INSTITUTE OF PHYSICS PUBLISHING INVERSE PROBLEMS

Inverse Problems **19** (2003) S165–S187 PII: S0266-5611(03)60935-4

Edge-preserving and scale-dependent properties of total variation regularization

David Strong[1] **and Tony Chan**[2]

[1] Department of Mathematics, Pepperdine University, Malibu, CA 90263, USA
[2] Department of Mathematics, University of California Los Angeles, Los Angeles, CA 90095, USA

E-mail: David.Strong@pepperdine.edu and chan@math.ucla.edu

Received 14 March 2003, in final form 29 September 2003
Published 12 November 2003
Online at stacks.iop.org/IP/19/S165

Abstract

We give and prove two new and fundamental properties of total-variation-minimizing function regularization (TV regularization): edge locations of function features tend to be preserved, and under certain conditions are preserved *exactly*; intensity change experienced by individual features is inversely proportional to the scale of each feature. We give and prove exact analytic solutions to the TV regularization problem for simple but important cases. These can also be used to better understand the effects of TV regularization for more general cases. Our results explain why and how TV-minimizing image restoration can remove noise while leaving relatively intact larger-scaled image features, and thus why TV image restoration is especially effective in restoring images with larger-scaled features. Although TV regularization is a global problem, our results show that the effects of TV regularization on individual image features are often quite local. Our results give us a better understanding of what types of images and what types of image degradation are most effectively improved by TV-minimizing image restoration schemes, and they potentially lead to more intelligently designed TV-minimizing restoration schemes.

1. Introduction

Consider the problem of restoring a noise-contaminated image in R^1, R^2 or R^3: find an approximation $u(\vec{x})$ to the true image $u_{\text{true}}(\vec{x})$, given the measured (noisy) image $u_0(\vec{x})$, where $u_0 = Ku_{\text{true}} + \eta$ and $\eta(\vec{x})$ is the noise or other degradation in the image. In order to develop more precise and directly useful results, we consider only the case in which the blurring operator K is the identity (in which case, strictly speaking, the problem could be considered a

0266-5611/03/060165+23$30.00

filtering problem, rather than a regularization problem). Rudin *et al* [33] proposed to remove the noise by minimizing the total variation $TV(u)$ in the image, where

$$TV(u) \equiv \int |\nabla u(\vec{x})| \, \mathrm{d}\vec{x}, \tag{1}$$

while preserving some fit to the original (e.g. measured) data u_0. The total variation (1) of u is also often known as the bounded variation of u. There are two common formulations of this function regularization problem. The formulation which we primarily consider in this paper is the unconstrained or Tikhonov [22] problem,

$$\min_u \int \tfrac{1}{2}[u(\vec{x}) - u_0(\vec{x})]^2 + \alpha(\vec{x})|\nabla u(\vec{x})| \, \mathrm{d}\vec{x}. \tag{2}$$

When $\alpha(\vec{x}) \equiv \alpha$ is constant, (2) is simply

$$\min_u \tfrac{1}{2}\|u - u_0\|^2 + \alpha TV(u), \tag{3}$$

where $\|\cdot\|$ is the L_2 norm and, unless otherwise specified, all functions we consider are of bounded variation. Another common formulation of this problem is the noise-constrained problem

$$\min_u \int |\nabla u(\vec{x})| \, \mathrm{d}\vec{x} \qquad \text{subject to } \|u - u_0\|^2 = \sigma^2, \tag{4}$$

where the error (noise) level σ^2 is assumed to be known. As shown in [9], solving (3) is equivalent to solving (4) when $\alpha \equiv \frac{1}{\lambda}$, where λ is the Lagrange multiplier found in solving (4). All of (2)–(4) are solved assuming Neumann boundary conditions on u.

In (2) and (3), $\alpha > 0$ is the regularization parameter that determines the balance between goodness of fit to the measured data and the amount of regularization done to the measured data u_0 in order to produce u. TV regularization looks for an approximation u to the original (e.g. noisy) function u_0 which has minimal total variation, but with no particular bias toward a sharp (i.e. discontinuous) or smooth solution. The measured data u_0, as well as the regularization parameter α when solving (2) or (3) or the estimated noise level σ^2 when solving (4), determine the sharpness or smoothness of the restored function. Larger values of α result in more regularization and less goodness of fit of u to the original data u_0. We can choose $\alpha(\vec{x})$ to be spatially varying in (2) or add a weighting term to the integrand in (4) for spatially *adaptive* image restoration or function regularization (e.g. [36–38]).

As we subsequently show, TV regularization, while not discriminating against smoothness, is particular adept at recovering sharp edges in an image. We do not attempt to exhaustively compare TV regularization to edge-preserving methods, but we do point out a few papers in which one may find further discussions. Obviously much has been written about median filters (e.g. [39]). *Anisotropic diffusion* (and several variations), introduced by Perona and Malik in [31], is increasingly popular and effective [8, 23, 26, 34, 44]. (In fact, as described in [38], TV regularization can actually be viewed as a model case of anisotropic diffusion, and consequently our results are also of use in understanding anistropic diffusion.) In [27], Mammen and van de Geer discuss least-squares penalized regression estimates with total variation penalties. In [17], Davies and Kovac look at the problem of non-parametric regression with emphasis on controlling the number of local extrema, and in particular consider the run method and the taut string-wavelet method. In [42] and [43], Winkler and Liebscher discuss the Potts and L1 procedures, including their multi-scale aspects, where a hyper-parameter is varied to control the strength of penalization.

Numerical results have shown that TV regularization is quite useful in image restoration (see e.g. [33, 36, 41]). However, while a relatively large amount of effort

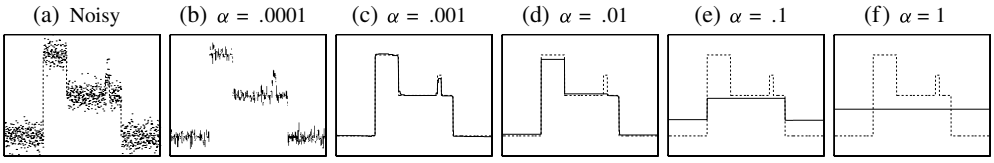

Figure 1. A comparison of results when using different values of α, when solving (3), in applying TV regularization to a noisy function in R^1. In each plot, the dashed line is the true function. In (a), the dotted line is the noisy function and in (b)–(f) the solid line is the regularized function. Larger values of α result in more regularization being applied to the noisy function. An unintelligent choice of α can result in insufficient noise removal or excessive loss of smaller-scaled features and other detail.

has been put into developing faster numerical schemes for solving the TV regularization problem [10, 12, 13, 25, 28, 33, 41], and to a lesser extent into the existence and uniqueness of solutions to (2)–(4) [1, 9, 19, 20], a relatively limited amount of study has been devoted to understanding more precisely how TV regularization affects images or other functions.

A more precise understanding of TV regularization is useful and is even necessary to provide a stronger mathematical justification for using TV minimization in image processing or other function regularization. A limited understanding of TV regularization can lead to unclear and even disastrous results. For example, figure 1 illustrates the problems that can occur when applying TV regularization to a function without understanding how the regularization affects features of varying scales or how the regularization depends on the value of α (we also see that for appropriate values of α, TV regularization is quite effective). Moreover, if we understand more precisely how TV regularization affects an image, then we can develop schemes that better exploit the properties of TV regularization that make it effective in image processing. Such schemes include TV regularization for vector-valued (e.g. colour) images [6], TV regularization for deblurring an image when the type of blurring experienced by the image is unknown [16] and new applications of the TV model for restoring nonflat image features, such as optical flows and chromaticity [15, 30, 40]. Additionally, the results which we subsequently develop in this paper can be used to construct spatially adaptive TV-minimizing image restoration schemes [36–38], as previously mentioned.

It is due to the nonlinearity of problems (2)–(4) that it is possible for TV regularization to effectively recover sharp edges in an image while simultaneously not penalizing smooth image features. Unfortunately, it is also because of this nonlinearity that it is extremely difficult (if not impossible) in the general case—and certainly non-trivial even in simple cases—to develop meaningful analytic theory which describes the effects of TV regularization in a simple, exact and useful way.

There is a relatively little literature which describes exactly why and how TV regularization works. Moreover, the work that has been done, while of some use, is also somewhat abstract and theoretical. In [7], Bellettini *et al* discuss conservation of shape and scale parameters controlling evolution within the TV-flow context. A nice study of TV regularization is made by Chambolle and Lions in [9]. In [18], Dobson and Santosa consider how TV regularization affects frequency distributions to show that the effectiveness of TV-minimizing image restoration depends on the 'mass' (essentially the size and grey-scale intensity) of the image feature relative to its total variation, and conclude that TV regularization is particularly effective in restoring 'blocky' images. In [21], Gousseau and Morel discuss why TV-minimizing techniques are effective for blocky images, but not as effective for detailed or textured images. In [4], Andreu *et al* prove that the solution to (3) for sufficiently large α

is simply the average of the initial data. These results are helpful; still, the effects of TV regularization should be understood more completely and more exactly.

In this paper we develop a precise and simple quantitative theory which describes the effects of TV regularization in some specific and important simple cases. Our theory also gives us a more thorough and precise qualitative understanding of how TV regularization affects images or other functions in the general case. We give and prove analytic formulae which describe the effects of applying TV regularization to a function (e.g. a noisy image), and which are useful for both understanding and predicting the effects of TV regularization.

Our theory shows the following:

(1) TV regularization tends to preserve edge locations, and under certain conditions it preserves edge locations *exactly*;
(2) intensity change is *exactly* inversely proportional to local feature scale (which helps to explain why TV image restoration, such as noise removal, can remove smaller-scaled noise, while leaving larger-scaled features essentially intact), is independent of original intensity and is directly proportional to the regularization parameter α;
(3) for smooth radially symmetric function features, function intensity change is inversely proportional to radial distance and directly proportional to α and
(4) TV regularization is somewhat local in its effects on image features; that is, under certain conditions, the effects of TV regularization on one feature in the image have little or no correlation with the effects of TV regularization on another feature.

This last result not only gives us a better understanding of how TV regularization affects an image, but it is potentially useful in developing faster numerical schemes (e.g. domain decomposition), which is important because of the relatively high computational costs in solving (2)–(4).

Our results are given for the continuous TV regularization problem (2) rather than the discrete version of the problem. Treating the continuous problem is often done for at least two reasons: first, the solution to the discrete problem will depend on how the problem is discretized, and second, the numerical problem is sometimes solved using some non-standard numerical scheme (for example, a scheme which tends to artificially preserve edges). In other words, the continuous problem is the most general case: it is the starting point regardless of how the problem is discretized or what numerical scheme is used to compute the solution. Our analytic results for the continuous problem agree with (in fact, they predict) the results of numerically solving the discrete problem.

