Adv Comput Math (2009) 31:61–85
DOI 10.1007/s10444-008-9082-7

Graduated adaptive image denoising: local compromise between total variation and isotropic diffusion

Erik M. Bollt · Rick Chartrand · Selim Esedoḡlu · Pete Schultz · Kevin R. Vixie

Received: 12 February 2007 / Accepted: 23 October 2007 /
Published online: 3 July 2008

Abstract We introduce variants of the variational image denoising method proposed by Blomgren et al. (In: Numerical Analysis 1999 (Dundee), pp. 43–67. Chapman & Hall, Boca Raton, FL, 2000), which interpolates between total-variation denoising and isotropic diffusion denoising. We study how parameter choices affect results and allow tuning between TV denoising and isotropic diffusion for respecting texture on one spatial scale while denoising features assumed to be noise on finer spatial scales. Furthermore, we prove existence and (where appropriate) uniqueness of minimizers. We consider both L^2 and L^1 data fidelity terms.

Keywords Graduated adaptive image denoising · Total variation · Isotropic diffusion

Mathematics Subject Classifications (2000) 35R30 · 65K10 · 49Q10 · 49L99 · 65J20

Communicated by Lixin Shen and Yuesheng Xu.

E. M. Bollt (✉)
Clarkson University, P.O. Box 5815, Potsdam, NY 13699, USA
e-mail: bolltem@clarkson.edu

R. Chartrand · P. Schultz · K. R. Vixie
Los Alamos National Laboratory, Theoretical Division, MS B284,
Los Alamos, NM 87545, USA

S. Esedoḡlu
Department of Mathematics, University of Michigan,
2074 East Hall 530 Church Street, Ann Arbor, MI 48109, USA

1 Introduction

Measured images typically provide us with noisy, discretized versions of an observed scene. In many important cases, high noise levels can make recognition of important features very difficult. Therefore, effective denoising plays an important role in the processing and understanding of images. Approaches to image denoising have been developed along three main lines: (1) wavelet-based methods—for example, methods based on wavelet thresholding introduced by Donoho and Johnstone [7], (2) stochastic or statistical methods—a prominent example is the Markov random field (MRF) approach introduced by Geman and Geman [11], and (3) PDE/variational approaches. One variational approach that has attracted a great deal of attention is the total variation method of Rudin, Osher and Fatemi [13] (ROF). The variational problem is

$$\min_u J(u) = \int_\Omega |\nabla u| + \frac{\lambda}{2}\int_\Omega |d-u|^2. \tag{1}$$

We are representing true or reconstructed images by a function $u : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$. The function d represents the measured image, which we suppose is of the form $u + \eta$, the sum of the true image and a noise component. Much of the attention the ROF approach has attracted is due to the fact that the method simultaneously reduces noise and preserve edges. An older approach is

$$\min_u J(u) = \int_\Omega |\nabla u|^2 + \frac{\lambda}{2}\int_\Omega |d-u|^2, \tag{2}$$

which differs from ROF in having the Dirichlet integral of u in place of the total variation of u. The Euler-Lagrange equation for this model is a diffusion equation. The method is very effective at removing noise but has the drawback of smearing or degrading edges in the image. It is interesting to note that this smearing out depends on the curvature of the edges, so that thresholding a smeared out edge does not recover the true edge or even a uniformly displaced true edge.

The ROF method above suffers from *staircasing*, which is a noise induced introduction of artificial steps or discontinuities into the reconstructed or denoised image. In an attempt to reduce this effect, Blomgren et al. [2] suggested letting the exponent in the regularization term depend on the image. Their functional,

$$J(u) = \int_\Omega |\nabla u|^{p(|\nabla u|)} + \frac{\lambda}{2}\int_\Omega |d-u|^2, \tag{3}$$

uses a decreasing function p such that $p(0) = 2$ and $p(x) = 1$ for all x greater than some M to adapt the smoothing to the gradient in the image. In their paper, they show a numerical example suggesting a reduction of staircasing.

More recently, Levine et al. [5, 12] have looked at a variant of the method suggested by Blomgren et al. They consider the case in which the exponent $p(x)$ depends only on the measured data d. To be more precise, they study

$$J(u) = \int_\Omega \phi(x, \nabla u) + \frac{\lambda}{2}\int_\Omega |u-d|^2, \tag{4}$$

where

$$\phi(x,r) = \begin{cases} \frac{1}{p(x)}|r|^{p(x)} & \text{if } |r| < \epsilon \\ |r| - \frac{\epsilon p(x) - \epsilon^{p(x)}}{p(x)} & \text{if } |r| \geq \epsilon \end{cases}, \tag{5}$$

$\epsilon > 0$ is fixed, and $p(x)$ is based on a smoothed version of the observed image d:

$$p(x) = \frac{1}{1 + k|\nabla G_\sigma * d(x)|^2}, \tag{6}$$

where k and σ are parameters, and

$$G_\sigma(x) = \frac{1}{\sigma} e^{\frac{-|x|^2}{4\sigma^2}} \tag{7}$$

is a Gaussian smoothing kernel. The authors show existence and uniqueness of minimizers for this functional, and develop a numerical method for computing them based on gradient descent.

It is also interesting to note that in their discussion of inf-convolution for the combination of the two convex functionals $\int |\nabla u|$ and $\int |\nabla u|^2$, Chambolle and Lions [3] introduce

$$F(u) = \frac{1}{2\epsilon} \int_{|\nabla u| < \epsilon} |\nabla u|^2 + \int_{|\nabla u| \geq \epsilon} \left(|\nabla u| - \frac{\epsilon}{2} \right) + \int_\Omega |u - d|^2 \tag{8}$$

where ϵ is a parameter that must be chosen a priori. This is suggestive of, though not identical to, the case of

$$p = \begin{cases} 2 & \text{if } s < \epsilon \\ 1 & \text{if } s \geq \epsilon \end{cases}. \tag{9}$$

2 Outline of paper

In this paper, we study

$$\min_u J(u) = \int_\Omega |\nabla u|^{p(|\nabla u|)} + \frac{\lambda}{2} \int_\Omega |d - u|^q, \ q = 1 \text{ or } 2 \tag{10}$$

in two cases:

- Case 1: $p = P(|\nabla(G_\delta * d)(x)|)$, where $s = |\nabla(G_\delta * d)(x)|$; and
- Case 2: $p = P(|\nabla(G_\delta * u)(x)|)$ where $s = |\nabla(G_\delta * u)(x)|$,

where G_δ is a smoothing kernel.

