilde{\epsilon} - (\min_{\Omega_2}\tilde{\epsilon} - J(\epsilon)) + \alpha(TV(\tilde{\epsilon}) + J(\epsilon) - 2J(\epsilon)) + \tfrac{1}{2}\int_\Omega r\epsilon^2\,dx \\
&= \alpha|\partial\Omega_{1,2}|(\min_{\Omega_1}\tilde{\epsilon} - \max_{\Omega_2}\tilde{\epsilon}) + \alpha TV(\tilde{\epsilon}) + \tfrac{1}{2}\int_\Omega r\,\epsilon^2\,dx \\
&\geq \tfrac{1}{2}\int_\Omega r\,\epsilon^2\,dx \\
&> 0 \qquad (\text{unless}\ \ \epsilon \equiv 0).
\end{aligned}
$$

So we've shown that $f(u+\epsilon) = f(u) + g(u,\epsilon)$, where $g(u,\epsilon) > 0$ unless $\epsilon \equiv 0$. Therefore u is the unique minimizer of this problem.

□

This simple result can be extended to more complicated piecewise continuous images, as follows:

Theorem 5 *Suppose the measured image $z(x)$ is defined as shown in Figure 10, so that*

$$z(r) = \begin{cases} \frac{1}{2} & \textit{if } x \,\epsilon\, \Omega_1 \\ 1 & \textit{if } x \,\epsilon\, \Omega_2 \\ \frac{1}{2} & \textit{if } x \,\epsilon\, \Omega_3 \\ 0 & \textit{if } x \,\epsilon\, \Omega_4 \\ \frac{1}{2} & \textit{if } x \,\epsilon\, \Omega_5 \\ 0 & \textit{if } x \,\epsilon\, \Omega_6 \end{cases}$$

Assume that α is chosen sufficiently small (so as not to remove any edges). Then the unique argument to the minimization problem

$$\min_u \ \tfrac{1}{2}\|u - z\|^2_\Omega + \alpha\, TV(u),$$

is given by

$$u(r) = \begin{cases} \frac{1}{2} + \frac{|\partial\Omega_1|}{|\Omega_1|}\,\alpha & \textit{if } x \,\epsilon\, \Omega_1 \\ 1 - \frac{|\partial\Omega_2|}{|\Omega_2|}\,\alpha & \textit{if } x \,\epsilon\, \Omega_2 \\ \frac{1}{2} - \frac{|\partial\Omega_{3,4}| - |\partial\Omega_{2,3}|}{|\Omega_3|} & \textit{if } x \,\epsilon\, \Omega_3 \\ 0 + \frac{|\partial\Omega_4|}{|\Omega_4|}\,\alpha & \textit{if } x \,\epsilon\, \Omega_4 \\ 1 - \frac{|\partial\Omega_5|}{|\Omega_5|}\,\alpha & \textit{if } x \,\epsilon\, \Omega_5 \\ 0 + \frac{|\partial\Omega_{5,6}|}{|\Omega_6|}\,\alpha & \textit{if } x \,\epsilon\, \Omega_6 \end{cases}$$

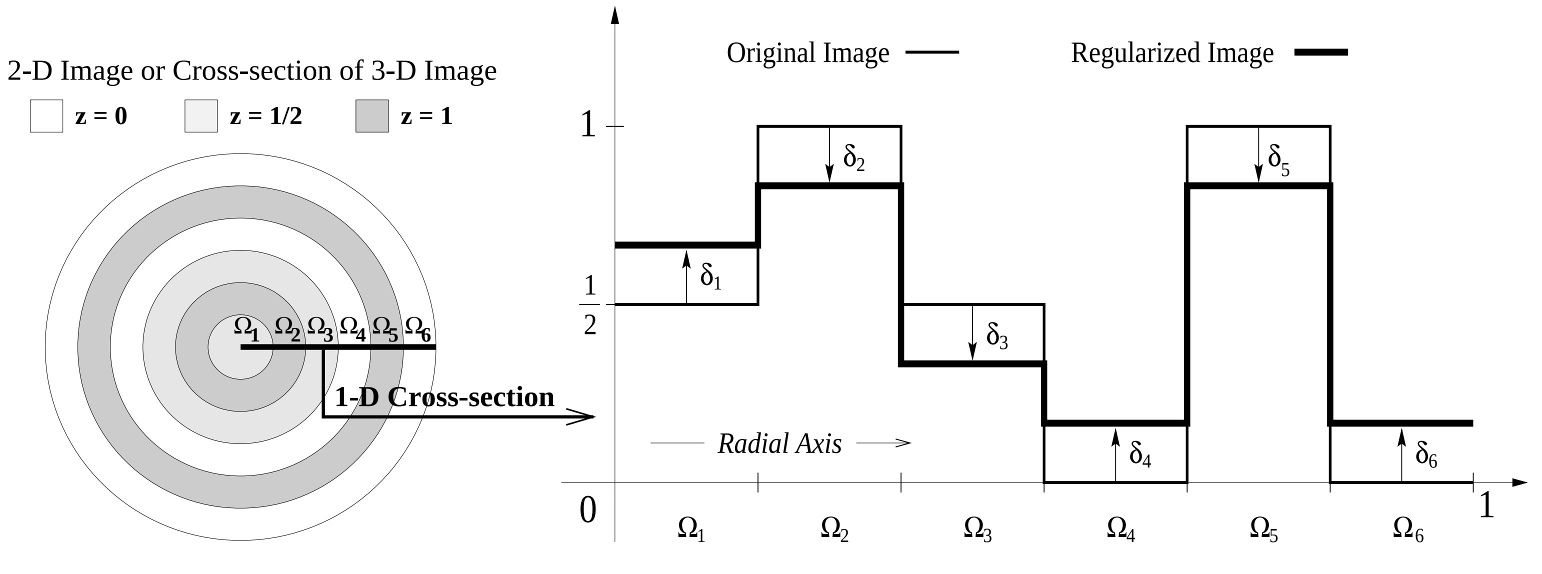

Figure 10: *1-D cross-section of radially symmetric 2-D/3-D region-wise constant*

The results from Theorem 4 can be used to prove the results claimed in Theorem 5. Also, this theorem could be generalized from the case "$z(r) \equiv z_i$ in region Ω_i" to the case where "$mean(z(r)) = z_i$ in region Ω_i", with the same resulting solution u. Finally, Theorem 5 and its extension to the noisy case can be extended to any radially piecewise constant function in two or three dimensions. See Figure 11 for a comparison of regularized noise-free images to regularized noisy images.