We outline the remainder of our paper. In section 2 we give and prove formulae which describe precisely the basic effects of TV regularization when dealing with piecewise constant functions, and we examine some special cases of these formulae. The two *fundamental properties* of TV regularization are proved by these formulae:

(1) *edge* location (e.g. the boundary of a feature in an image) is preserved exactly, and
(2) the change in intensity experienced by individual image features due to regularization is inversely proportional to the *scale* of each feature.

In section 3 we discuss this second fundamental property and some of its implications. In section 4 we extend our formulae to the simplest case of smooth functions. A few important conclusions and a discussion of future work motivated by our results are given in section 5.

2. The two fundamental properties of TV regularization

We first analyse the effects of TV regularization on noise-contaminated radially symmetric piecewise constant functions. We do this because image features are often partially or entirely

piecewise constant, and because for the radially symmetric piecewise constant case we can find exact results that are impossible to derive in the general case. Our results can then be extended to more general cases. We consider the unconstrained problem (2). This makes the effects of TV regularization more obvious (than if we were to solve (4)), since in (2) the regularization is directly dependent on the regularization parameter α. Our results, of course, apply to (4), in which the regularization parameter is implicit in the form of the Lagrange multiplier.

In section 2.1, we give the general formulae which describe the effects of TV regularization on a radially symmetric piecewise constant function. The key results of this section are the two fundamental properties of TV regularization, that edge location is preserved exactly and that the change in function intensity is exactly inversely proportional to the scale of individual image features. Prior to the theoretical results we give below, the exact preservation of edge location was seemingly evident in numerical results and suspected by many, but, in fact, was still doubted by some. The results of 2.1 are consequently quite significant to the basic understanding of TV regularization. In section 2.2, we examine our theoretical results for simple functions in R^1, R^2 and R^3, with constant α, and for each of these cases we give numerical examples which illustrate and confirm our analytic results.

2.1. Formulae and proofs

With radially symmetry, it is not difficult to transform the minimization problem (2) in R^d, $1 \leqslant d \leqslant 3$, to a one-dimensional problem,

$$\min_u \int_r \tfrac{1}{2}[u(r) - u_0(r)]^2 + \alpha(r)|u_r(r)|\,\mathrm{d}\Omega(r) \tag{5}$$

where

$$\mathrm{d}\Omega(r) = \beta_d r^{d-1}\,\mathrm{d}r \qquad \text{and} \qquad \beta_d = \begin{cases} 1 & \text{for } d = 1 \\ 2\pi & \text{for } d = 2 \\ 4\pi & \text{for } d = 3. \end{cases} \tag{6}$$

One consequence of assuming that u_0 is radially symmetric (where $u_0 = u_{\text{true}} + \text{noise}$, and u_{true} is radially symmetric) is that we are also assuming that the noise present in the image is radially symmetric. Of course, for most problems noise is not radially symmetric. We make our assumption in order to make our mathematical analysis and results possible. Ultimately (and quite necessarily), numerical results have consistently shown that our formulae describe the results in the usual case of non-radially symmetric noise. It is reasonable to take α in (2) to be radially symmetric if u_0 is, and it is reasonable to assume that u is radially symmetric in solving (2) if both u_0 and α are (which, in fact, can be proved). We also note that although the analytic results we give are proved for the continuous problem when assuming radial symmetry, our analytic results predict quite well the results of solving the discrete problem (e.g. for digital images) with no radial symmetry. Examples of this are given in section 3.2.

Radially symmetric piecewise constant functions are comprised of three types of features: 'extrema', 'steps', and 'boundary' regions, as illustrated in figure 2. The functions in figure 2 are simply the R^1 functions that correspond to their radially symmetric counterparts in R^d. The formulae subsequently given in theorems 1–3 describe the effects of TV regularization on each of these types of feature as well as on the noise. Theorem 1 describes the effects of TV regularization when $u_0(r)$ is a noisy monotonic (i.e. monotonic before noise is added) step function, as illustrated in figure 2(a). Using theorem 1 in the case where $n = 2$, we prove theorem 2, which describes the effects of TV regularization when $u_0(r)$ is a hat function (before noise is added), that is, when $u_0(r)$ has a single (piecewise constant) extremum as illustrated in figure 2(b). Theorem 3, which is illustrated in figure 2(c), summarizes theorems 1 and 2.

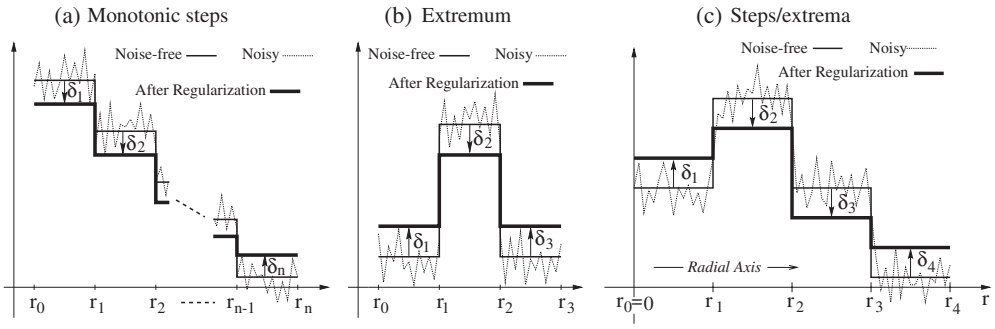

Figure 2. Noisy piecewise constant functions, before and after TV regularization, which results in piecewise constant functions (noise is removed) with exactly preserved edge location and reduced contrast. Intensity change is inversely proportional to the scale of the image features. The R^1 functions shown correspond to radially symmetric R^d functions.

We give a few words on our notation: Ω_r is the circle in R^2 or the sphere in R^3 with radius r, and $\partial\Omega_r$ is the boundary of Ω_r. Then where $r_1 \leqslant r_2$, let $\Omega_{r_1,r_2} = \Omega_{r_2} - \Omega_{r_1}$; that is, Ω_{r_1,r_2} is the region enclosed by the two boundaries $\partial\Omega_{r_1}$ and $\partial\Omega_{r_2}$. Then $|\partial\Omega_r| = \beta_d r^{d-1}$ and $|\Omega_{r_1,r_2}| = \frac{\beta_d}{d}(r_2^d - r_1^d)$, where β_d is given by (6).

Theorem 1 (Noisy monotonic step function). *For R^d, $1 \leqslant d \leqslant 3$, let $u_0(r)$ be defined on $[r_0, r_n]$. For $1 \leqslant i \leqslant n-1$ suppose that $U_i \geqslant U_{i+1}$, where*

$$U_i \equiv \frac{\int_{r_{i-1}}^{r_i} u_0(r)\,\mathrm{d}\Omega(r)}{\int_{r_{i-1}}^{r_i} \mathrm{d}\Omega(r)} \qquad \text{for } 1 \leqslant i \leqslant n. \tag{7}$$

If

$$\begin{aligned} &(i) \qquad U_i + \delta_i \geqslant U_{i+1} + \delta_{i+1} \qquad \text{for } 1 \leqslant i \leqslant n-1 \\ &(\ddot{u}) \qquad |\delta_i| \geqslant \max_{r_{i-1} \leqslant r \leqslant r_i} |u_0(r) - U_i| \qquad \text{for } 1 \leqslant i \leqslant n \end{aligned} \tag{8}$$

and

$$\alpha(r) \geqslant \alpha_{\max} \equiv \max_{1 \leqslant i \leqslant n-1} \{\alpha_i\} \qquad \text{for } r \notin \{r_i\}_{1 \leqslant i \leqslant n-1} \tag{9}$$

where $\alpha_i = \alpha(r_i)$ for $1 \leqslant i \leqslant n-1$, then the solution to (5) is given by

$$u(r) = U_i + \delta_i \qquad \text{for } r \in [r_{i-1}, r_i], \quad 1 \leqslant i \leqslant n \tag{10}$$

where

$$\delta_i = \begin{cases} \dfrac{-\alpha_i |\partial\Omega_{r_i}|}{|\Omega_{r_{i-1},r_i}|} & \text{for } i = 1 \\ \dfrac{\alpha_{i-1}|\partial\Omega_{r_{i-1}}| - \alpha_i |\partial\Omega_{r_i}|}{|\Omega_{r_{i-1},r_i}|} & \text{for } 2 \leqslant i \leqslant n-1 \\ \dfrac{\alpha_{i-1}|\partial\Omega_{r_{i-1}}|}{|\Omega_{r_{i-1},r_i}|} & \text{for } i = n. \end{cases} \tag{11}$$

Because of its length, the proof for theorem 1 is given in the appendix.

To make theorem 1 easier to understand and appreciate, we make a few notes. First, U_i in (7) is simply the mean of u_0 in the region Ω_{r_{i-1},r_i}. A simple case (but not the only case) of this is if $u_{\text{true}}(r)$ is a monotonically decreasing step function $u_{\text{true}}(r)$ whose discontinuities

are at $\{r_i\}$, and whose value in region $[r_{i-1}, r_i]$ is U_i. If noise, whose mean in each region Ω_{r_{i-1},r_i} is zero, is added to create $u_0 = u_{\text{true}} + \text{noise}$, then the mean of u_0 over Ω_{r_{i-1},r_i} is U_i. (Results are only slightly affected if the mean of the noise over a particular region is not exactly zero.) Under the conditions given in the theorem, the basic result of TV regularization is then a monotonically decreasing step function $u(r)$ whose discontinuities are also at $\{r_i\}$ and whose contrasts (i.e. size of the discontinuities at $\{r_i\}$) are less than they were in u_{true}. This is illustrated in figure 2(a). In theorem 1, the regularized function $u(r)$ is given by (10) and the changes in intensity (which reduce the contrast in the image) are given by (11). Condition (9) simply quantifies the condition that the values of spatially varying $\alpha(r)$ away from the discontinuities $\{r_i\}$ are at least as large as the values of $\alpha(r)$ at the discontinuities $\{r_i\}$ (it does not matter what the values of $\alpha(r)$ are at $r = r_0$ and r_n, since we have Neumann boundary conditions). A typical example of this is when $\alpha(r) \equiv \alpha$ is constant.