We begin by introducing a simple numerical method for minimizing (10) for Case 1, and explore the behavior of this functional and its dependence on the various parameters that define the smoothing kernel G_σ and the exponent function $p(x)$. We then prove existence and, in the case of $q = 2$, uniqueness of minimizers in both cases.

3 The case $p = P(|\nabla(G_\delta * d)(x)|)$

We choose $q = 2$ and p given by

$$p = P_M\left(|\nabla(G_\delta * d)(x)|^2\right) \tag{11}$$

where $G_\delta : \mathbb{R}^2 \to \mathbb{R}$ is a symmetric mollifier centered at 0 and supported on a disk of radius δ with maximum height K. $P_M : \mathbb{R}^+ \to [1, 2]$ is defined by

$$P_M(s) = \begin{cases} 2 - \frac{10s^3}{M^3} + \frac{15s^4}{M^4} - \frac{6s^5}{M^5} & \text{if } s \le M \\ 0 & \text{if } s > M \end{cases}. \tag{12}$$

The polynomial is the quintic polynomial with the properties $P_M(0) = 2$, $P_M(M) = 1$, $P'_M(0) = 0$, $P'_M(M) = 0$, $P''_M(0) = 0$, and $P''_M(M) = 0$. See Fig. 1.

The parameter M prescribes the size of a derivative required to consider total-variation denoising to be locally the best option.

Note that our functional,

$$J(u) = \int_\Omega |\nabla u|^{P_M(|\nabla(G_\delta * d)|^2)} + \frac{\lambda}{2}\|d - u\|^2_{L^2(\Omega)}, \tag{13}$$

uses only the initial data d to control the level of regularization.

We should note that this case is significantly simpler than Case 2, in which d is replaced by u. One way to see this is by comparing the formal Euler-Lagrange equation for the two cases. For Case 1 which we are considering here, the Euler-Lagrange equation is

$$0 = -\nabla \cdot \left[\frac{\nabla u}{|\nabla u|} P_M\left(|\nabla(G_\delta * d)|^2\right) |\nabla u|^{P_M(|\nabla(G_\delta * d)|^2)-1}\right] + \lambda(u - d). \tag{14}$$

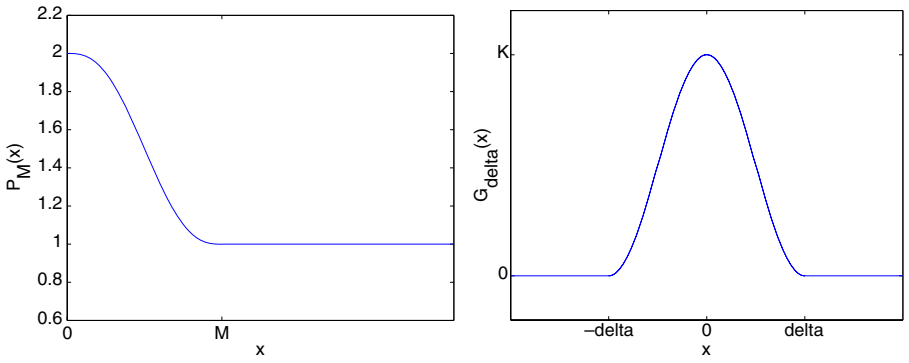

Fig. 1 *Left* P_M function from (12). *Right* $G_\delta(x)$ function for regularizing gradients in the exponent in $J(u)$

For Case 2, the Euler-Lagrange equation is

$$0 = -G_\delta * \nabla \cdot \left[|\nabla u|^{P_M(|\nabla(G_\delta * u)|^2)} P'(|\nabla(G_\delta * u)|^2) 2\nabla(G_\delta * u) \right]$$
$$-\nabla \cdot \left[|\nabla u|^{P_M(|\nabla(G_\delta * u)|^2)} \frac{\nabla u}{|\nabla u|} \right] + \lambda(u - d). \tag{15}$$

Our implementation of gradient descent using the Euler-Lagrange equation is entirely straightforward. Taking $g^n_{i,j}$ to be a standard finite difference approximation of the right hand side of (14) at $x_{i,j}$ and t_n, we get an Euler-like updating scheme,

$$u^{n+1}_{i,j} = u^n_{i,j} + \Delta t g^n_{i,j}. \tag{16}$$

We use an adaptive step size scheme. The new value $u^{n+1}_{i,j}$ is accepted for each step in which the cost is improved, $J(u^{n+1}_{i,j}) < J(u^n_{i,j})$, and the step Δt is increased by a factor $\Delta t \to s\Delta t$, $s > 1$. For each unsuccessful step where $J(u^{n+1}_{i,j}) \geq J(u^n_{i,j})$, the trial step is not used, and the step size is decreased, $\Delta t \to s\Delta t$, $s < 1$.

3.1 Binary image example: edges, texture and noise

In this section we investigate the ability of the functional in (13) to preserve or recover edges, smooth sections, and texture, while removing noise. Making the necessary distinctions is an inherently ill-defined task. Noise, when viewed at one scale, can look very much like texture on another scale. Separating noise and texture depends on having a separation of scale between them. Thus, the best we can hope for is a method which preserves texture on one spatial scale, while removing noise at smaller spatial scales.