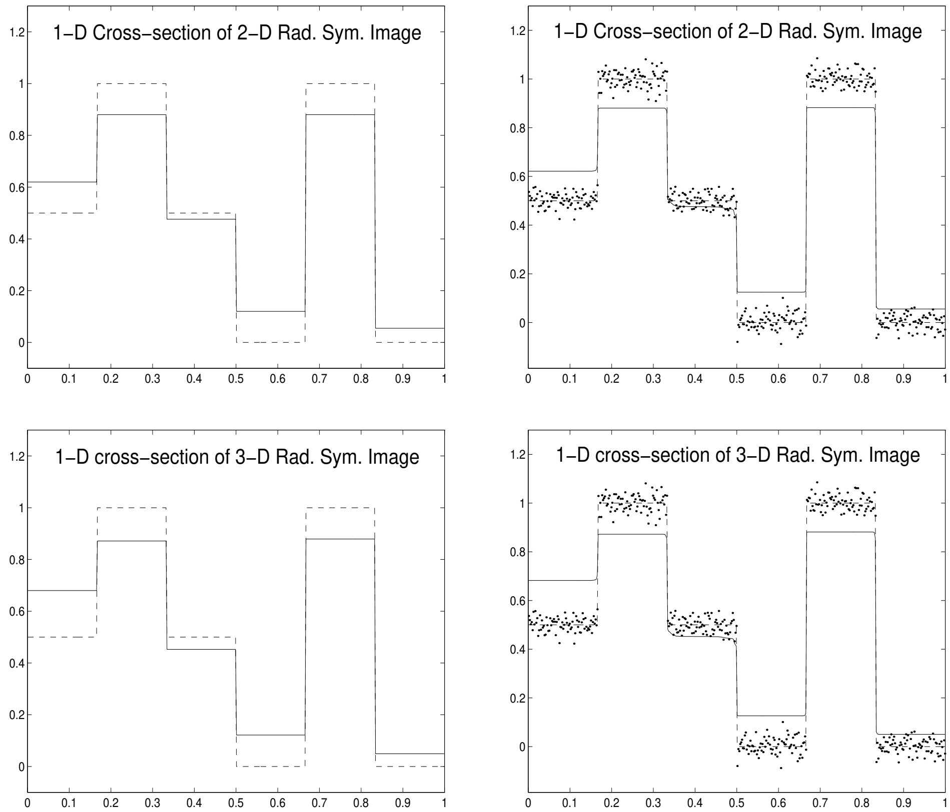

Figure 11: *Noise-free images (2-D and 3-D) and the same images with noise added, before and after regularization. In all images, the dashed line is the true image, the solid line is the regularized or denoised image, and in the images on the right, the dotted line is the noisy image.*

2.4 Numerical Approximation of Continuous Problem

The theoretical results of the simple examples discussed in Section 2.2 can be tested numerically. At this point we mention that the problem that the minimization problem which we solve in practice is slightly different from the continuous one introduced in Section 1.

In order to improve the conditioning of the problem, the TV functional (3) may be replaced by the smooth approximation

$$J(u) = \int_\Omega \sqrt{|\nabla u|^2 + \beta^2}\, dx, \tag{18}$$

without significantly affecting the sharpness of the image, if β is not too large [14]. Equation (1) then becomes

$$\min_u \frac{1}{2}\|u - z\|^2 + \alpha\, J(u). \tag{19}$$

We then wish to find the image u which minimizes (19). By finding the gradient of (19), we find that a necessary condition for a minimizer is

$$f(u) = (u - z) + \alpha\, L(u) = 0, \tag{20}$$

where

$$L(u) = -\nabla \cdot \left(\frac{\nabla u}{\sqrt{|\nabla u|^2 + \beta^2}}\right), \quad \text{with } \frac{\partial u}{\partial n} = 0 \quad \text{on} \quad \partial\Omega.$$

In practice, u is actually found using the modified formulation of the problem (19), but the agreement between the solution to the original problem (1) and the modified problem (19) is nearly exact, especially for small values of β.

In Figure 12 we give an example of a discrete image in which we see that the boundary of the object in the image is actually deformed as a result of TV regularization. The smoothing of the boundary is also a function of our choice of β. The object of interest in the image is a rectangle with width $\frac{2}{3}$ and length $\frac{1}{3}$. The boundary of the image is deformed as the (sharp) corners of the rectangle are smoothed. In this discrete case, we can see that this boundary deformation at the corners is accomplished by the pixel values at the corners being reduced relatively more than those away from the corners. Because the pixels near the corners are reduced in value by *more* than theory would predict (i.e. the theory that is based on the assumption that the boundary is not deformed), the amount by which the pixel values are reduced in the center of the object is *less* than our theory would have predicted. It is the average amount (the average inside the original boundary of the object) by which these pixels are reduced in value that is nearly equal to what our simple theory predicted. Notice that for larger values of β, the regularized image has lost some of it sharpness, while for smaller values of β (so that the discrete problem better approximates the continuous problem) the only loss of sharpness is due to the deformation of the boundary, as we would expect in the continuous problem.

It would seem that an exact theory (of the type developed when we assume that the boundary is not deformed, such as when the object of interest is symmetric) is impossible to develop in a simple form for the general case. But it seems that our theory from the ideal case can be extended to the general case, although not exactly, in a useful way.

2.5 A Heuristic for Estimating the Scale of Local Image Features

The effect of TV based regularization on a particular object or local feature is inversely proportional to its scale, as defined in (14), (15). In this section we briefly discuss how one