We next make a few observations. First, notice that if we take $|\partial\Omega_{r_0}|$ and $|\partial\Omega_{r_n}|$ to be zero (which it turns out is equivalent to solving (5) using Neumann boundary conditions), then all three cases of (11) can be written as

$$\delta_i = \frac{\alpha_{i-1}|\partial\Omega_{r_{i-1}}| - \alpha_i|\partial\Omega_{r_i}|}{|\Omega_{r_{i-1},r_i}|} \qquad \text{for } 1 \leqslant i \leqslant n.$$

Second, notice that with the regularized image u as defined in (10), (7) and (11),

$$\int_{r_0}^{r_n} u(r)\,\mathrm{d}\Omega(r) = \int_{r_0}^{r_n} u_0(r)\,\mathrm{d}\Omega(r). \tag{12}$$

This conservation of mass occurs while simultaneously $TV(u) < TV(u_0)$. For α sufficiently large (but finite), the regularized image $u(r)$ is constant and is simply the mean of the measured image $u_0(r)$ over the entire domain. This interesting result is discussed in more detail and from a different perspective in [4]. Observe this phenomenon in figure 1(f). Third, we note that theorem 1 is stated and proved for monotonically *decreasing* step functions (that is, before noise is added the function u_{true} is a monotonic step function); however, the results and corresponding proof are analogous for monotonically *increasing* step functions, which we consequently assume and later use without proof. Fourth, as images become more complicated, the formulae we find for δ are for the amount of change, but not necessarily the direction (i.e. the sign) of the change. In general, the direction of change will be such as to reduce the contrast in the image between neighbouring features. Finally, we note that the only condition placed on the original image u_0 is found in (8)(ii), and the conditions indirectly related to (8)(ii).

Condition (8) in theorem 1 ensures what is subsequently referred to as the *α-condition*, when applying TV regularization to noisy (as well as noise-free) piecewise constant functions. The α-condition is described *qualitatively*, and if desired can be defined *quantitatively* for any given noisy function $u(r)$.

Definition 1 (α-condition). *The regularization parameter $\alpha(\vec{x})$ meets the α-condition for a noise-contaminated piecewise constant function if*

(i) $\alpha(\vec{x})$ is sufficiently small that all discontinuities in u_{true} are present in u and
(ii) $\alpha(\vec{x})$ is sufficiently large that the noise is completely removed, resulting in a regularized function that is piecewise constant with reduced contrast, and that has discontinuities exactly where they were in u_{true}.

This *α-condition* is generally an intuitive guide rather than a strict bound on α. In fact, often there are no theoretical bounds on α that would ensure that both stipulations of the α-condition are satisfied. Still, when in practice there happens to be no theoretical region for α which ensures the two conditions of the α-condition, the results predicted by the formulae

in theorems 1–3 are still approximately or even exactly the actual results of TV regularization. This is the case for the subsequent examples illustrated in figures 3(d)–(f).

Theorem 2 is a natural consequence of applying theorem 1 in the case that $n = 2$ to a noise-contaminated radially symmetric piecewise constant function with a single extremum, as illustrated in figure 2(b). Theorem 1 in the case $n = 2$ is a step function with a single *negative* step. The result for the case of a step function with a single *positive* step is analogous and we use without proof this analogous result in proving theorem 2.

Theorem 2 (Hat function: single extremum). *For R^d, $1 \leqslant d \leqslant 3$, let (7) and (9) hold for $n = 3$. If $\max_{r\in[r_{i-1},r_i]} u_0(r) \leqslant U_i + \delta_i \leqslant U_2 + \delta_2 \leqslant \min_{r\in[r_1,r_2]} u_0(r)$ for $i = 1, 3$ (which ensures that the α-condition is met), then the solution to (5) is given by*

$$u(r) = U_i + \delta_i \qquad \text{for } r \in [r_{i-1}, r_i], \quad i = 1, 2 \tag{13}$$

where

$$\delta_i = \begin{cases} \dfrac{\alpha_1|\partial\Omega_{r_1}|}{|\Omega_{r_0,r_1}|} & \text{for } i = 1 \\ -\dfrac{\alpha_1|\partial\Omega_{r_1}| + \alpha_2|\partial\Omega_{r_2}|}{|\Omega_{r_1,r_2}|} & \text{for } i = 2 \\ \dfrac{\alpha_2|\partial\Omega_{r_2}|}{|\Omega_{r_2,r_3}|} & \text{for } i = 3. \end{cases}$$

Proof. Choose $\tilde{r}$ such that $\tilde{r}^d = (\alpha_1 r_1^{d-1} r_2^d + \alpha_2 r_2^{d-1} r_1^d)/(\alpha_1 r_1^{d-1} + \alpha_2 r_2^{d-1})$ (it not difficult to show that $r_1^d < \tilde{r}^d < r_2^d$ and thus that $r_1 < \tilde{r} < r_2$) so that

$$\frac{\alpha_1|\partial\Omega_{r_1}| + \alpha_2|\partial\Omega_{r_2}|}{|\Omega_{r_1,r_2}|} = \frac{\alpha_1|\partial\Omega_{r_1}|}{|\Omega_{r_1,\tilde{r}}|} = \frac{\alpha_2|\partial\Omega_{r_2}|}{|\Omega_{\tilde{r},r_2}|}. \tag{14}$$

Define

$$f(v) = \int_{r_0}^{r_3} \tfrac{1}{2}[v(r) - u_0(r)]^2 + \alpha(r)\,|v_r(r)|\,\mathrm{d}\Omega(r). \tag{15}$$

Let

$$f_1(v) = \int_{[r_0,\tilde{r})} \tfrac{1}{2}[v(r) - u_0(r)]^2 + \alpha(r)|v_r(r)|\,\mathrm{d}\Omega(r),$$

$$f_2(v) = \int_{(\tilde{r},r_3]} \tfrac{1}{2}[v(r) - u_0(r)]^2 + \alpha(r)\,|v_r(r)|\,\mathrm{d}\Omega(r).$$

Then $f(v) = f_1(v) + f_2(v) + \int_{\{\tilde{r}\}} \alpha(r)|v_r(r)|\,\mathrm{d}\Omega(r)$, so that

$$\begin{aligned} \min_v f(v) &= \min_v \left\{ f_1(v) + f_2(v) + \int_{\{\tilde{r}\}} \alpha(r)|v_r(r)|\,\mathrm{d}\Omega(r) \right\} \\ &\geqslant \min_v \{f_1(v) + f_2(v)\} \qquad (\text{since } \alpha(r) \geqslant 0) \\ &\geqslant \min_v f_1(v) + \min_v f_2(v). \end{aligned}$$

Using theorem 1 when $n = 2$ and using (14), we find that f_1 and f_2 are each minimized by u as defined in (13). Since this u is continuous at $r = \tilde{r}$, then f in (15) is minimized by this u. □

Theorems 1 and 2 are combined and slightly simplified (by taking $r_0 = 0$) in theorem 3, which is illustrated in figure 2(c).

Theorem 3 (General piecewise constant function). *Given the conditions of theorems 1 and 2, the unique argument u to (5) is as given in those theorems, where the change in function intensity (not necessarily including its sign) is given by*

$$\delta_i = \begin{cases} \dfrac{\alpha_{i-1}|\partial\Omega_{r_{i-1}}| + \alpha_i|\partial\Omega_{r_i}|}{|\Omega_{r_{i-1},r_i}|} & \textit{extremum regions} \\ \dfrac{\alpha_{i-1}|\partial\Omega_{r_{i-1}}| - \alpha_i|\partial\Omega_{r_i}|}{|\Omega_{r_{i-1},r_i}|} & \textit{step regions} \\ \dfrac{\alpha_{i-1}|\partial\Omega_{r_{i-1}}|}{|\Omega_{r_{i-1},r_i}|} & \textit{boundary region}. \end{cases} \tag{16}$$

Proof. This theorem can be proved for any noise-contaminated radially symmetric piecewise constant function in R^d by dividing the function into its three types of component (extremum, step and boundary regions), similar to what was done in proving theorem 2. □

The changes in intensity $\{\delta_i\}$ depend only on the value of $\alpha(r)$ at $\{r_i\}$ (the discontinuities of u_{true}), as long as $\alpha(r) \geqslant \max_{r\in\{r_i\}} \alpha(r)$. We take advantage of this fact in constructing spatially adaptive image restoration schemes in [36, 37]. Also, our results show that the regularized image is the same when restoring a noise-free image and when restoring the noisy version of that image, for sufficiently large values of α, if the mean of the noise in each region of the image is zero. This is subsequently illustrated in figure 3. We note that the sign of the change in intensity δ_i in each region will in general be such as to reduce contrast in the image.

2.2. Theorem 3 with constant α

Let the conditions of theorem 3 be satisfied, and let $\alpha(x) \equiv \alpha$ be constant. Then (16) is given by

$$\delta_i = \begin{array}{ccccc} R^1 & R^2 & R^3 & \text{Type of region} \\ \hline \dfrac{2}{x_i - x_{i-1}}\alpha & \dfrac{2(r_i + r_{i-1})}{r_i^2 - r_{i-1}^2}\alpha & \dfrac{3(r_i^2 + r_{i-1}^2)}{r_i^3 - r_{i-1}^3}\alpha & \text{Extremum} \\ \dfrac{0}{x_i - x_{i-1}}\alpha \equiv 0 & \dfrac{2(r_i - r_{i-1})}{r_i^2 - r_{i-1}^2}\alpha & \dfrac{3(r_i^2 - r_{i-1}^2)}{r_i^3 - r_{i-1}^3}\alpha & \text{Step} \\ \dfrac{1}{x_i - x_{i-1}}\alpha & \dfrac{2r_{i-1}}{r_i^2 - r_{i-1}^2}\alpha & \dfrac{3r_{i-1}^2}{r_i^3 - r_{i-1}^3}\alpha & \text{Boundary} \end{array} \tag{17}$$

Example 1. In R^1, R^2 and R^3, we take the original image (before noise is added) to be

$$u_0(r) = \begin{cases} 0.5 & \text{for } 0.00 \leqslant r < 0.25 \\ 1.0 & \text{for } 0.25 \leqslant r < 0.50 \\ 0.5 & \text{for } 0.50 \leqslant r < 0.75 \\ 0.0 & \text{for } 0.75 \leqslant r \leqslant 1.00. \end{cases}$$

In the top row of images (a)–(c) in figure 3 we give the numerical results in applying TV regularization to $u_0 = u_{\text{true}}$ (i.e. without noise), solving (5) using $\alpha = 0.01$. We also apply TV regularization to $u_0 = u_{\text{true}} + \eta$, a noise-added version of this image, and obtain virtually identical results, as shown in the second row of images (d)–(f). For all six cases, the numerical results match the results predicted by (17). The juxtaposition of the regularized noisy images and the regularized noise-free images helps to clarify how TV regularization affects both the noise in the image and the image itself.