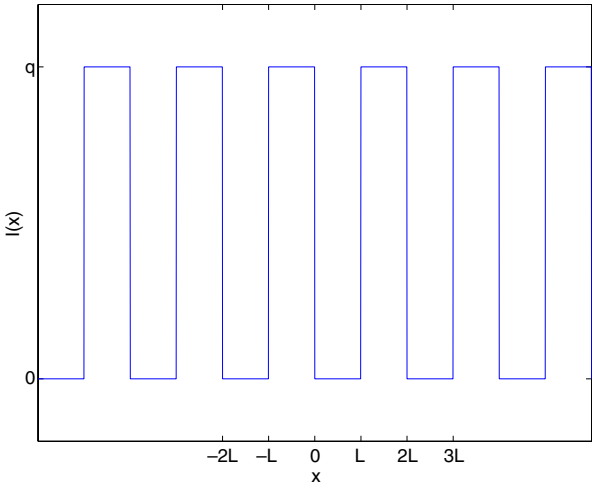

Fig. 2 Binary test image in one dimension designed to explore noise versus texture scales in denoising procedure when choosing parameters δ, M, and K

We now turn to a detailed exploration of a simple example. Figure 2 shows I, our one-dimensional binary test image. I is given by

$$I(x) = A \,\text{floor}(x/L) \mod 2, \tag{17}$$

and is a one-dimensional model of a noiseless image with texture of spatial period $2L$ and amplitude A. The measured data d is given by

$$d(x) = I(x) + \text{noise} = I(x) + \mathcal{N}(0, \sigma), \tag{18}$$

where $\mathcal{N}(0, \sigma)$ is Gaussian noise with standard deviation σ and mean 0. The resulting 7 parameters—M, H, δ, L, A, σ, and λ—interact to influence the precise nature of the minimizer to (13).

There are at least two parametric dichotomies that are relevant:

- $AH > M$ vs. $AH < M$: this dichotomy tells us whether or not edges of the test image are preserved or degraded.
  - If $AH > M$, there is an interval around each edge where $|\nabla(G_\delta * I)(x)|^2 > M$. If the noise level is not too large, then $|\nabla(G_\delta * d)(x)|^2 > M$ as well, so that $p(x) = 1$. This means that the denoising is locally

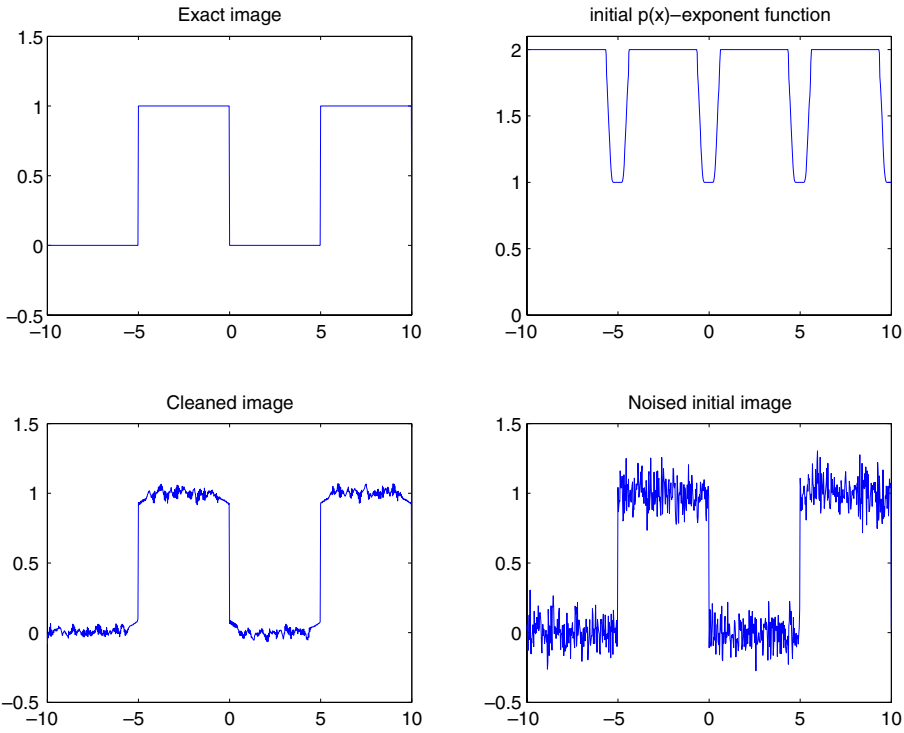

Fig. 3 Edges preserved: The $p(x)$ function resulting from the *data image in lower right* is shown in the *upper right*. $AH > M$ implies an interval around edges where optimization of the functional (13) behaves locally as total-variation denoising. Notice the *edges at each L* are well preserved as desired, and the *smaller-scaled noise* is smoothed

like total-variation denoising around each edge, and the edges are preserved. See Fig. 3.
  - If $AH < M$, then $|\nabla(G_\delta * I)(x)|^2$ is always less than M. Assuming a reasonable noise level, $p(x) > 1$ for all x in the image. The resulting diffusion not only smooths out the noise, but also rounds off the edges. See Fig. 4.

- $2\delta >> L$ vs. $2\delta \leq L$: This determines whether or not the square wave, now thought of as texture, is preserved or smoothed out.
  - If $2\delta \leq L$, the smoothing kernel acts on a spacial scale that is too small to smooth out the texture. Examples are Figs. 3 and 4, where the texture is preserved by the denoising process.
  - If $2\delta >> L$, the smoothing kernel does not distinguish the texture from noise. See Fig. 5.

Finally there is the issue of the relation between noise and features such as edges or texture. If there is a separation in spatial scale between the noise