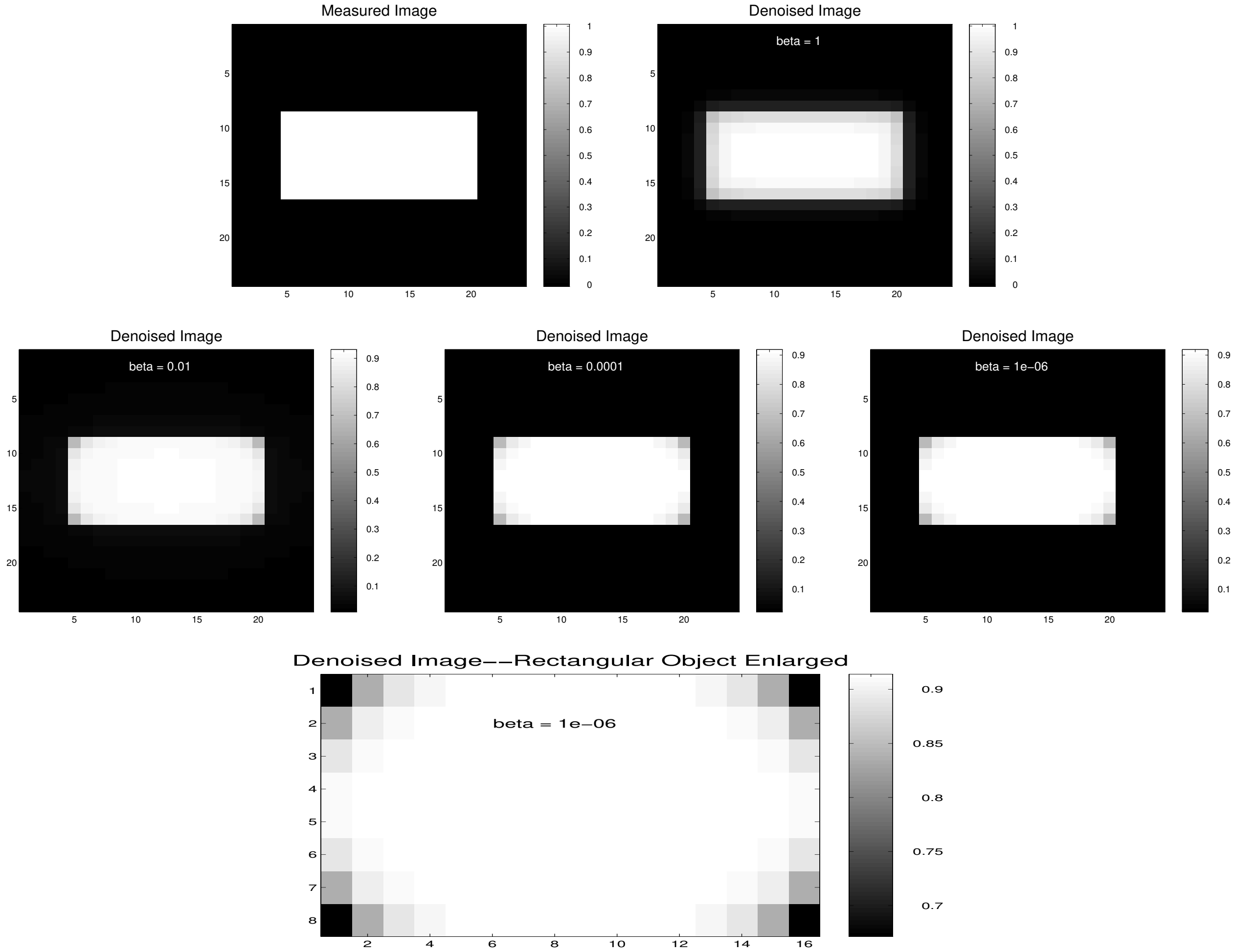

Figure 12: *The effect of varying values of β is seen in these pictures. The final picture is a zoom-in on the rectangular figure where $\beta = .000001$, so as to see in better detail the deformation of the boundary of the rectangular figure.*

might estimate the scale by knowing only the "width" of the object. We could also treat the various parts of a single object as individual features with different scales. This estimation could then be used in determing the appropriate choice for α, when choosing α based on the equation (13).

For a rectangle, let the "width" be the length of its shorter side. The scale of an object may be thought of as relating directly to its width. This idea could also be applied to objects of less uniform scale by looking at some sort of average width. We could say that objects of the same width but of differing lengths are of the same or similar scales.

In an n x n discrete image, for an object of pixel width 1 and length k, we have

$$s_k = \frac{|\Omega|}{|\partial\Omega|} = \frac{\frac{k}{n^2}}{\frac{2k+2}{n}} = \frac{k}{(2k+2)n}.$$

When $k = 1$ this ratio is $s_1 = \frac{1}{4n}$, while as $k \to \infty$ we have $s_{k\to\infty} = \frac{1}{2n}$. So for this example, for any k we will have $\frac{1}{4n} \leq ratio_k \leq \frac{1}{2n}$. Since most objects (of pixel width 1) will not have $k = 1$, we note that $s_2 = \frac{1}{3n}$ and $s_4 = \frac{2}{5n}$. So if the the only information that we had (or that we could estimate) about an object or local feature were that it is 1 pixel in width, then we might choose α based on the assumption that for this object, $s \approx \frac{2}{5n}$ may be a good estimate.

This estimate of $s = \frac{|\partial\Omega|}{|\Omega|}$ will vary depending on the particular scale. Also, in locally adaptive regularization, the estimated length of the local features can also be used to find a more precise and appropriate value of α locally. This regularization-scale relationship should be combined with any knowledge we have about the image, whether known apriori or by some sort of estimative analysis, to get the most useful results.

3 Adaptive Regularization Based on Exact Solution Theory

With the understanding of some basic theory of the effect of TV regularization on objects of any given scale, we can now show some applications of the regularization-scale relationship which we have now defined. We will consider the case in which there is more than one scale in a simple image, first without noise, and then with noise, for comparison. We discuss how to choose α based solely on the noise level, by applying the relationship in (13). We then outline a scheme for choosing α based on both the noise level and the scale of local features of the image. We then apply this scheme to the task of denoising a brain scan image, as an example of applying this scheme to a more complex image.

3.1 Multiple Local Scales in an Image

We now consider an image which is comprised of objects of differing scale. Consider the image shown in Figure 13. The two rectangular regions have widths of 3 pixels and 5 pixels. In these two regions $z = 1$, and outside of the regions $z = 0$. For now we will add no noise in order to show the effect the regularization has on the contrast in the image. In Figure 13 we show the results of TV regularization when α is chosen according to a specific scale and desired amount of contrast to preserve. Two values of α are used. We use definition (16) to find the appropriate choice for α. We first choose α based on the scale of the thinner rectangle, which is 3 pixels wide by 10 pixels long, so that

$$\alpha_{.75} = \frac{(1 - .75) \cdot 1 \cdot \frac{30}{25^2}}{\frac{26}{25}} \approx .0115$$

is chosen to preserve approximately 75% of the magnitude of the thinner rectangle, while $\alpha = .0167$ is chosen to preserve approximately 75% of the magnitude of the wider rectangle. In the final picture α varies according to the its position in the domain, and its values are chosen to preserve approximately 75% of the magnitude of both rectangles. Again, due to the deformation of the boundaries, the amount by which the intensity reduction in the

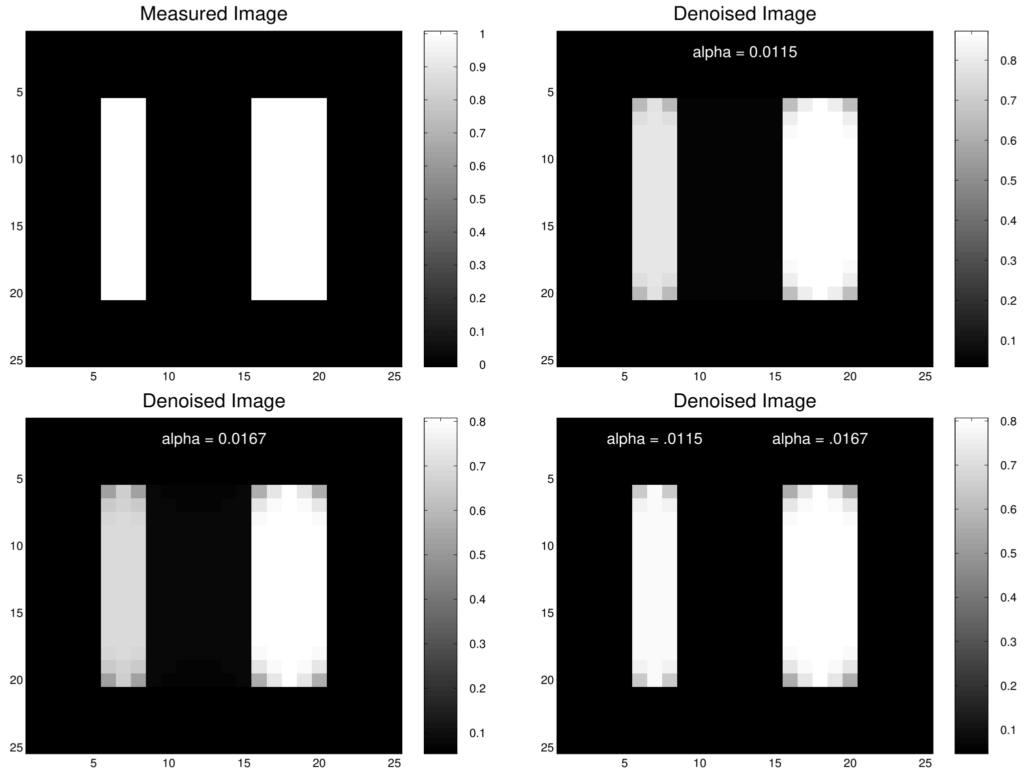

Figure 13: *A noise-free image is TV regularized to show the effects of choosing specific values of α, according to the scale of individual objects or features.*

interior of the image is slightly less than our theory would predict (as well as the intensity level at the corners of the boundaries being reduced by more than predicted, if the entire region were to remain constant-valued). This example gives a simple case of how finding a precise relationship between α and the scale may be very useful in doing multi-scale and adaptive regularization, where choice of regularization parameter depends on the scale of the individual features of the image.