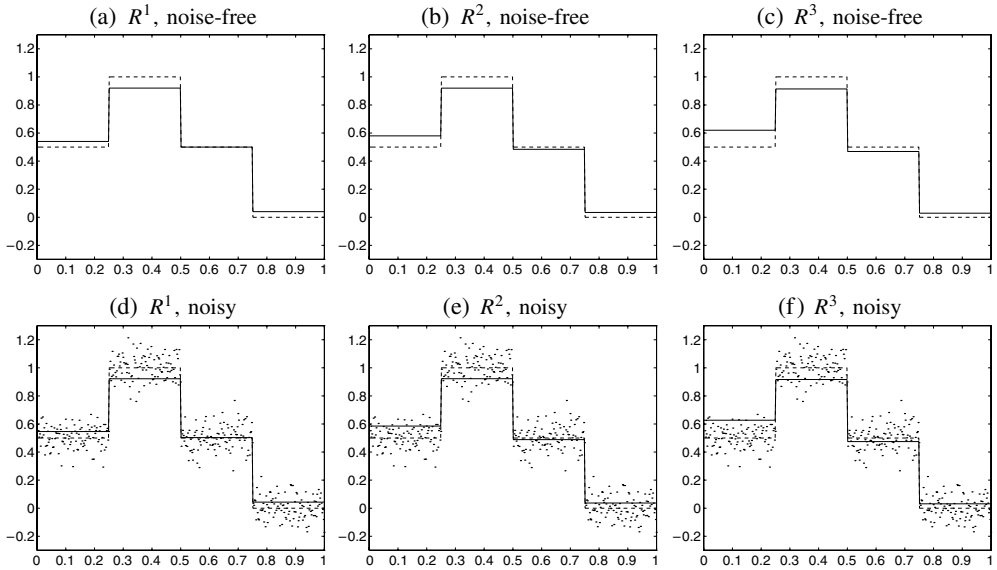

Figure 3. Example 1. Results of solving (5) using $\alpha = 0.01$. (a) and (d) are R^1 functions, and (b)–(e) are the R^1 functions corresponding to radially symmetric R^2 and R^3 functions, noise-free and noisy, before and after TV regularization. The dashed line is the noise-free function, the solid line is the regularized function and in (d)–(f) the dotted line is the noisy function. In each case the change in intensity level is as predicted by (17). Notice the equivalence between the regularized functions in the noise-free case and the noisy case.

For this example theoretically there is no region of values for α in which both conditions of the α-condition could be met; this illustrates that in practice the theoretical bounds on α that would be required in order to ensure the α-condition do not necessarily need to be met in order to approximately or even exactly obtain the results predicted by (17), or more generally (16).

Finally, we note that, as we have formulae which describe exactly how much the intensity changes in each region, theoretically it is possible to perfectly (or approximately) reconstruct the true image, by adding back to the regularized image the loss in intensity caused by regularization. This turns out to be more difficult that one might first suspect, at least for more complicated (e.g. more realistic) images. We are currently studying the possibilities of this near perfect reconstruction given our precise understanding of how TV regularization changes intensity levels.

3. Change in image intensity as a function of image feature scale

One of the most important and useful results of section 2 is the fact that the change in image intensity due to TV regularization is directly proportional to the regularization parameter α and inversely proportional to the scale of the image feature, as we now describe in section 3.1. This explains why noise (which can often be thought of as very small-scaled image features) is removed while the image features (whose scales tend to be large relative to the noise) are preserved. In section 3.2 we apply this relationship to three discrete examples in R^2, which are images with non-radially symmetric features, and we give numerical results. In section 3.3 we give a few simple applications of this interesting property of TV regularization. In section 3.4 we briefly discuss how our results demonstrate the somewhat local nature of TV regularization.

3.1. Change in image intensity $\delta = \alpha/scale$

In the case where α is constant, note for the formula given in (16) for extremum regions that the change in intensity simplifies to $\delta_i = (|\partial\Omega_{r_{i-1},r_i}|/|\Omega_{r_{i-1},r_i}|)\alpha$ where $|\partial\Omega_{r_{i-1},r_i}| = |\partial\Omega_{r_{i-1}}| + |\partial\Omega_{r_i}|$. If we let Ω represent a radially symmetric constant region, then this can be written $\delta = (|\partial\Omega|/|\Omega|)\alpha$, where $|\partial\Omega|$ is the size of the entire (inner + outer) boundary of the feature. We define the scale of a piecewise constant image feature as the ratio of the area in R^2 (volume in R^3) of the feature to its boundary length (surface area in R^3); that is, scale $= |\Omega|/|\partial\Omega|$. (Note that this definition also applies to non-symmetric features.) For example, in R^2 a circle of radius r would have scale $= \frac{\pi r^2}{2\pi r} = \frac{r}{2}$, so that its scale is linearly proportional to r.

With this definition of scale, the change in intensity δ can be rewritten as

$$\delta = \frac{\alpha}{\text{scale}}, \tag{18}$$

so that *the change in image intensity*, δ, *is inversely proportional to scale and directly proportional to* α. This is perhaps the most basic and important property of TV regularization and how it affects an image: *TV regularization causes smaller-scaled features (such as noise) to be partially or entirely removed while larger-scaled image features are relatively unaffected*. In both [18] and [21] it is concluded that TV image regularization is well suited for denoising images with large-scale features; the above results now explain this more precisely and in a perhaps more exploitable way.

3.2. Numerical results for discrete non-radially symmetric image

So far we have considered some special cases of the *continuous* TV regularization problem, namely radially symmetric piecewise constant functions. We have been able to find *analytic* solutions to (2) for these cases. In the general case, however, we are not able to explicitly find analytic solutions to (2). Also, in practice we must of course *numerically* solve the *discrete* version of the (usually non-radially symmetric) continuous problem, for example, when denoising a digital image. We now briefly show the agreement between the theory and the numerical solutions to the discrete version of (2) for three images, two of which are not radially symmetric. We limit ourselves to the three images shown in figure 4 in order to keep the discussion relatively simple. Our primary purpose in this paper has been to develop a simple and precise theory for the special cases which we have treated, which may be used to understand how TV regularization affects the various types of images dealt with in the general case. Our purpose is not to attempt to exhaustively examine the exact effects of TV regularization in the general case.

Example 2. We briefly examine the agreement between the change in intensity in the feature of interest and the change predicted by (18) for three images. Before noise is added the region of interest (the shape) has unit intensity and the background has zero intensity. The intensities in the image after noise is added and after regularization occurs can be seen in each image. The feature of interest Ω in the first image, shown in figure 4(a), is a circle of radius $\frac{1}{3}$ in a 1×1 square domain, discretized at 128×128. We note that the finer the discretization, the more similar to the continuous problem the discretized problem would be, and thus the better the agreement between the theoretically predicted results for the continous problem and the actual numerical results for the discretized problem.

To each image we add Gaussian noise, a different amount to each image for variety (the particular level of noise is not terribly important for our purposes). Of course, in general, more

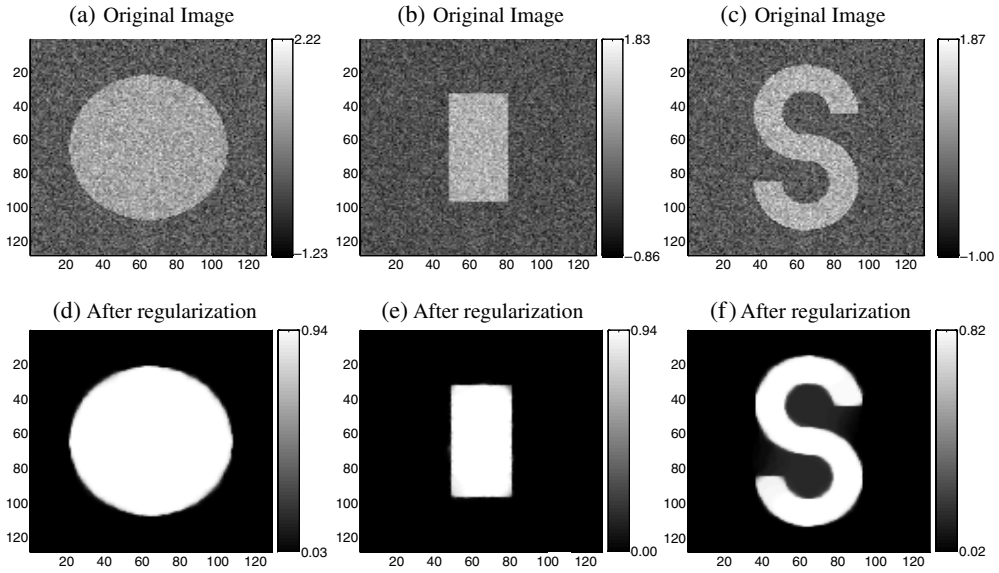

Figure 4. Example 2. Refer to table 1. The results of applying TV regularization to general piecewise constant functions. For each image feature, noise is removed, edge location is preserved and the change in intensity level in the image feature and the background is approximately as predicted.

noise means that more regularization is necessary, which in this case means a larger value of α. For the first image we use $\alpha = 0.01$, so that in the circular region the change in intensity $\delta_{\text{predicted}}$ as predicted by (18) should be $\delta_{\text{predicted}} = \frac{\alpha}{\text{scale}} = \frac{|\partial\Omega|}{|\Omega|}\alpha = [2\pi(\frac{1}{3})/\pi(\frac{1}{3})^2](0.01) = 0.06$, so that in the circular region the intensity level should be 0.94 after regularization. This is nearly exactly the case, as shown in figure 4(d).

Notice that the boundary of the circle has been slightly smoothed. This is because of the noise and because the discretized circle is not perfectly radially symmetric. If desired this slight boundary smoothing in the discrete case can be mitigated or entirely removed by employing an edge-preserving numerical scheme for solving (2)–(4), such as the *minmod* scheme used in [33]. Because of the effectiveness and the popularity of edge-preserving numerical schemes, we are currently investigating whether it is possible to find analytic formulae which describe the effects of TV regularization when using such schemes.

We can also find the predicted change in the background of the image as well. Since Neumann boundary conditions are inherent in solving the TV regularization problem, the only boundary of the background is the boundary between the background and the circle, in which case we have $\delta_{\text{predicted}} = \frac{\alpha}{\text{scale}} = \frac{|\partial\Omega|}{|\Omega|}\alpha = [2\pi(\frac{1}{3})/(1 - \pi(\frac{1}{3})^2)](0.01) \approx 0.03$, so that the intensity of the background should be approximately 0.03 after regularization, which is the case.

We also apply TV regularization to two images whose features in the continuous case are not radially symmetric: a rectangle and an 'S'. We note that in the non-radially symmetric case, TV-minimizing techniques tend to smooth out rough or non-radially symmetric boundaries [18, 36], and can consequently result in the deformation of boundaries in the function (unless an edge-preserving numerical scheme is used). This boundary deformation occurs because the total variation of a feature is directly proportional to its boundary size, e.g. for a piecewise constant feature [35], so that one way of minimizing the total variation of

Table 1. Example 2. Refer to figure 4. A comparison of predicted changes in intensity to the results found numerically in applying TV regularization to discrete images. The numerical results nearly exactly match the predicted results, with the small discrepancy being due to the noise not having exactly zero mean and due to the slight deformation of the boundaries in the images with non-radially symmetric features.