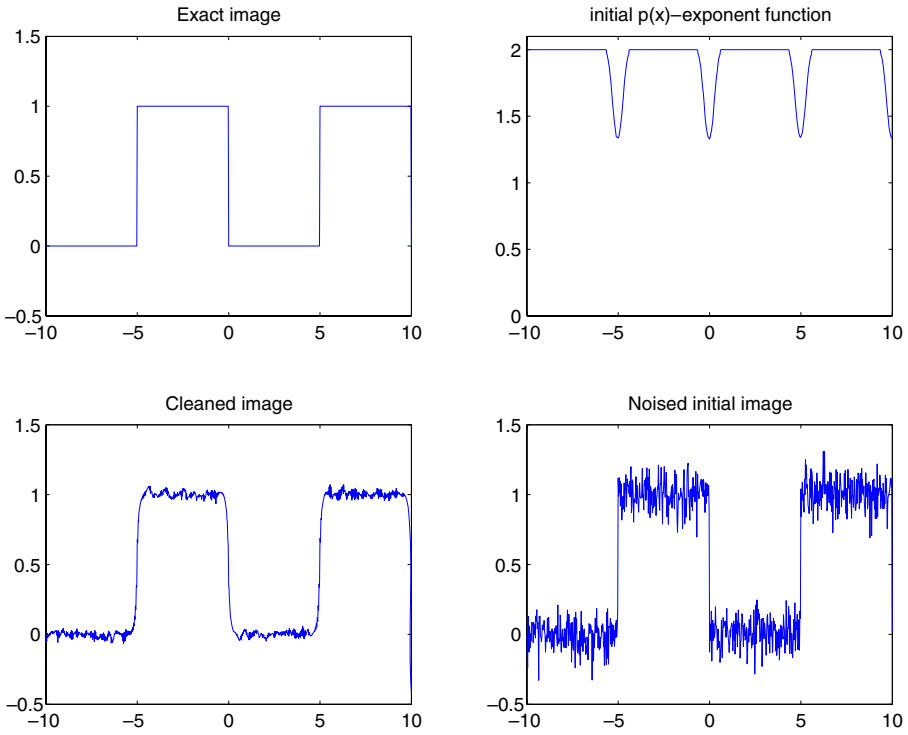

Fig. 4 *Edges degraded* The $p(x)$ function resulting from the *data image in lower right* is shown in the *upper right*. $AH < M$ implies that around the edges, optimization of the functional (13) behaves locally as a diffusion. Notice the edges at each L are *rounded* due to the fact that diffusion does not respect edges, as does total-variation denoising

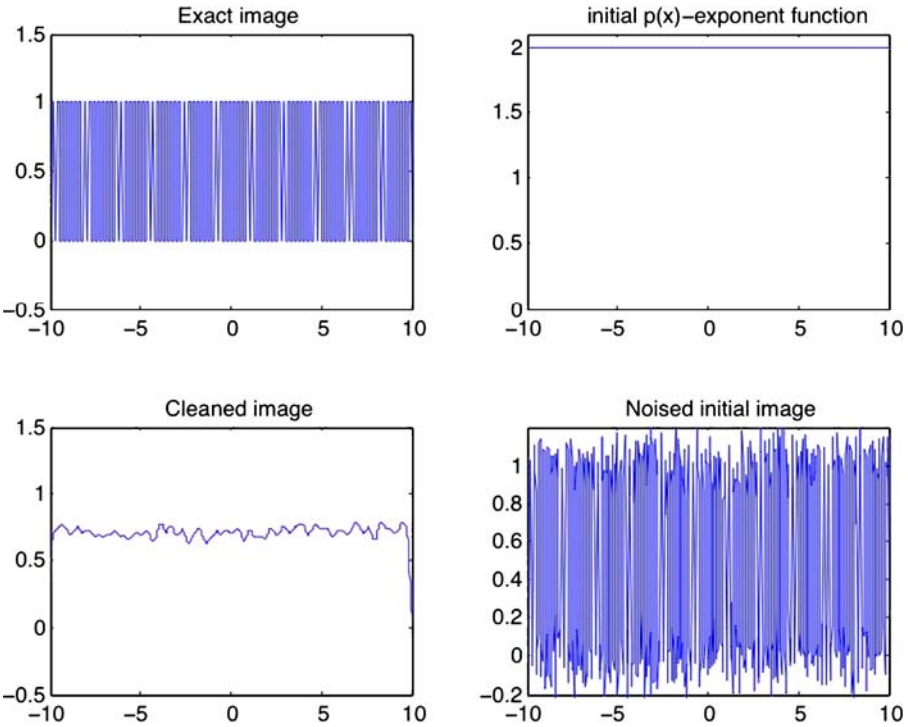

Fig. 5 *Texture smoothed* the small scale of the texture ($L = 1/4$) in relation to the scale of the smoothing kernel ($\delta = 10$) results in the texture being smoothed away like noise

and the texture or if there is a separation in amplitude between the noise and edge heights, δ, H, and M can be chosen to eliminate noise while preserving edges or texture or both. We must have δ large enough to smooth out the noise oscillation without being so large that either the edges are degraded or the texture smoothed. Since this depends on σ, A, and L characterizing the measured image, we cannot choose our parameters δ, H, and M a priori.

3.2 Natural image examples

We now show the results of the data driven exponent function $p(x)$ on natural images. See Figs. 6, 7, 8, and 9.

As expected, results depend on our ability to tune δ, K, and M to recover texture and edges while removing noise. We see in the images shown that this closely relates to building a P_M function which is essentially equivalent to edge detection. In particular, in Figs. 10, 11, we see the effect of a study of increasing the kernel standard deviation from 0.01 to 0.1, on noisy figures derived from the famous image of Lena, and another figure of the Jemez Mountains near Los Alamos New Mexico. Both images present particular challenges to standard