For comparison, we next add Gaussian noise (mean 0) to the image, and we regularize this noisy image by using the same values of α as we used in Figure 13. In Figure 14 we see that the results are nearly the same as the results seen in Figure 13. The noise is essentially removed, and the amount of contrast after regularization is very similar to the contrast in the noise-free figures. The level of similarity will depend on the amount of noise present and the value of the regularization parameter α.

3.2 Choosing the Regularization Parameter Based on Noise in the Image

Often we base our choice of α on the noise in the image. We may try to choose α to be equal to (or as close as possible to) the Lagrange multiplier in the constrained problem (2), so that the solution to the Tikhanov regularization problem (1) is identical to the solution

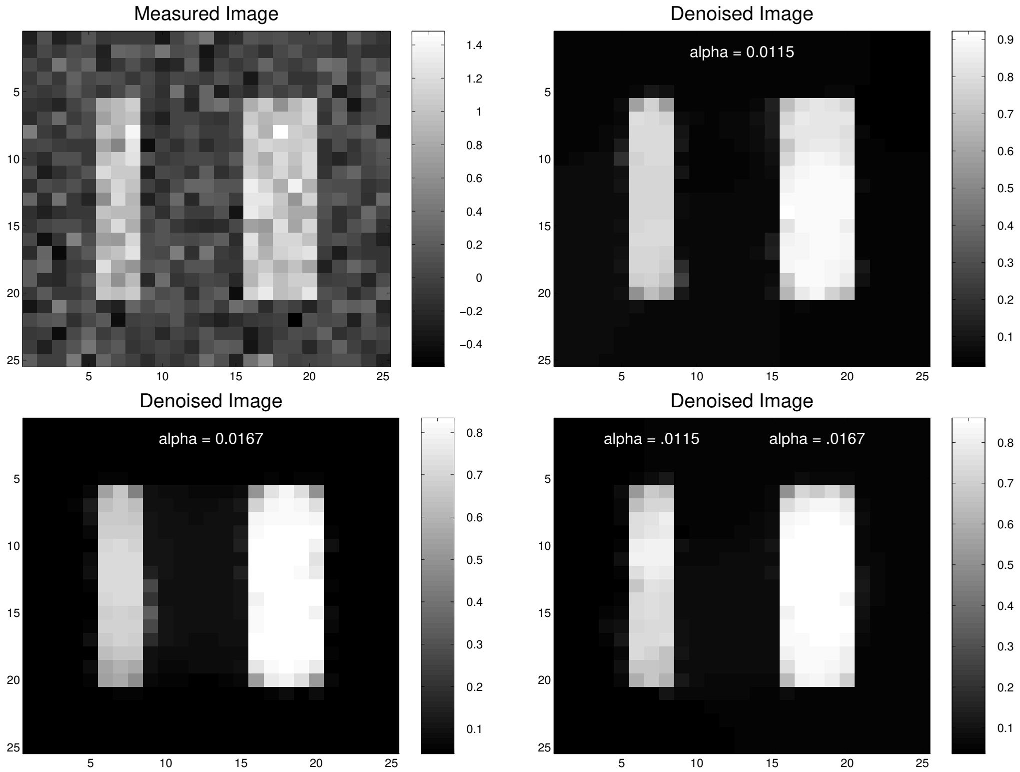

Figure 14: *A noisy image is TV regularized to show the effects of choosing specific values of α, according to the scale of individual objects or features.*

of the constrained problem (2). (There are other methods for choosing the regularization parameter—see [7]—if the noise level is not known or if it cannot be accurately estimated.) In this section we discuss an application of our regularization-scale relationship which can be used to base α on the noise in the image. This approach requires no computational work to find α if the noise level (and type of noise) is known, and it requires only the work to estimate the type and level of noise if it is unknown.

First we briefly examine the noise itself. Take for example the image in Figure 15, which is simply noise. This noise has been constructed so that the values at adjacent pixels alternate between positive and negative values (like a checkerboard). The noise values have a magnitude which are uniformly distributed between 0 and 1.

In this case the purpose is to find the smallest α, α^{noise}, such that all of the noise will be essentially removed. (Or find α^{noise} with some other heuristic—the point is to base the choice of α^{noise} only on the noise.) If α^{noise} can be found for a given level of noise, then given an image where the noise level is known or can be estimated, we can use this value of α^{noise}, and expect that the noise would be essentially smoothed out, while hopefully preserving as much detail of the image as possible. If we were to choose $\alpha > \alpha^{noise}$, then we would have at least as much smoothing as needed (in theory) to remove all of the noise, but we would also remove more of the detail. If we were to choose $\alpha < \alpha^{noise}$, then we would preserve more detail, but at the expense of allowing more noise to remain.

For this first example, we can use the results from the first example of Section 2.2. We first look at the case when the sign of the noise alternates between adjacent pixels, like a checkerboard. Then with this "checkerboard" assumption, we simply need to find the value of α that will completely smooth out the noise element with largest magnitude, which in this case is 1. Since each noise element is a single $\frac{1}{n}$ x $\frac{1}{n}$ pixel, we can use (16) to find that

$$\begin{aligned} \alpha_0 &= \frac{(1-0)\cdot 1 \cdot \frac{1}{n^2}}{\frac{4}{n}} \\ &= \frac{1}{4n}. \end{aligned}$$

We found that the theoretical value of α_0 was not quite large enough to remove all of the noise. In practice, we found that a slightly larger value of α_0 was needed. The results of letting $\alpha = \frac{1.1}{4n}$ are shown in Figure 15. Notice that the order of magnitude of the denoised image is 10^{-4} so that essentially all of the noise was smoothed out.

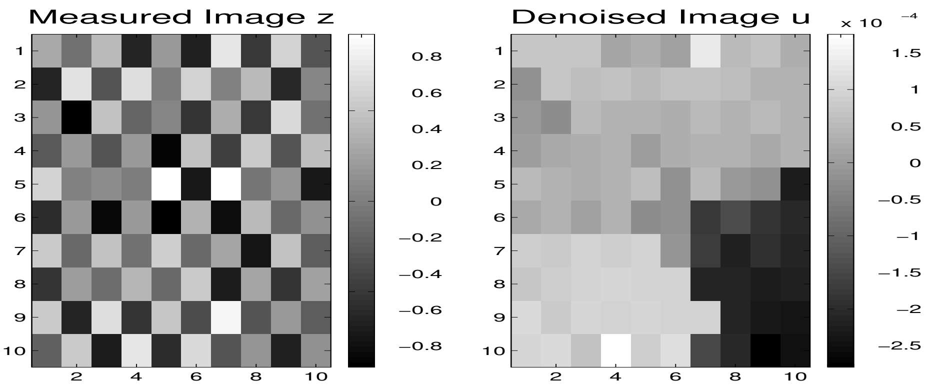

Figure 15: *The measured image is simply "checkerboard" noise. The denoised image is the results of applying TV regularization to the "checkerboard" noise, and it is essentially constant at 0. (Notice that the order of magnitude of the denoised image is* 10^{-4}*).*

The "checkerboard" noise in the previous example is generally not encountered in practice, but is a useful and demonstrative example. A more realistic example is given in Figure 16. Here the noise is Gaussian (mean 0) with standard deviation of $\frac{1}{3}$. We can then assume that most of the noise has magnitude no greater than 1. If we again choose $\alpha = \frac{1.1}{4n}$ as before, we see the results are still quite good. There is still some noise that was not smoothed out due to two things. First, some of the noise had magnitude greater than 1. Second, since we no longer had the "checkerboard" assumption, there were adjacent pixels with noise values of the same sign, which in effect created local objects of scale or width greater than one pixel. Again we could choose α to be as large as needed to completely smooth out the noise, but if we were actually removing noise from an image, larger values of α could also cause more detail to be removed.