		Shape		Background	
Image	α	Predicted δ	Computed δ	Predicted δ	Computed δ
Circle	0.010	$=0.06$	≈ 0.06	≈ 0.03	≈ 0.03
Rectangle	0.005	$=0.06$	≈ 0.06	≈ 0.01	≈ 0.00
'S'	0.010	≈ 0.18	≈ 0.18	≈ 0.03	≈ 0.02

the feature would be to reduce its boundary size, in particular by smoothing corners. Because this boundary deformation occurs in the non-radially symmetric case, the results given in section 2.1 (in particular, that there is no boundary deformation in the radially symmetric case) are even more significant.

For each image, we give the value predicted by (18) for the change in intensity, and find the change of intensity when solving the problem numerically. Since there is a bit of smoothing near the edges, we take the value away from the edges (the image is essentially constant away from the edges) as the computed value. The results are given in figure 4 and table 1. The slight discrepancies between predicted and computed changes in intensity level are due mostly to the noise not having exactly zero mean (this is especially the case for the background), as well as the boundary deformation (not a change or shift in boundary location, but the boundary smoothing—this distinction is significant) in the two images with non-radially-symmetric features. Overall, the agreement between the theory and the numerical results is nearly exact.

3.3. Adaptive TV regularization using $\delta = \alpha/scale$

There is good potential for exploiting the second fundamental property (18) of TV regularization. For example, if we can locally measure the change in intensity level due to TV regularization, then we can find the scale of various image features, by re-writing (18) as

$$\text{scale}(\vec{x}) = \frac{\alpha}{\delta(\vec{x})}. \tag{19}$$

In [36] we use (19) to construct automatic scale recognition schemes, where we first apply TV regularization to an image, then find $\delta(x)$ throughout the image, then use (19) to find scale(x). Also in [36] we use (19) to construct scale-sensitive adaptive image restoration schemes, where the basic approach is to choose smaller $\alpha(r)$ in regions with smaller-scaled features in order to better preserve detail, and larger $\alpha(r)$ in regions with larger-scaled features in order to better remove noise. Finally, it turns out that TV regularization can be used to resolve an image into its various scales. A simple illustration of this is seen in figure 1, where for larger values of α, the smaller-scaled image features (including noise) are removed. Because of our fairly precise understanding of how the value of α will determine the effect of TV regularization on features of any given scale, it becomes possible to resolve an image into its features with scale larger than any desired threshold, as is simplistically illustrated in figure 1. We are currently studying these various ideas that are the direct result of (18) and (19).

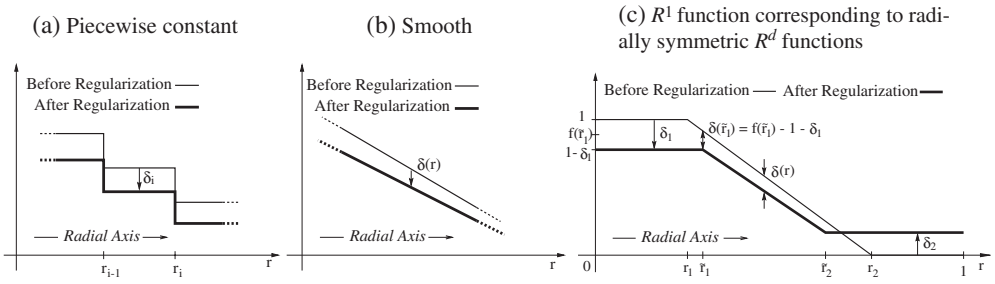

Figure 5. With radial symmetry, the effect of TV regularization in the smooth case (b) can be taken as a limit of the piecewise constant case (a). In (b) and (c) $\delta(r) = 0$ in R^1, $\delta(r) = \frac{1}{r}\alpha$ in R^2 and $\delta(r) = \frac{2}{r}\alpha$ in R^3. In (c) for R^1, formulae for finding $\tilde{r}_i$ and δ_i are given in (21) and (22); formulae for finding $\tilde{r}_i$ in R^2 and R^3 are given in (23), while δ_1 and δ_2 must be numerically approximated.

3.4. The local nature of TV regularization

Although (5), and more generally (2), are global problems, the effects of TV regularization are often quite local. Notice in (16) that the change δ_i in each region Ω_{r_{i-1},r_i} depends only on the region itself and on α. In particular, δ_i does not depend on the effects of the regularization felt in any other region. This is especially so for extremum regions, for which we found (18). For extremum regions, δ depends only on the scale of the feature and α, but again δ does not depend on the effects of the regularization felt in any other region. To be clear, TV regularization is *not* local to a particular size of subdomain (e.g. 5×5 pixels or 10×10 pixels). Rather, it is somewhat local to individual image features, whether the features are small or large. Of course the conditions of theorems 1–3 are often not exactly met in practice. Still, this local regularization characteristic of TV regularization is quite evident when applying TV regularization to general images. It is our hope that this additional understanding of the localness of TV regularization can lead to better manipulation of the effects of TV regularization, as well as to faster numerical schemes which apply regularization to individual parts of the image (e.g., domain decomposition-like schemes).

4. A note on TV regularization of smooth functions

We briefly consider how the theory developed in section 2 helps describe the effects of TV regularization on 'smooth' (i.e. non-piecewise constant) features. Our results are for functions in R^1 and for radially symmetric functions in R^2 and R^3. We consider the case where $\alpha(\vec{x}) \equiv \alpha$ is constant. We do not intend to provide a completely rigorous treatment of the smooth case here, but rather simply to provide some insight into how TV regularization affects smooth functions by using the theory we have already developed.

4.1. Smooth functions as the 'limit' of piecewise constant functions

We first consider the portion of a radially symmetric function with constant, decreasing slope, as illustrated in figure 5(b). For now we assume that no noise is present. We can take this function as a sort of limit of the step function shown in figure 5(a). In section 2 we found that if α is constant, then δ_i in figure 5(a) is given in (16) by $\delta_i = [(|\partial\Omega_{r_i}| - |\partial\Omega_{r_{i-1}}|)/|\Omega_{r_{i-1},r_i}|]\alpha$. As $r_{i-1} \longrightarrow r_i$ (and as the width of each step descreases, obviously the number of individual steps in the step function would need to increase to maintain the same basic function), we find

that $\delta(r)$ for $r \in [r_{i-1}, r_i]$ is given by

$$\delta(r) = \frac{|\partial\Omega_{r_i}| - |\partial\Omega_{r_{i-1}}|}{|\Omega_{r_{i-1},r_i}|}\alpha = \begin{cases} \dfrac{1-1}{r_i - r_{i-1}}\alpha \equiv 0 & \text{in } R^1 \\ \dfrac{2\pi r_i - 2\pi r_{i-1}}{\pi r_i^2 - \pi r_{i-1}^2}\alpha \longrightarrow \dfrac{1}{r}\alpha & \text{in } R^2 \\ \dfrac{4\pi r_i^2 - 4\pi r_{i-1}^2}{\frac{4}{3}\pi r_i^3 - \frac{4}{3}\pi r_{i-1}^3}\alpha \longrightarrow \dfrac{2}{r}\alpha & \text{in } R^3. \end{cases} \tag{20}$$

We consider the effects of TV regularization on the noise-free function shown in figure 5(c). We refer to it as 'smooth' in the sense that it is continuous, with no sharp discontinuous edges. Although this function is not actually smooth in the usual sense, the results we develop using this example are applicable to functions which are smooth in the usual sense. Moreover, the notion of smoothness is somewhat irrelevant in the discrete implementation of the problem anyway. By viewing this function as the limit of a step function, the results developed in section 2 and in (20) predict the regularized function shown in figure 5(c).

4.2. Smooth functions in R^1

In R^1, assuming that the regularization function u that is the solution to (3) is as shown in figure 5(c), (20) tells us that $\delta(r) \equiv 0$. Also, it is clear that

$$\tilde{r}_i = r_i - \frac{\delta_i}{m}, \tag{21}$$

and in solving (3) it is fairly easy to find that

$$\begin{aligned} \delta_1 &= +m|\Omega_{0,r_1}| - \sqrt{(m|\Omega_{0,r_1}|)^2 + 2m\alpha} \\ \delta_2 &= -m|\Omega_{r_2,1}| + \sqrt{(m|\Omega_{r_2,1}|)^2 + 2m\alpha}, \end{aligned} \tag{22}$$

where $m = \frac{1}{r_1 - r_2}$ (simply the slope of the non-constant part of the function). If we let $m \longrightarrow -\infty$ (which is equivalent to $r_1 \longrightarrow r_2$), we would have $\delta_1 \longrightarrow -\alpha/|\Omega_{0,r_2}|$ and $\delta_2 \longrightarrow \alpha/|\Omega_{r_2,1}|$, as found for the piecewise constant case in section 2.

Example 3. Using $r_1 = 0.25$, $m = -2$, $r_2 = 0.75$ and $\alpha = 0.03$, we find that the values predicted by (21) and (22), $\delta_1 = -0.139$, $\tilde{r}_1 = 0.320$, $\delta_2 = 0.139$ and $\tilde{r}_2 = 0.680$, agree exactly with the regularized function found by solving (3), as shown in figure 6(a). We do not give further proof to verify these results, as they are a fairly natural extension of results given in section 2 for the R^1 piecewise constant case. For comparison, we apply TV regularization to a noisy version of the function, and observe similar results in figure 6(d), except for the unwanted effect of 'staircasing'. (This staircasing effect is well known and is currently being studied [5, 9, 11, 29, 32].) These results show that, for smooth image features, TV regularization can result in a slight shifting of the boundary/edge location of the feature. On the other hand, for this type of function/image, there are not really sharp edges to begin with.