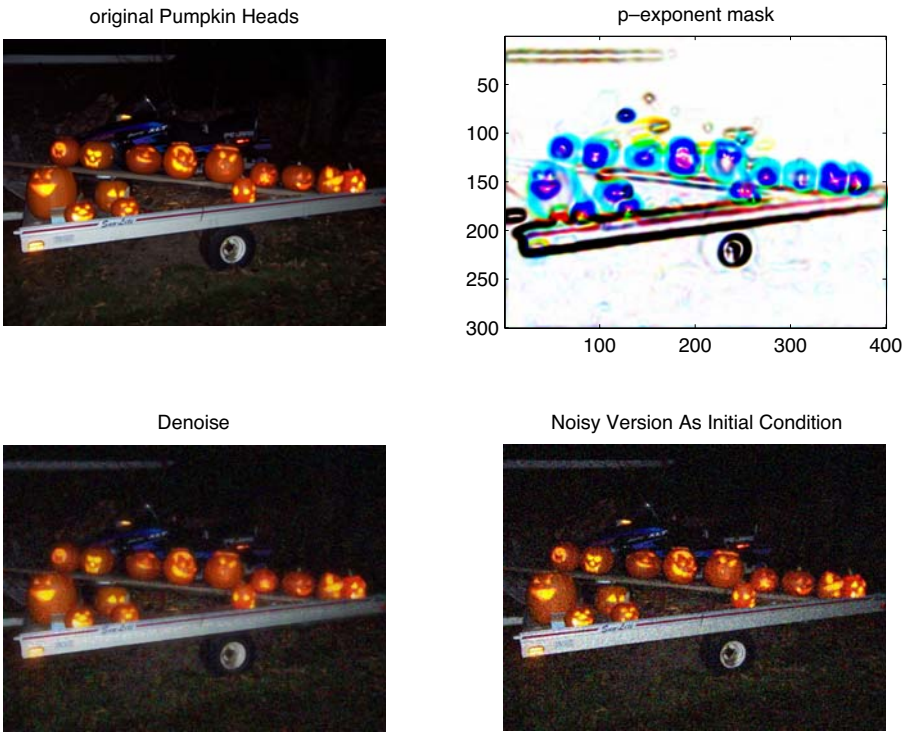

Fig. 6 *Upper left* the original picture "Pumpkin Heads." The *red horizontal line through the middle* depicts the location of the one-dimensional slice shown in Fig. 7. *Upper right* the exponent $p(x)$ is computed based on the polynomial (12); this exponent "mask" is shown for each of the three color planes, for the most recent iteration, using the scheme that the exponent $p(x)$ is a function of the current best u. See also a slice of $p(x)$ in Fig. 7. *Lower left* the cleaned image. Apparently, the fact that different edges are detected in each of the different color planes, shown (*upper right*), lends to a more natural result, with a more three-dimensional appearance. *Lower right* the noisy initial condition, $d(x) = I(x) + \text{noise}$

methods as they display both fine scale and large scale structures. We see how the changing of the kernel width can emphasize these scales, which can compete with the noise scale, as described herein. Taking this hint, the next subsection discusses a surrogate for the exponent function $p(x)$ based on a simple edge detection method.

3.3 A surrogate for $p = P_M(|\nabla(G_\sigma * d)(x)|^2)$: $P_E(x)$ from Canny edge detection

A simplification can be obtained by first detecting edges and then building the exponent function based on the results from the edge detection. For convenience, we chose to pass an image to the Canny edge detection algorithm in the Matlab image processing toolbox. Let $\mathbb{Z}_{256}^{m,n}$ be the space of 8-bit, gray scale, m by n pixellated images. Then a three color image can be represented

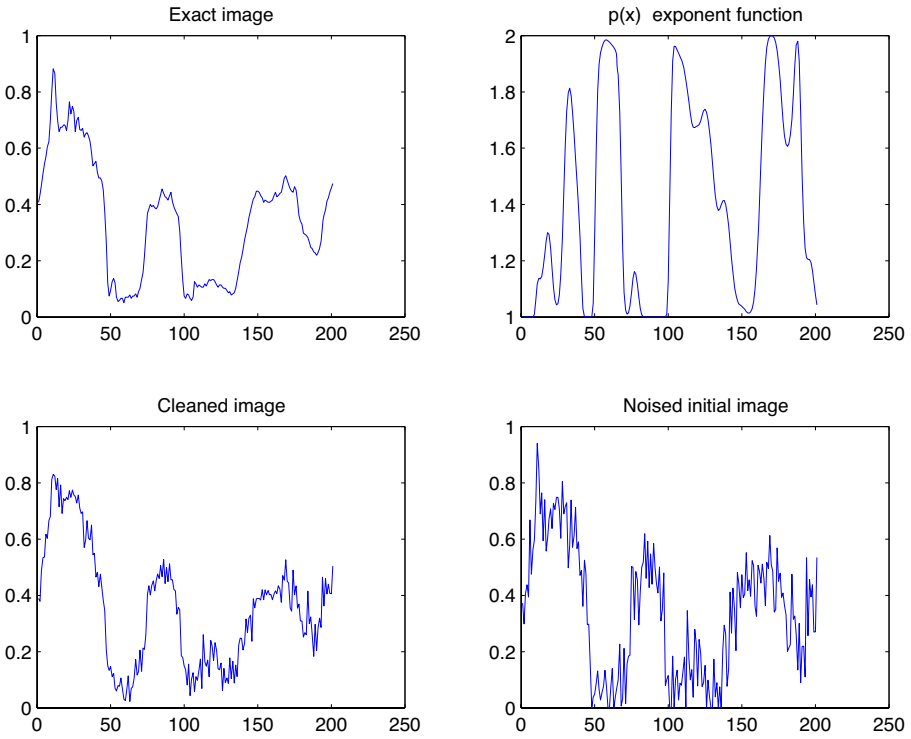

Fig. 7 A one-dimensional slice of the Pumpkins Head data shown in Fig. 6 shown at the position along the *horizontal red line* depicted in Fig. 6 (*upper left*). Note the exponent ranges from $1 \leq p(x) \leq 2$

by $I \in (\mathbb{Z}_{256}^{m,n})^3$ in which each of the color planes describes one of the RGB channels. An edge detection algorithm can be represented by a function

$$E : \left(\mathbb{Z}_{256}^{m,n}\right)^3 \to \left(\mathbb{Z}_2^{m,n}\right)^3. \tag{19}$$

The 3 edge images or planes are binary images in each color plane with a value 0 away from edges and a value 1 at detected edges. Using this we define

$$P_E(I) = \mathbf{2} - G_\delta * E(I), \tag{20}$$

where $\mathbf{2}$ is an $m \times n \times 3$ array of 2s, and $E(I)$ is the edge image from the edge detection scheme. The convolution with G_δ smooths the transition from detected edges to flat regions. The results are shown in Figs. 12 and 13.

The edge detection algorithm finds different edges in each of the three color planes. It is our opinion that the different edges in each of the three color planes leads to a more natural (three-dimensional) appearance of the denoised minimizer u shown in Fig. 12.