We now consider the problem of recovering an image by removing noise from the measured data. We will first consider an image to which Gaussian noise (mean 0, two levels of variance)

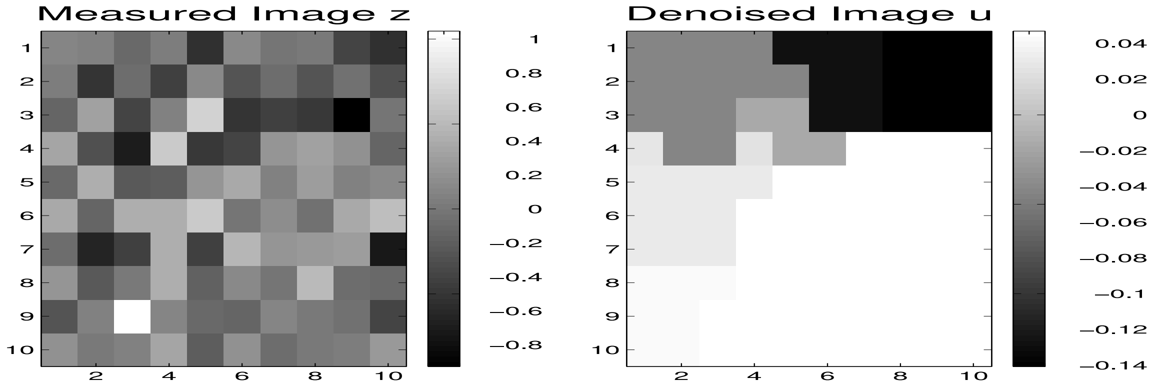

Figure 16: *The measured image is simply Gaussian noise. The denoised image is the result of TV regularization applied to the Gaussian noise.*

has been added, as shown in Figure 17. In each case, we determine a value of α^{noise}, so as to remove essentially all of the noise, but without consideration of the scale of the objects. Noise with standard deviation of $\frac{1}{3}$ was added to the first image, so that we would expect most or all of the noise to have magnitude less than 1. In this case we choose α^{noise} as before, so that $\alpha^{noise} = \frac{1.1}{4n} = .009167$, since $n = 30$. Noise of standard deviation $\frac{1}{2}$ is added to the image in the second case, so that we now choose $\alpha^{noise} = \frac{1.5 \cdot 1.1}{4n} = .01375$.

We can see that when the noise level is smaller, we could actually increase α in the region which contains the widest rectangle to remove more noise. However, where the noise level is high, we are starting to lose too much contrast in the object of smallest scale, so perhaps α should be chosen smaller in this region to preserve more contrast. We must choose α based on both noise (level and type) and the local scale of the image. One idea for doing this is described in the next section.

3.3 Choosing the Regularization Parameter Based on Both Noise and Scale

If the measured image is extremely noisy, or there is very fine detail in the image, it is not possible to remove all of the noise while preserving all of the detail of the true image. As a result of this trade off, there is a balance which must be chosen between noise reduction and detail preservation. There is no general rule to determine the optimal balance. It will depend on several factors, including the type of image being denoised and the purpose for noise removal. This should be done locally for best results.

If in a particular region or for a particular feature the expected scale and contrast are known apriori, one could put a limit on the amount of smoothing to be done in order to preserve a minimum amount of detail, even if this means allowing more noise to remain than we would otherwise allow. For example, one could decide that the maximum amount of intensity reduction of the feature is to be half of the original intensity, so as to maintain edges and reasonable contrast. This corresponds to imposing the restriction that $\delta \leq \frac{1}{2}I$.

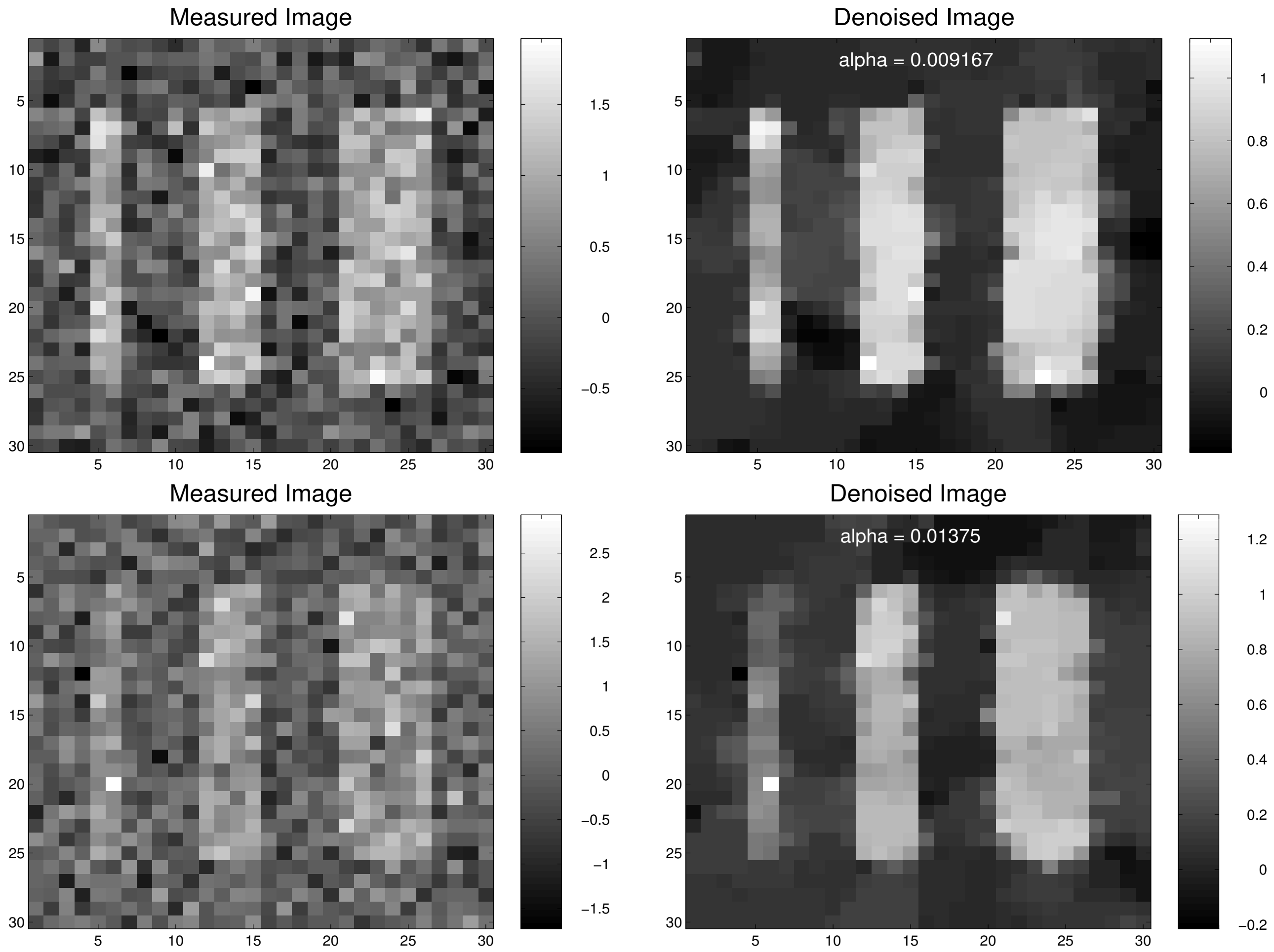

Figure 17: *An example of choosing α based only on the noise level to illustrate that the regularization parameter should also be based on the local scale in the image.*