4.3. Smooth functions in R^2 and R^3

For R^2 and R^3 we again consider the function shown in figure 5(c). The R^1 functions in figure 5 can be viewed as the R^1 representation of radially symmetric R^2 or R^3 functions. Similar to the R^1 case, in R^d we can find $\tilde{r}_i$ as a function of δ_i:

$$\tilde{r}_i = \frac{mr_i + \delta_i - \sqrt{(\delta_i + mr_i)^2 + 2^d m\alpha}}{2m}. \tag{23}$$

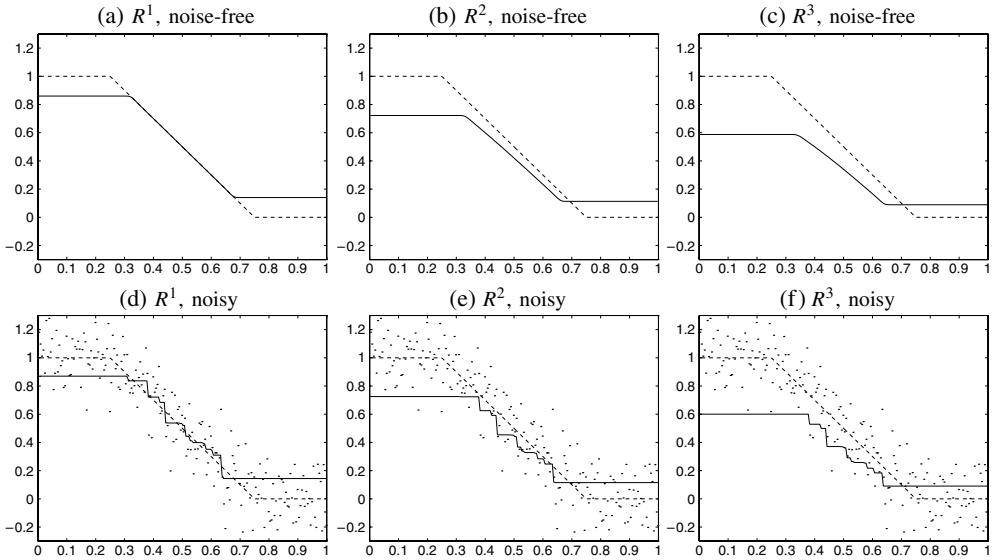

Figure 6. Examples 3–5. TV regularization of 'smooth' noise-free and noisy functions. The dashed line is the true function, the solid line is the regularized function and in (d)–(f) the dotted line is the noisy function. Exaggerated 'staircasing' in (e) and (f) is due in large part to the noise being taken as radially symmetric for these examples. In (b), (c), (e) and (f), the R^1 functions correspond to radially symmetric functions in R^2 and R^3.

However, for the R^2 and R^3 cases we cannot find explicit equations for δ_i, as we were able to do in finding (22) in R^1. The best we can do is write (5) as a function of δ_1 and δ_2 (as well as other known values)

$$\min_{\delta_1,\delta_2} \tfrac{1}{2} \int [u(\delta_1, \delta_2) - u_0]^2 + \alpha TV(u(\delta_1, \delta_2)), \tag{24}$$

and numerically approximate the minimizing values of δ_1 and δ_2. It is reasonable to assume that we can numerically find unique minimizing values for δ_1 and δ_2 since it has been shown [1, 20] that a unique minimizer to (3) exists under certain typical conditions. If our predictions about the form of u, as given in figure 5(c), are correct, then the function that results from solving (24) should be identical to the function that results from solving (5), which is the case, as demonstrated in the two examples below.

Example 4. In R^2, where $r_1 = 0.25$, $m = -2$, $r_2 = 0.75$ and $\alpha = 0.04$, we numerically approximate the values of δ_1 and δ_2 which minimize (24), use their values to compute $\tilde{r}_1$ and $\tilde{r}_2$ in (23) and find $\delta_1 = -0.278$, $\tilde{r}_1 = 0.328$, $\delta_2 = 0.113$ and $\tilde{r}_2 = 0.663$. These values agree exactly with the regularized function found by solving (5), with no assumptions made about what form u should take, as shown in figure 6(b). Results for the noisy R^2 case are shown in figure 6(e). The staircasing observed in figure 6(e) is exaggerated, since we took the noise to be symmetric (in order to solve this radially symmetric R^2 problem as an R^1 problem). In the actual R^2 problem, the noise of course is generally not radially symmetric, and the staircasing, although still present, is not nearly as pronounced as seen in the figure.

Example 5. In R^3, using the same values as used in example 4, we numerically approximate the values of δ_1 and δ_2 which minimize (24), use these values to compute $\tilde{r}_1$ and $\tilde{r}_2$ in (23) and find $\delta_1 = -0.413$, $\tilde{r}_1 = 0.338$, $\delta_2 = 0.089$ and $\tilde{r}_2 = 0.643$, which agree with the function u

found by solving (5), as shown in figure 6(c). Results for the noisy R^3 case are shown in figure 6(f). Once again, the staircasing is exaggerated, due to the radial symmetry of the noise.

5. Conclusion and future work

As was suspected by many (but, in fact, doubted by a few), TV regularization tends to preserve edges in general, and in certain cases does indeed preserve edges exactly. Consequently, TV regularization is very well suited for edge-preserving image processing, while not discriminating against smoothness in an image. Our results show that the basic effect of TV regularization is to change intensity in an image to reduce the contrast (i.e. the variation) between individual features in the image. Also, TV regularization tends to affect individual image features, and the change in intensity experienced by each feature is inversely proportional to the scale of that feature and directly proportional to the regularization parameter α.

There are various avenues to explore that are motivated by our results:

(1) developing spatially adaptive and more effective non-adaptive TV-minimizing image restoration schemes, in particular, by making use of (19);
(2) developing more localized versions of TV regularization (which would consequently be quicker and cheaper than the standard global TV regularization), taking advantage of the localness of TV regularization; and
(3) developing faster numerical schemes for solving TV regularization problems by exploiting both fundamental properties of TV regularization described in this paper.

Acknowledgments

This work was supported by the ONR under contract N00014-96-1-0277, by the NSF under grant DMS-9626755 and by Los Alamos National Laboratory under grant LDRD-DR 20030037DR.

Appendix. Proof of theorem 1

To prove theorem 1, we will need the following technical result.

Lemma 1. *Let $u(r)$, $\varepsilon(r)$ and $\alpha(r) > 0$ be defined on $[r_0, r_n]$. If $u(r)$ is a monotonically decreasing step function with discontinuities only at $S = \{r_i\}_{i=1}^{n-1}$, then (where $\mathrm{d}\Omega(r)$ is given by (6))*

$$\int_{[r_0,r_n]} \alpha(r)|u_r(r) + \varepsilon_r(r)|\,\mathrm{d}\Omega(r) \geqslant \int_{[r_0,r_n]} \alpha(r)|u_r(r)|\,\mathrm{d}\Omega(r) + \int_{[r_0,r_n]-S} \alpha(r)|\varepsilon_r(r)|\,\mathrm{d}\Omega(r) - \int_S \alpha(r)\varepsilon_r(r)\,\mathrm{d}\Omega(r).$$

Proof of lemma 1. First a note on $\int_{\{\hat{r}\}} f(r)\,\mathrm{d}\Omega(r)$. Suppose $f(r)$ is in BV. Then $\mathrm{d}f/\mathrm{d}r$ is a signed Radon measure which we denote μ_{f_r} and we define $\int_{\{\hat{r}\}} f_r(r)\,\mathrm{d}\Omega(r) \equiv \lim_{\hat{r}_-,\hat{r}_+\to\hat{r}} \mu_{f_r}([\hat{r}_-, \hat{r}_+]) = \mu_{f_r}(\{\hat{r}\})$. Likewise, for $|f_r| = |\mathrm{d}f/\mathrm{d}r|$, we get a (positive) Radon measure and the definition of $\int_{\{\hat{r}\}} |f_r(r)|\,\mathrm{d}\Omega(r) = \mu_{|f_r|}(\{\hat{r}\})$. Next, let

$$S_+ = \left\{ r_i \in S, 1 \leqslant i \leqslant n-1 : \int_{\{r_i\}} |u_r(r)| - \varepsilon_r(r)\,\mathrm{d}\Omega(r) \geqslant 0 \right\},$$

$$S_- = \left\{ r_i \in S, 1 \leqslant i \leqslant n-1 : \int_{\{r_i\}} |u_r(r)| - \varepsilon_r(r)\,\mathrm{d}\Omega(r) < 0 \right\}.$$

Then we have

$$\begin{aligned}
&\int_{[r_0,r_n]} \alpha(r)|u_r(r)+\varepsilon_r(r)|\,\mathrm{d}\Omega(r) \\
&\qquad = \int_{[r_0,r_n]-S} \alpha(r)|\varepsilon_r(r)|\,\mathrm{d}\Omega(r) + \int_S |u_r(r)+\varepsilon_r(r)|\,\mathrm{d}\Omega(r) \\
&\qquad\quad (\text{since } u_r(r)=0 \text{ in } [r_0,r_n]-S) \\
&\qquad = \int_{[r_0,r_n]-S} \alpha(r)|\varepsilon_r(r)|\,\mathrm{d}\Omega(r) + \int_{S_+} \alpha(r)[|u_r(r)|-\varepsilon_r(r)]\,\mathrm{d}\Omega(r) \\
&\qquad\quad + \int_{S_-} \alpha(r)[\varepsilon_r(r)-|u_r(r)|]\,\mathrm{d}\Omega(r) \\
&\qquad\quad \left(\text{since } \alpha(r)>0 \text{ and } \int_{\{r_i\}} u_r(r)\,\mathrm{d}\Omega(r) < 0\right) \\
&\qquad \geqslant \int_{[r_0,r_n]-S} \alpha(r)|\varepsilon_r(r)|\,\mathrm{d}\Omega(r) + \int_S \alpha(r)[|u_r(r)|-\varepsilon_r(r)]\,\mathrm{d}\Omega(r) \\
&\qquad = \int_{[r_0,r_n]-S} \alpha(r)|\varepsilon_r(r)|\,\mathrm{d}\Omega(r) + \int_{[r_0,r_n]} \alpha(r)|u_r(r)|\,\mathrm{d}\Omega(r) \\
&\qquad\quad - \int_S \alpha(r)\varepsilon_r(r)\,\mathrm{d}\Omega(r) \qquad (\text{since } u_r(r)=0 \text{ in } [r_0,r_n]-S).
\end{aligned}$$

□

Proof of theorem 1. Define

$$f(v) = \int_{r_0}^{r_n} \tfrac{1}{2}[v(r)-u_0(r)]^2 + \alpha(r)|v_r(r)|\,\mathrm{d}\Omega(r),$$

which is simply the function to minimize in (5). We show that $f(u+\varepsilon) > f(u)$ for any $\varepsilon = \varepsilon(r)$ unless $\varepsilon \equiv 0$, where u is as defined in (10). Using lemma 1 we have

$$\begin{aligned}
f(u+\varepsilon) &= \int_{r_0}^{r_n} \tfrac{1}{2}[u(r)+\varepsilon(r)-u_0(r)]^2 + \alpha(r)|u_r(r)+\varepsilon_r(r)|\,\mathrm{d}\Omega(r) \\
&\geqslant f(u) + g(u,\varepsilon)
\end{aligned}$$

where

$$\begin{aligned}
g(u,\varepsilon) = &\int_{r_0}^{r_n} \varepsilon(r)[u(r)-u_0(r)]\,\mathrm{d}\Omega(r) + \int_{[r_0,r_n]-S} \alpha(r)|\varepsilon_r(r)|\,\mathrm{d}\Omega(r) \\
&- \int_S \alpha(r)\,\varepsilon_r(r)\,\mathrm{d}\Omega(r) + \int_{r_0}^{r_n} \tfrac{1}{2}[\varepsilon(r)]^2\,\mathrm{d}\Omega(r).
\end{aligned}$$

We must show that $g(u,\varepsilon) > 0$ unless $\varepsilon \equiv 0$.