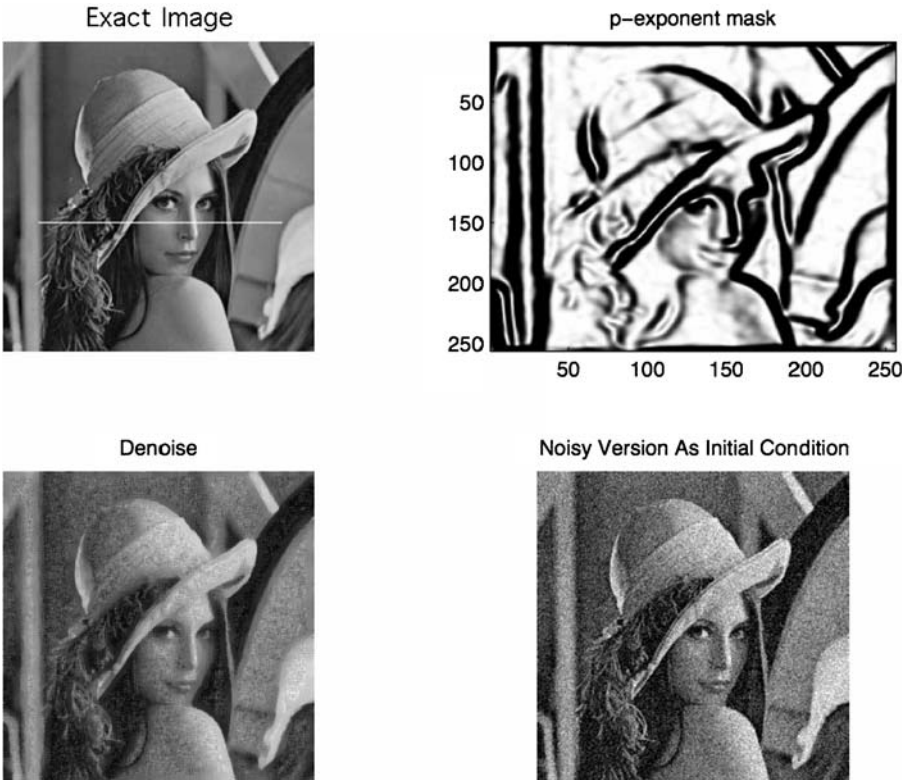

Fig. 8 No paper in image processing is complete without a picture of this lady, the ubiquitous Lena. The *white horizontal line* shows the location of the one-dimensional slice plots shown in Fig. 9. All relevant arrangements are as in Pumpkin Heads, Fig. 6

4 Existence and uniqueness of minimizers

In this section we give proof of existence of a minimizer of the variable exponent gradient based functional. References for the theorems and results used in this section are [6, 8–10]. Throughout, Ω is a bounded open subset of $\mathbb{R}^m$. For d and u in $BV(\Omega)$, consider

$$J(u) = F(u) + Q(u), \tag{21}$$

where $F(u)$ is a generalized regularization term

$$F(u) = \int_\Omega |\nabla u|^{\tilde{p}} \tag{22}$$

and the data fidelity term $Q(u)$ is either

$$Q(u) = \int_\Omega |d - u| = \|d - u\|_1, \tag{23}$$

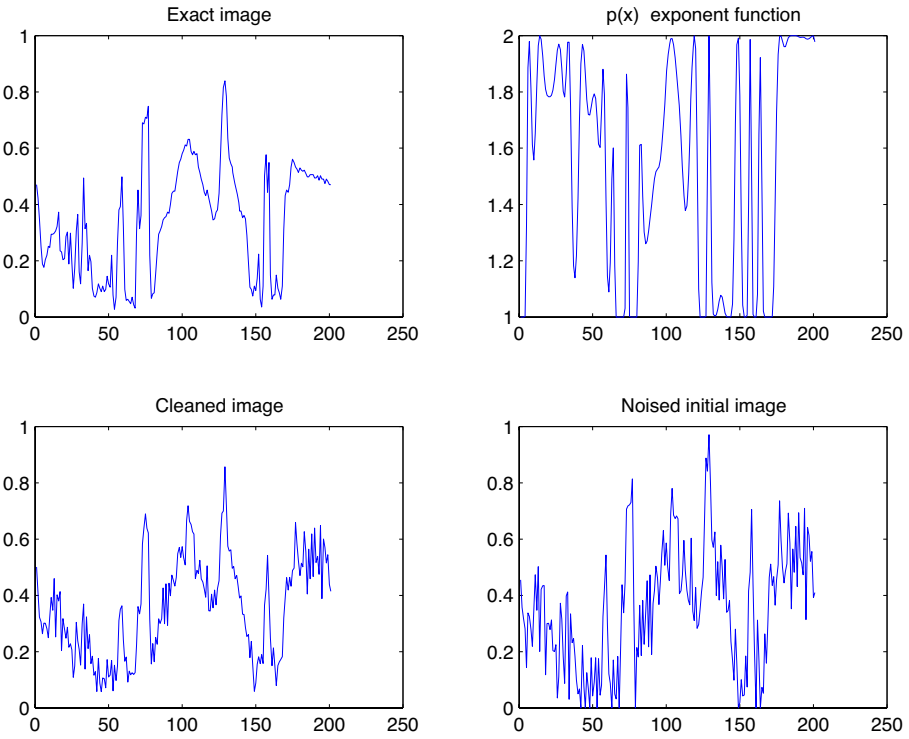

Fig. 9 The one-dimensional slice of Lena, at the location of the *horizontal white line* across Lena's nose in Fig. 8 (*upper left*)

or

$$Q(u) = \int_{\Omega} |d-u|^2 = \|d-u\|_2^2. \tag{24}$$

We are most interested in the two cases

$$\tilde{p} = p(u) = p_0\left(|\nabla(G_\delta * u)|^2\right), \tag{25}$$

$$\tilde{p} = p(d) = p_0\left(|\nabla(G_\delta * d)|^2\right) \tag{26}$$

where G_δ is a $C^2 \cap W^{3,2}$ mollifier supported on $|x| \leq \delta$ and $p_0 : [0, \infty) \to [1, 2]$ is a nonincreasing C^2 function with $p_0(M) = 1$ for some $M > 0$.