The result is that α is restricted by some α_{max} so that

$$\alpha \leq \alpha_{\frac{1}{2}}, \tag{21}$$

where $\alpha_{\frac{1}{2}}$ is defined as in (16). Then α is chosen, depending on the noise in the measured image, but with the new restriction (21). This may be done for various regions of a single image, depending on the amount of detail we wish to preserve in each region. This is especially important in adaptive regularization.

One method in which we choose α based on both noise and local scale is now outlined. We also find the locally adaptive α_{max}^{scale} which is based solely on the scale of the image, as outlined above. We determine the value of α^{noise} which we would choose based solely on the noise level, or based on some other scheme if needed (the point is that α^{noise} is not based on scale). Then the locally adaptive regularization parameter is given by

$$\alpha(\vec{x}) = f(\alpha_{max}^{scale}(\vec{x}), \alpha^{noise}).$$

(Actually α^{noise} could be spatially variant as well.) One idea for choosing α based on both noise and local scale is

$$\alpha(\vec{x}) = \min(\alpha^{scale}_{max}(\vec{x}), \alpha^{noise}). \tag{22}$$

Other ideas could include some sort of weighted average of $\alpha^{scale}_{max}(\vec{x})$ and α^{noise} in determining $\alpha(\vec{x})$. Choosing α in this way allows us to preserve edges and to maintain a more controlled level of contrast. This could be very useful if coupled with some kind of edge or contrast enhancement done after the regularization.

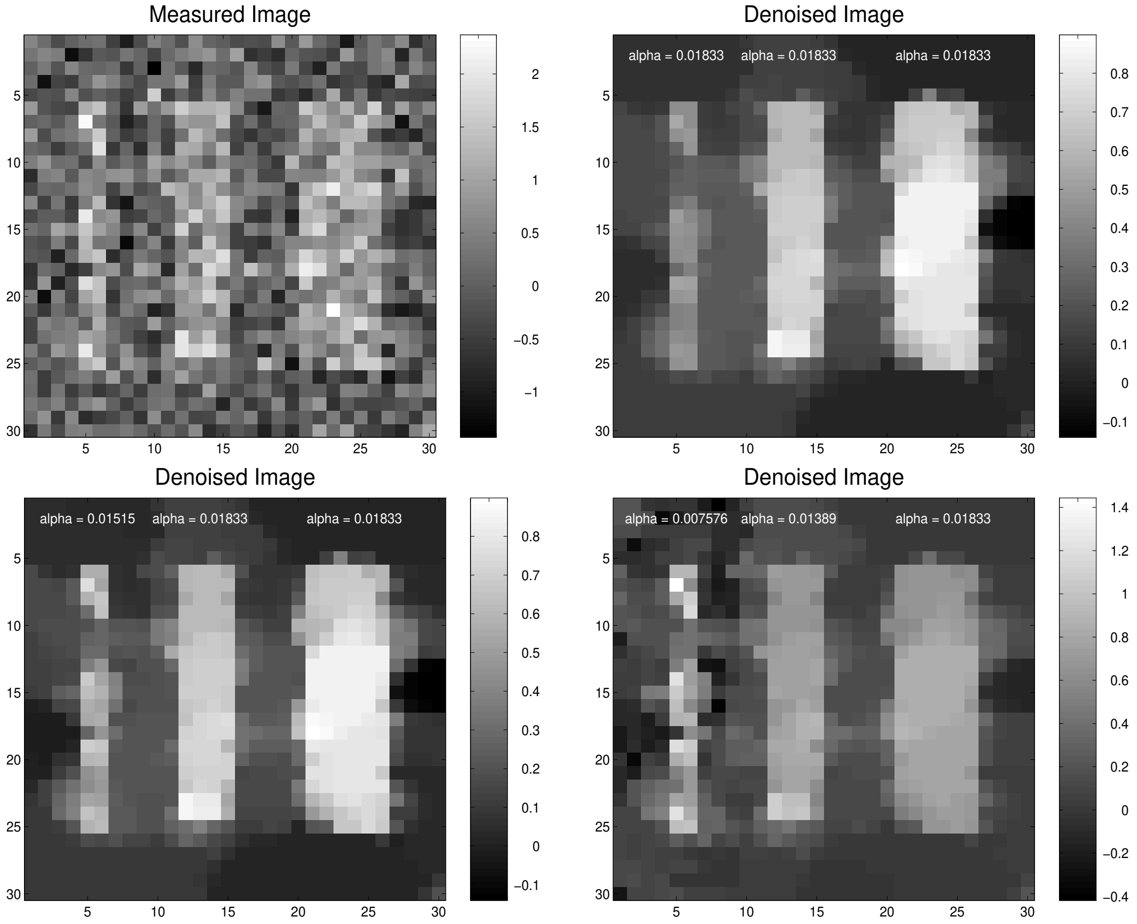

Figure 18: *In this figure, α is based on both noise and scale. We use the formula $\alpha(\vec{x}) = min(\alpha^{scale}_{max}(\vec{x}), \alpha^{noise})$. For the first image where α is allowed to vary, we choose $\alpha^{scale}_{max}(\vec{x}) = \alpha_{.5}$ in each of the three regions corresponding to each object. In the second adaptively regularized image, $\alpha^{scale}_{max}(\vec{x}) = \alpha_{.75}$*

Results from defining α according to (22), using $\alpha_{\frac{3}{4}}$ and $\alpha_{\frac{1}{2}}$ as α^{scale}_{max}, for two noise levels, are given in Figure 18. To find α^{noise} we follow the idea presented in Section 3.2 when discussing the "checkerboard" noise. The standard deviation of the noise is $\frac{1}{2}$, we would expect that most or all of the noise will have magnitude of less than 1.5. However, we choose α^{noise} to be $\alpha^{noise} = \frac{2 \cdot 1.1}{4n} = .01833$ instead of $\alpha^{noise} = \frac{1.5 \cdot 1.1}{4n} = .01375$, so as to better illustrate the results of basing our choice of α on noise and scale. The case where we use $\alpha^{scale}_{max} = \alpha_{\frac{1}{2}}$ gives us the best results. We could choose f (in α_f) smaller to remove even more noise, but we must be careful not to over-regularize resulting in excessive loss of contrast in smaller-scaled features.

3.4 Application to Brain Scan Image

We have seen that total variation based noise removal is very effective in removing noise from simple images, particularly when we use our regularization-scale relationship. We will now attempt to remove the noise from a noisy brain scan, which is a much more complicated image. This image is very useful as a test case because of its complexity.

The discrete data of the true image is a 256 x 256 array. For a better analysis, we will consider only a portion taken from the center part of the true brain image. The discrete data for this image is a 40 x 40 array. This image is corrupted with Gaussian noise (mean 0). The noise level is chosen to be relatively low, in order to better see the adaptive capabilities of this approach.