We first define

$$\tilde{\varepsilon}(r) \equiv \varepsilon(r) - \int_{S_{i-1}} \varepsilon_r(r)\,\mathrm{d}\Omega(r) \qquad \text{for } r \in [r_{i-1}, r_i), \quad 1 \leqslant i \leqslant n$$

where $S_i = \{r_j\}_{j=1}^{i}$ (S_0 is empty), so that $\tilde{\varepsilon}_r(r)$ is continuous at each $r_i \in S$ and

$$\begin{aligned}
\int_{\{r_i\}} \tilde{\varepsilon}_r(r)\,\mathrm{d}\Omega(r) = 0 &\Longrightarrow \int_{\{r_i\}} \alpha(r)|\tilde{\varepsilon}_r(r)|\,\mathrm{d}\Omega(r) = 0 \\
&\Longrightarrow \int_S \alpha(r)|\tilde{\varepsilon}_r(r)|\,\mathrm{d}\Omega(r) = 0. \qquad (A.1)
\end{aligned}$$

Also, with $\tilde{\varepsilon}$ so defined, $\tilde{\varepsilon}_r(r) = \varepsilon_r(r)$ for all $r \notin S$, so that

$$\int_{[r_0,r_n]-S} \alpha(r)|\varepsilon_r(r)|\,\mathrm{d}\Omega(r) = \int_{[r_0,r_n]-S} \alpha(r)|\tilde{\varepsilon}_r(r)|\,\mathrm{d}\Omega(r) = \int_{r_0}^{r_n} \alpha(r)|\tilde{\varepsilon}_r(r)|\,\mathrm{d}\Omega(r) \quad \text{(using (A.1))}. \tag{A.2}$$

We also have

$$\begin{aligned}
&\int_{r_0}^{r_n} \varepsilon(r)[u(r) - u_0(r)]\,\mathrm{d}\Omega(r) \\
&\quad = \int_{r_0}^{r_n} \tilde{\varepsilon}(r)[u(r) - u_0(r)]\,\mathrm{d}\Omega(r) + \int_{r_0}^{r_n} [\varepsilon(r) - \tilde{\varepsilon}(r)][u(r) - u_0(r)]\,\mathrm{d}\Omega(r) \\
&\quad = \int_{r_0}^{r_n} \tilde{\varepsilon}(r)[u(r) - u_0(r)]\,\mathrm{d}\Omega(r) \\
&\qquad + \sum_{i=1}^{n} \left\{ \int_{r_{i-1}}^{r_i} [\varepsilon(r) - \tilde{\varepsilon}(r)][u(r) - u_0(r)]\,\mathrm{d}\Omega(r) \right\} \\
&\quad = \int_{r_0}^{r_n} \tilde{\varepsilon}(r)[u(r) - u_0(r)]\,\mathrm{d}\Omega(r) \\
&\qquad + \sum_{i=1}^{n} \left\{ \int_{r_{i-1}}^{r_i} \left[\int_{S_{i-1}} \varepsilon_t(t)\,\mathrm{d}\Omega(t) \right] [u(r) - u_0(r)]\,\mathrm{d}\Omega(r) \right\} \\
&\quad = \int_{r_0}^{r_n} \tilde{\varepsilon}(r)[u(r) - u_0(r)]\,\mathrm{d}\Omega(r) \\
&\qquad + \sum_{i=2}^{n-1} \left\{ [\alpha_{i-1}|\partial\Omega_{r_{i-1}}| - \alpha_i|\partial\Omega_{r_i}|] \left[\int_{S_{i-1}} \varepsilon_t(t)\,\mathrm{d}\Omega(t) \right] \right\} \\
&\qquad + \alpha_{n-1}|\partial\Omega_{r_{n-1}}| \int_{S_{n-1}} \varepsilon_r(r)\,\mathrm{d}\Omega(r) \quad \text{(using (7), (10), (11), and } S_0 = \emptyset) \\
&\quad = \int_{r_0}^{r_n} \tilde{\varepsilon}(r)[u(r) - u_0(r)]\,\mathrm{d}\Omega(r) \\
&\qquad + \sum_{i=2}^{n-1} \left\{ [\alpha_{i-1}|\partial\Omega_{r_{i-1}}| - \alpha_i|\partial\Omega_{r_i}|] \left[\sum_{j=1}^{i-1} \int_{\{r_j\}} \varepsilon_t(t)\,\mathrm{d}\Omega(t) \right] \right\} \\
&\qquad + \alpha_{n-1}|\partial\Omega_{r_{n-1}}| \int_{S_{n-1}} \varepsilon_r(r)\,\mathrm{d}\Omega(r) \\
&\quad = \int_{r_0}^{r_n} \tilde{\varepsilon}(r)[u(r) - u_0(r)]\,\mathrm{d}\Omega(r) \\
&\qquad + \sum_{j=1}^{n-2} \left\{ \int_{\{r_j\}} \varepsilon_t(t)\,\mathrm{d}\Omega(t) \sum_{i=j+1}^{n-1} [\alpha_{i-1}|\partial\Omega_{r_{i-1}}| - \alpha_i|\partial\Omega_{r_i}|] \right\} \\
&\qquad + \alpha_{n-1}|\partial\Omega_{r_{n-1}}| \int_{S_{n-1}} \varepsilon_r(r)\,\mathrm{d}\Omega(r) \\
&\quad = \int_{r_0}^{r_n} \tilde{\varepsilon}(r)[u(r) - u_0(r)]\,\mathrm{d}\Omega(r) + \sum_{i=1}^{n-1} \int_{\{r_i\}} \alpha(r)\varepsilon_r(r)\,\mathrm{d}\Omega(r) \\
&\quad = \int_{r_0}^{r_n} \tilde{\varepsilon}(r)[u(r) - u_0(r)]\,\mathrm{d}\Omega(r) + \int_S \alpha(r)\varepsilon_r(r)\,\mathrm{d}\Omega(r).
\end{aligned} \tag{A.3}$$

Next we define $\tilde{\tilde{\varepsilon}}(r)$ to be a vertically translated version of $\tilde{\varepsilon}(r)$,

$$\tilde{\tilde{\varepsilon}}(r) \equiv \tilde{\varepsilon}(r) - \tilde{\varepsilon}(r_{n-1})$$

so that $\tilde{\tilde{\varepsilon}}(r_{n-1}) = 0$, which allows us to write

$$\tilde{\tilde{\varepsilon}}(r) = \int_{t=r_{n-1}}^{r} \tilde{\tilde{\varepsilon}}_t(t)\,\mathrm{d}t, \tag{A.4}$$

and $\tilde{\tilde{\varepsilon}}_r(r) = \tilde{\varepsilon}_r(r)$, which leads to

$$\int_{r_0}^{r_n} \alpha(r)|\tilde{\varepsilon}_r(r)|\,\mathrm{d}\Omega(r) = \int_{r_0}^{r_n} \alpha(r)|\tilde{\tilde{\varepsilon}}_r(r)|\,\mathrm{d}\Omega(r). \tag{A.5}$$

Then

$$\begin{aligned}
&\int_{r_0}^{r_n} \tilde{\varepsilon}(r)[u(r) - u_0(r)]\,\mathrm{d}\Omega(r) \\
&\quad= \int_{r_0}^{r_n} \tilde{\tilde{\varepsilon}}(r)[u(r) - u_0(r)]\,\mathrm{d}\Omega(r) + \tilde{\varepsilon}(r_{n-1}) \int_{r_0}^{r_n} [u(r) - u_0(r)]\,\mathrm{d}\Omega(r) \\
&\quad= \int_{r_0}^{r_n} \tilde{\tilde{\varepsilon}}(r)[u(r) - u_0(r)]\,\mathrm{d}\Omega(r) \quad \text{(using (12))} \\
&\quad= \int_{r=r_0}^{r_n} \left[\int_{t=r_{n-1}}^{r} \tilde{\tilde{\varepsilon}}_t(t)\,\mathrm{d}t\right][u(r) - u_0(r)]\,\mathrm{d}\Omega(r) \quad \text{(using (A.4))} \\
&\quad= \int_{t=r_0}^{r_n} h(t)\tilde{\tilde{\varepsilon}}_t(t)\,\mathrm{d}t
\end{aligned} \tag{A.6}$$

where

$$h(t) = \begin{cases} -\displaystyle\int_{r=r_0}^{t} u(r) - u_0(r)\,\mathrm{d}\Omega(r) & \text{for } r_0 \leqslant t < r_{n-1} \\ \displaystyle\int_{r=t}^{r_n} u(r) - u_0(r)\,\mathrm{d}\Omega(r) & \text{for } r_{n-1} \leqslant t \leqslant r_n. \end{cases}$$

We show that $|h(t)| \leqslant \alpha_{\max}\beta_d t^{d-1}$ for $r \in [r_0, r_n]$, so that

$$-|h(t)| \geqslant -\alpha_{\max}\beta_d t^{d-1} \qquad \text{for } t \in [r_0, r_n]. \tag{A.7}$$

If $r_0 \leqslant t < r_1$, then

$$\begin{aligned}
|h(t)| &= \beta_d \left| \int_{r_0}^{t} r^{d-1}[u(r) - u_0(r)]\,\mathrm{d}r \right| \quad \text{(using (6))} \\
&\leqslant \alpha_1\beta_d \frac{r_1^{d-1}(t^d - r_0^d)}{r_1^d - r_0^d} \quad \text{(using (7), (10) and (11))} \\
&\leqslant \alpha_1\beta_d t^{d-1} \\
&\leqslant \alpha_{\max}\beta_d t^{d-1} \quad \text{(using (9))}.
\end{aligned}$$

Similarly, for each j, $1 \leqslant j \leqslant n-2$, if $r_j \leqslant t \leqslant r_{j+1}$, then

$$\begin{aligned}
|h(t)| &= \left| \left\{ \int_{r_0}^{r_1} + \sum_{k=2}^{j} \int_{r_{k-1}}^{r_k} + \int_{r_j}^{t} \right\} u(r) - u_0(r)\,\mathrm{d}\Omega(r) \right| \\
&\leqslant \left\{ \int_{r_0}^{r_1} + \sum_{k=2}^{j} \int_{r_{k-1}}^{r_k} + \int_{r_j}^{t} \right\} |u(r) - u_0(r)|\,\mathrm{d}\Omega(r)
\end{aligned}$$

$$\leqslant \beta_d \left[\alpha_1 r_1^{d-1} + \sum_{k=2}^{j} \alpha_k (r_k^{d-1} - r_{k-1}^{d-1}) + \alpha_j (r_{j+1}^{d-1} - r_j^{d-1}) \frac{t^d - r_j^d}{r_{j+1}^d - r_j^d} \right]$$