We now rewrite $F(u)$ in a form that will allow for computation of weak derivatives. For $p \geq 1$, the smooth convex function $x \mapsto x^p$ is the supremum of all its tangent lines. Computing the value at $x = |t|$ of the tangent line with slope $v \geq 0$ gives that

$$|t|^p = \max_v \left(pv^{(p-1)}|t| - (p-1)v^p\right). \tag{27}$$

Fig. 10 A study of changing kernel widths as applied to a noisy image of Lena. Changing the kernel clearly emphasizes structure scale versus noise scale trade-off as described herein

Thus, at a fixed x,

$$|\nabla u(x)|^{\tilde{p}} = \max_{\nu \geq 0} \left(\tilde{p} \nu^{\tilde{p}-1} |\nabla u(x)| - (\tilde{p}-1)\nu^{\tilde{p}} \right) \tag{28}$$

$$= \max_{\substack{\nu \geq 0 \\ |\sigma| \leq 1}} \left(\tilde{p} \nu^{\tilde{p}-1} \sigma \cdot \nabla u(x) - (\tilde{p}-1)\nu^{\tilde{p}} \right). \tag{29}$$

This suggests that we can write

$$F(u) = \max_{\substack{\nu(x) \geq 0 \\ |\sigma(x)| \leq 1}} \int \left(\tilde{p} \nu^{\tilde{p}-1} \sigma \cdot \nabla u - (\tilde{p}-1)\nu^{\tilde{p}} \right) \tag{30}$$

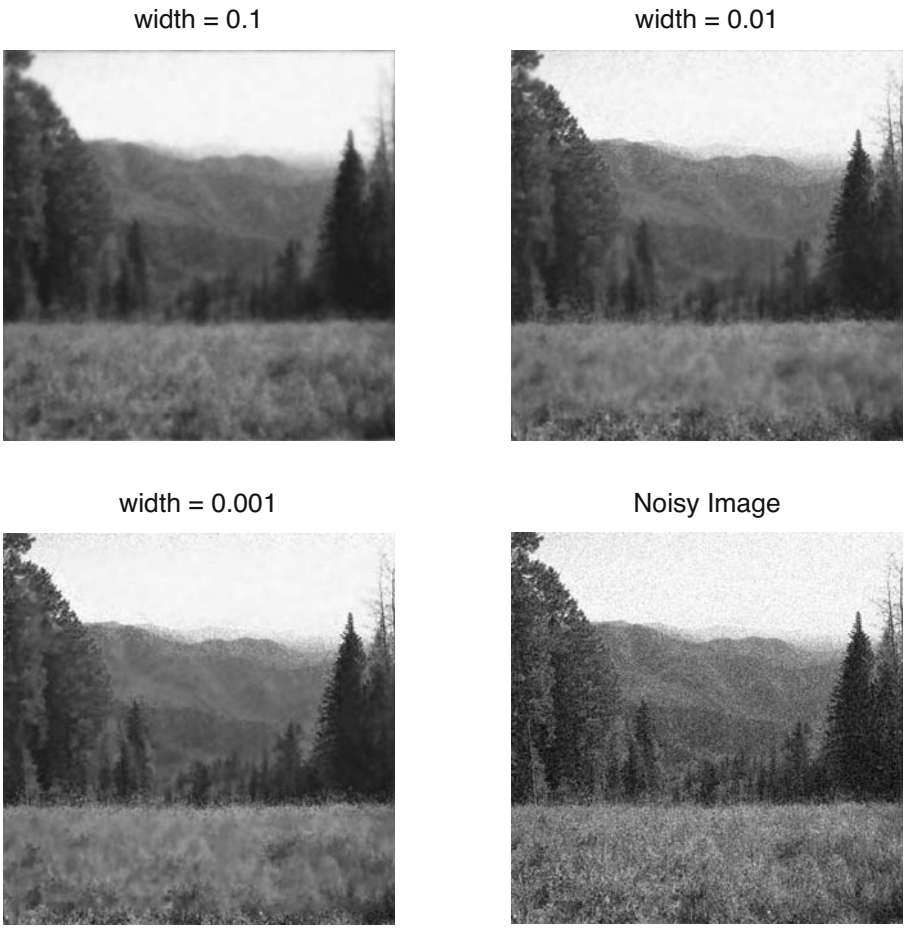

Fig. 11 A study of changing kernel widths as applied to a noisy image of the Jemez Mountains New Mexico, near Los Alamos National Laboratory. As with the image of Lena in Fig. 10, which also has fine scale structures which compete with the denoising algorithms attempt to remove noise, changing the kernel clearly emphasizes structure scale versus noise scale trade-off as described herein

and then integrate by parts to eliminate dependence on derivatives of u. The precise definition we will use is

$$F(u) = \sup_{v,\sigma} \int \left(\nabla \cdot (\tilde{p} v^{\tilde{p}-1} \sigma) u - (\tilde{p} - 1) v^{\tilde{p}}\right), \tag{31}$$

where v ranges over the set of $C^1(\overline{\Omega})$ functions with positive minimum, and σ ranges over functions in $\left(C_c^1(\Omega)\right)^m$ with $|\sigma| \leq 1$.

We will show that if u and $\tilde{p}$ are C^1, then equations (22) and (31) agree.