We first apply the TV regularization to this image where α is constant over the whole image. We do this for a few different values of α. The results of using constant α are shown in Figures 19 and 20. We can see that as α becomes larger, more noise is removed. This is beneficial in the larger-scaled regions, but in the smaller-scaled regions, too much detail has been lost. We need adaptive regularization for better results.

We next let α vary over the domain according to the scale of the local features of the image. We will denoise the image using the scheme described in Section 3.3. Since we have exact knowledge of the true image, we know the exact scales of the local features, which we can use to determine how to let α vary. We will find α^{scale}_{max}, as described in Section 3.3, which will be the maximum value α can assume in the various subregions of this image. We then choose $\alpha = min(\alpha^{scale}_{max}, \alpha^{noise})$, where α^{noise} is the value of α we would choose based solely on the noise level in the image. For this example we choose $\alpha^{noise} = 0.40$, which is large relative to the noise level. We do this for demonstrative reasons, since the result of choosing α^{noise} large is that $\alpha(\vec{x}) = \alpha^{scale}_{max}(\vec{x})$, except in the few relatively large regions, where we have $\alpha(\vec{x}) = \alpha^{noise}$.

For the first case where α is chosen in an adaptive way, we have chosen f in α_f to be .04 (see 16), and for the second case of adaptive α, f is chosen to be .08. This corresponds to $\min_x \alpha^{scale}_{max}(\vec{x}) = .025$ in the first case, and $\min_x \alpha^{scale}_{max}(\vec{x}) = .050$ in the second case. In both cases, since $\alpha^{noise} = 0.40$, we have $\max \alpha^{scale}_{max}(\vec{x}) = .400$. In Figure 20 we give the results of using an adaptive $\alpha(\vec{x})$. The values of $\alpha(\vec{x})$ that will control the local balance between fitting error and TV regularization, are also shown in Figure 20. The two images which were adaptively regularized give better results. The detail is better preserved in the smaller-scaled regions, while in the larger-scaled regions, better noise removal is done.

For this brain scan, $\alpha(\vec{x})$ was constructed essentially by hand, by looking at the local scales of the image, and determining from (13) how to to weight the fitting error relative to the total variation. (Of course in general we will not have the true image to use in determining

α^{scale}_{max}, so that this example is somewhat contrived). Determining (or as least estimating) the local characteristics in the image is one of the important challenges in adaptive or multi-scale regularization. If we can estimate local scales, we can effectively apply the results developed in this section. There is the need to be able to do this automatically and efficiently.

3.5 Applying Exact Solution to Solving Constrained Problem

The analysis of the scale in TV regularization which we have developed in the previous sections can be used to solve the constrained problem (2) as well. It can be used to determine how to locally weight the TV regularization in solving a modification of the orginal problem. As already mentioned in section 1.2, the modified constrained problem is

$$\min_u \int_\Omega \alpha(\vec{x}, u)|\nabla u(\vec{x})|\, d\vec{x} \quad \text{subject to} \quad \|u - z\|^2_\Omega = \sigma^2,$$

where $\alpha(\vec{x}, u)$ is a varying parameter which determines the local balance between goodness of fit and total variation reduction. This would cause more smoothing to be done to objects of larger scale, and less to objects of smaller scale, while still preserving the condition $\|u-z\|^2_\Omega = \sigma^2$.

3.6 Summary of Application of Analytic Theory

To summarize, by developing the exact solution theory of TV regularization, we can do adaptive regularization:

- Automatically: If we can automatically estimate the scale of the local features of the image, which we use to find $\alpha^{scale}(\vec{x})$. Then where α^{noise} is the value of α based only on noise, we choose $\alpha(\vec{x}) = f(\alpha^{scale}(\vec{x}), \alpha^{noise})$.

- Interactively: If we know what scale of object we're interested in, then we can choose α accordingly. It is possible to choose α so that the correct balance between noise removal and regularization is done, so as to preserve at least a minimum amount of detail and contrast. This can be done locally if there are multiple objects in the image.

- Automatically and Interactively: Automatic regularization could first be done, with interactive regularization done to recover the desired information from the image in a particular region with a particular scale.

- Combine the theoretical results with existing adaptive schemes for TV or other geometric diffusion regularization.

References

[1] L. Alvarez and J. Morel. Formalization and Computational Aspects of Image Analysis. *Acta Numerica*, pp. 1-59, 1994.

[2] M. R. Banham and A. K. Katsaggelos. *Spatially adaptive wavelet-based multiscale image restoration.* Technical report, Dept. of E.E. and C.S., Northwestern University, 1995. Submitted to IEEE Trans. on Image Processing.

[3] J. Biedmond and R. Lagendijk. *Iterative Identification and Restoration of Images.* Kluwer, 1991.

[4] D. Dobson and F. Santosa. Recovery of Blocky Images from Noisy and Blurred Data. *SIAM J. Appl. Math.* To appear.

[5] D. Dobson and C.R. Vogel. Convergence of An Iterative Method for Total Variation Denoising. *SIAM J. Num. Analysis*, submitted.

[6] R.C. Gonzalez and R.E. Woods. *Digital Image Processing.* Addison Wesley, 1993.

[7] M. Hanke and P.C. Hansen. Regularization Methods for Large Scale Problems. *Survey of Math. in Industry*, 3:253-315, 1993.

[8] P.L. Lions, S. Osher, and L. Rudin. Denoising and Deblurring Algorithms with Constrained Nonlinear PDE's. *SIAM J. Num. Analysis*, submitted.

[9] J.M. Morel and S. Solimini. *Variational Methods for Image Segmentation.* Birkhauser, Boston, 1995.

[10] P. Perona and J. Malik. Scale Space and Edge Detection using Anisotropic Diffusion. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 12 (7): 629-639, July 1990.

[11] L. Rudin, MTV-multiscale total variation principle for a PDE-based solution to non-smooth ill-posed problem. Tech. Report, Cognitech, Inc. Talk presented at the Workshop on Math. Methods in Computer Vision, Univ. of Minnesota, Sept. 11-15, 1995.

[12] L. Rudin, S. Osher, and E. Fatemi. Nonlinear Total Variation Based Noise Removal Algorithms. *Physica D*, 60:259-268, 1992.

[13] B.M. ter Haar Romeny. *Geometry-driven Diffusion in Computer Vision.* Kluwer Academic Publishers, 1994.

[14] C.R. Vogel and M.E. Oman. *Iterative Methods for Total Variation Denoising*, SIAM Journal on Scientific Computing, to appear.