(using (6), (7), (10) and (11))

$$\leqslant \alpha_{\max} \beta_d \left[r_1^{d-1} + \sum_{k=2}^{j} (r_k^{d-1} - r_{k-1}^{d-1}) + (r_{j+1}^{d-1} - r_j^{d-1}) \frac{t^d - r_j^d}{r_{j+1}^d - r_j^d} \right] \quad \text{(using (9))}$$

$$= \alpha_{\max} \beta_d \left[r_j^{d-1} + (r_{j+1}^{d-1} - r_j^{d-1}) \frac{t^d - r_j^d}{r_{j+1}^d - r_j^d} \right]$$

$$\leqslant \alpha_{\max} \beta_d (r_j^{d-1} + t^{d-1} - r_j^{d-1})$$

$$= \alpha_{\max} \beta_d t^{d-1}.$$

Finally, if $r_{n-1} \leqslant t \leqslant r_n$, then

$$|h(t)| = \beta_d \left| \int_{r=t}^{r_n} r^{d-1} [u(r) - u_0(r)] \, \mathrm{d}r \right| \quad \text{(using (6))}$$

$$= \alpha_{n-1} \beta_d \frac{r_{n-1}^{d-1} (r_n^d - t^d)}{r_n^d - r_{n-1}^d} \quad \text{(using (7), (10) and (11))}$$

$$\leqslant \alpha_{n-1} \beta_d t^{d-1}$$

$$\leqslant \alpha_{\max} \beta_d t^{d-1} \quad \text{(using (9))}.$$

So

$$\int_{r_0}^{r_n} \tilde{\varepsilon}(r)[u(r) - u_0(r)] \, \mathrm{d}\Omega(r)$$

$$= \int_{t=r_0}^{r_n} h(t) \tilde{\tilde{\varepsilon}}_t(t) \, \mathrm{d}t \quad \text{(using (A.6))}$$

$$\geqslant -\int_{r_0}^{r_n} |h(t)| |\tilde{\tilde{\varepsilon}}_t(t)| \, \mathrm{d}t$$

$$\geqslant -\alpha_{\max} \int_{r_0}^{r_n} |\tilde{\tilde{\varepsilon}}_t(t)| \, \mathrm{d}\Omega(t) \quad \text{(using (A.7))}$$

$$\geqslant -\int_{r_0}^{r_n} \alpha(t) |\tilde{\tilde{\varepsilon}}_t(t)| \, \mathrm{d}\Omega(t) \quad \text{(using (9) and (A.1))}$$

$$= -\int_{r_0}^{r_n} \alpha(t) |\tilde{\varepsilon}_t(t)| \, \mathrm{d}\Omega(t) \quad \text{(using (A.5))}. \tag{A.8}$$

Finally, we have

$$g(u, \varepsilon) = \int_{r_0}^{r_n} \tilde{\varepsilon}(r)[u(r) - u_0(r)] \, \mathrm{d}\Omega(r) + \int_{r_0}^{r_n} \alpha(r) |\tilde{\varepsilon}_r(r)| \, \mathrm{d}\Omega(r)$$

$$+ \int_{r_0}^{r_n} \tfrac{1}{2} [\varepsilon(r)]^2 \, \mathrm{d}\Omega(r) \quad \text{(using (A.2) and (A.3))}$$

$$\geqslant \int_{r_0}^{r_n} \tfrac{1}{2} [\varepsilon(r)]^2 \, \mathrm{d}\Omega(r) \quad \text{(using (A.8))}$$

$$> 0 \quad \text{(unless } \varepsilon \equiv 0\text{)}.$$

So $f(u + \varepsilon) \equiv f(u) + g(u, \varepsilon) > f(u)$, unless $\varepsilon \equiv 0$. Therefore u is the unique solution to (5). □

References

[1] Acar R and Vogel C 1994 Analysis of bounded variation penalty methods for ill-posed problems *Inverse Problems* **10** 1217–29

[2] Alvarez L, Guichard F, Lions P-L and Morel J-M 1993 Axioms and fundamental equations of image processing *Arch. Ration. Mech. Anal.* **123** 200–57

[3] Alvarez L and Morel J-M 1994 Formalization and computational aspects of image analysis *Acta Numerica* 1–59

[4] Andreu F, Caselles V, Diaz J and Mazon J 2002 Some qualitative properties for the total variation flow *J. Funct. Anal.* **188** 516–47

[5] Blomgren P 1998 Total variation methods for restoration of vector valued images *PhD Dissertation (UCLA Math Department CAM Report* 98-30*)*

[6] Blomgren P and Chan T 1998 Color TV: total variation methods for restoration of vector-valued images *IEEE Trans. Image Process.* **7** 304–9

[7] Bellettini G, Caselles V and Novaga M 2002 The total variation flow in $\mathbf{R}^N$ *J. Diff. Eqns* **184** 475–525

[8] Catté F, Lions P-L, Morel J-M and Coll T 1992 Image selective smoothing and edge detection by nonlinear diffusion *SIAM J. Numer. Anal.* **29** 182–93

[9] Chambolle A and Lions P-L 1997 Image recovery via total variation minimization and related problems *Numer. Math.* **76** 167–88

[10] Chan T, Golub G and Mulet P 1999 A nonlinear primal-dual method for total variation-based image restoration *SIAM J. Sci. Comput.* **20** 1964–77

[11] Chan T, Marquina A and Mulet P 2000 High-order total variation-based image restoration *SIAM J. Sci. Comput.* **22** 503–16

[12] Chan T and Mulet P 1999 On the convergence of the lagged diffusivity fixed point method in total variation image restoration *SIAM J. Numer. Anal.* **36** 354–67

[13] Chan T, Osher S and Shen J 2001 The digital TV filter and nonlinear denoising *IEEE Trans. Image Process.* **10** 231–41

[14] Chan T and Shen J 2002 Mathematical models for local nontexture inpaintings *SIAM J. Appl. Math.* **62** 1019–43

[15] Chan T and Shen J 2001 Variational restoration of nonflat image features: models and algorithms *SIAM J. Appl. Math.* **61** 1338–61

[16] Chan T and Wong C-K 1998 Total variation blind deconvolution *IEEE Trans. Image Process.* **7** 370–5

[17] Davies P L and Kovac A 2001 Local extremes, runs, string and multi-resolution *Ann. Stat.* **29** 1–65

[18] Dobson D and Santosa F 1996 Recovery of blocky images from noisy and blurred data *SIAM J. Appl. Math.* **56** 1181–98

[19] Dobson D and Scherzer O 1996 Analysis of regularized total variation penalty methods for denoising *Inverse Problems* **12** 601–17

[20] Dobson D and Vogel C R 1997 Convergence of an iterative method for total variation denoising *SIAM J. Numer. Anal.* **34** 1779–91

[21] Gousseau Y and Morel J-M 2001 Are natural images of bounded variation? *SIAM J. Math. Anal.* **33** 634–48

[22] Groetsch C 1984 *The Theory of Tikhonov Regularization for Fredholm Integral Equations of the First Kind* (Boston, MA: Pitman)

[23] ter Haar Romeny B M (ed) 1994 *Geometry-Driven Diffusion in Computer Vision* (Dordrecht: Kluwer)

[24] Ito K and Kunisch K 2000 BV-type regularization methods for convoluted objects with edge, flat and grey scales *Inverse Problems* **16** 909–28

[25] Karkkainen T, Majava K and Makela M 2001 Comparison of formulations and solution methods for image restoration problems *Inverse Problems* **17** 1977–95

[26] Keeling S and Stollberger R 2002 Nonlinear anisotropic diffusion filtering for multiscale edge enhancement *Inverse Problems* **18** 175–90

[27] Mammen E and van de Geer S 1997 Locally adaptive regression splines *Ann. Stat.* **25** 387–413

[28] Marquina A and Osher S 2000 Explicit algorithms for a new time dependent model based on level set motion for nonlinear deblurring and noise removal *SIAM J. Sci. Comput.* **22** 387–405

[29] Nikolova M 1998 Local strong homogeneity of a regularized estimator *SIAM J. Appl. Math.* **61** 633–58

[30] Perona P 1998 Orientation diffusion *IEEE Trans. Image Process.* **7** 457–67

[31] Perona P and Malik J 1990 Scale space and edge detection using anisotropic diffusion *IEEE Trans. Pattern Anal. Mach. Intell.* **12** 629–39

[32] Ring W 1999 Structural properties of solutions of total variation regularization problems *Technical report* University of Graz, Austria (Available at http://www.kfunigraz.ac.at/imawww/ring)

[33] Rudin L, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica* D **60** 259–68

[34] Sapiro G and Ringach D 1996 Anisotropic diffusion of multivalued images with applications to color filtering *IEEE Trans. Image Process.* **5** 1582–6
[35] Strang G 1982 L^1 and L^∞ approximation of vector fields in the plane *Nonlinear Partial Differential Equations in Applied Science* (*Lecture Notes in Numer. Appl. Anal.* vol 5) pp 273–88
[36] Strong D 1997 Adaptive total variation minimizing image restoration *PhD Dissertation (UCLA Math. Department CAM Report* 97-38*)*
[37] Strong D, Blomgren P and Chan T 1997 Feature-driven adaptive total variation minimizing image restoration *Proc. SPIE* **3167** 222–33
[38] Strong D and Chan T 1996 Spatially and scale adaptive total variation based regularization and anisotropic diffusion in image processing *UCLA Math. Department CAM Report* 96-46
[39] Tyan S G 1981 Median filter: deterministic properties *Two-Dimensional Digital Signal Processing II. Transforms and Median Filters* (*Springer Topics in Applied Physics* vol 43) (Berlin: Springer) chapter 6 pp 197–218
[40] Tang B, Sapiro G and Caselles V 1999 Direction diffusion *ICCV* **2** 1245–52
[41] Vogel C and Oman M 1996 Iterative methods for total variation denoising *SIAM J. Sci. Comput.* **17** 227–38
[42] Winkler G 2003 *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction* (*Springer Applications of Mathematics* vol 27) (New York: Springer)
[43] Winkler G and Liebscher V 2002 Smoothers for discontinuous signals *J. Nonparametric Stat.* **14** 203–22
[44] You Y, Xu W, Tannenbaum A and Kaveh M 1996 Behavioral analysis of anisotropic diffusion in image processing *IEEE Trans. Image Process.* **5** 1539–53