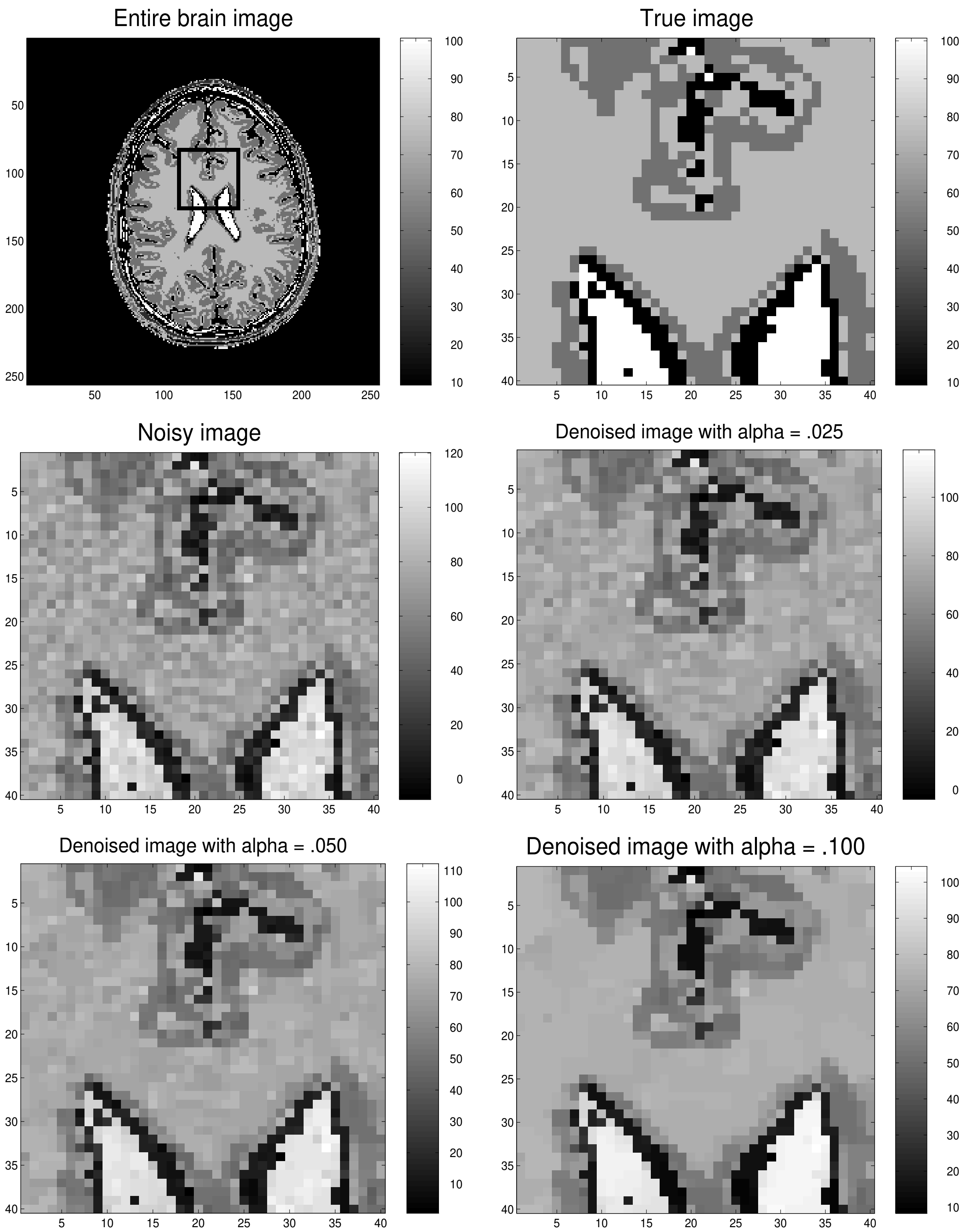

Figure 19: *The true and noisy images of a portion of a brain scan. Denoised images for three values of constant α are shown. More denoised images are shown on the next page.*

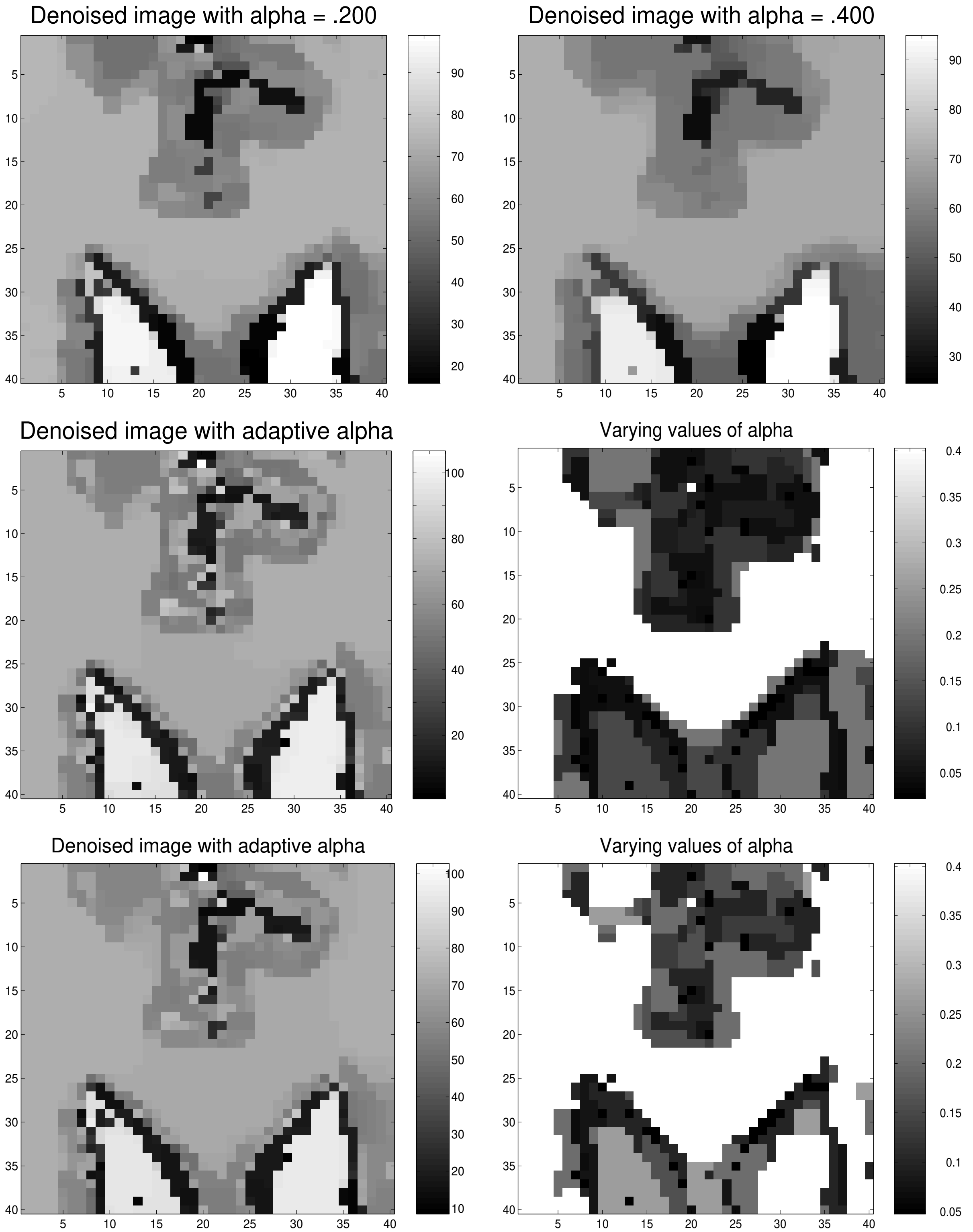

Figure 20: *Continued from the previous page. More denoised images of a portion of a brain scan using a constant value of $\alpha(x)$ are given. On the middle row we give the result of using an adaptive $\alpha(x)$, and the plot of this $\alpha(x)$. The bottom row shows the result of choosing another $\alpha(x)$, which is relatively larger than the first choice of $\alpha(x)$, and the plot of $\alpha(x)$ in this case.*