Note that for $1 \leq p \leq 2$ and $x \geq 0$,

$$x \leq x^p + \frac{1}{4}. \tag{32}$$

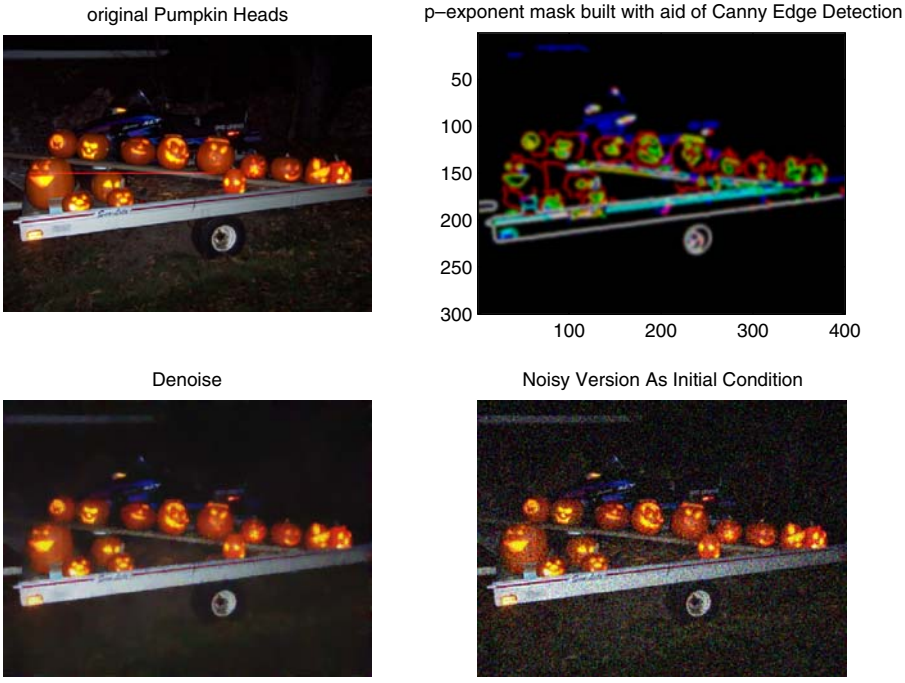

Fig. 12 The exponent $p(x)$ is pre-computed based on the popular Canny edge detection scheme. In the *upper right*, the edges detected from the noisy image in the *lower left*, are somewhat different in each of the three color planes. This lends to a different balancing of diffusion and total-variation-based smoothing in each of the three color planes, yielding a more natural looking final denoised image, with a better three-dimensional appearance, shown in lower left. The *white horizontal line* through the middle of the clean image in the upper right clean image shows in which one-dimensional slice image Fig. 13 is made

Therefore, for either $\tilde{p} = p(u)$ or $\tilde{p} = p(d)$,

$$TV(u) = \int |\nabla u| \leq \int |\nabla u|^{\tilde{p}} + \frac{1}{4}\mathrm{Vol}(\Omega). \tag{33}$$

This inequality holds in the weak formulation as well: choose $v = \tilde{p}^{-1/(\tilde{p}-1)}$. Then

$$\nabla \cdot \left(\tilde{p} v^{\tilde{p}-1}\sigma\right) u - (\tilde{p}-1)v^{\tilde{p}} = (\nabla \cdot \sigma)u - \left[\left(\tilde{p}^{-1/(\tilde{p}-1)}\right) - \left(\tilde{p}^{-1/(\tilde{p}-1)}\right)^{\tilde{p}}\right] \geq (\nabla \cdot \sigma)u - \frac{1}{4}. \tag{34}$$

Integrating and taking the supremum gives

$$F(u) \geq TV(u) - \frac{1}{4}\mathrm{Vol}(\Omega), \tag{35}$$

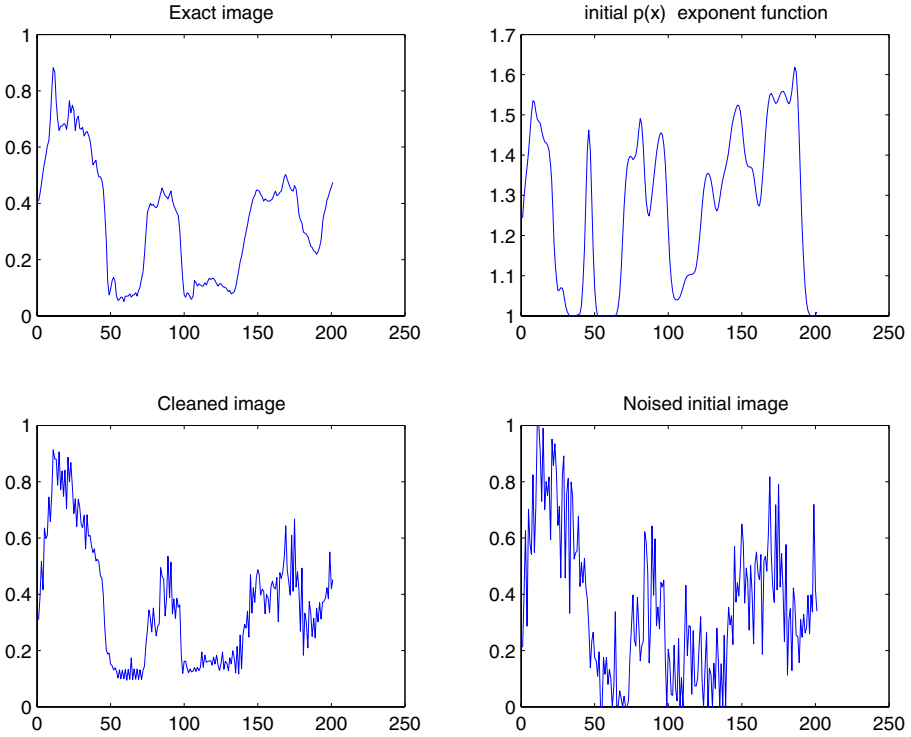

Fig. 13 A one-dimensional slice of the Pumpkins image Fig. 12 is made, at the position of the *white line* shown in Fig. 12

since the total variation is defined weakly [1] as

$$TV(u) = \sup \left\{ \int_\Omega u \nabla \cdot \sigma : \sigma \in C_c^1(\Omega), |\sigma| \leq 1 \right\}. \tag{36}$$

4.1 The weak-derivative formulation of $F(u)$

In this section we show that when u and p are smooth functions, then the supremum in (31) is $\int |\nabla u|^{\tilde{p}}(x)$. The advantage, of course, is that (31) is defined on a larger class of functions. Let u, $\tilde{p}$ be in $C^1(\Omega)$ with $1 \leq \tilde{p} \leq 2$. Also let ν and σ be functions with $\nu(x) > 0$ and $|\sigma(x)| \leq 1$ for all x. Fix $x \in \Omega$; then (29) shows that

$$|\nabla u(x)|^{\tilde{p}(x)} \geq \tilde{p}(x)\nu(x)^{\tilde{p}(x)-1}\sigma(x) \cdot \nabla u(x) - (\tilde{p}(x) - 1)\nu(x)^{\tilde{p}(x)}, \tag{37}$$

and therefore for C^1 functions,

$$\int_\Omega |\nabla u|^{\tilde{p}} \geq \sup \int \left(\nabla \cdot (\tilde{p}\nu^{\tilde{p}-1}\sigma)u - (\tilde{p} - 1)\nu^{\tilde{p}}\right), \tag{38}$$

since the minus sign from integrating by parts can be absorbed into σ.

Lemma 1 *Let $u \in C^1(\overline{\Omega})$. Let $\epsilon > 0$. Then there is a $\sigma \in \left(C_c^1(\Omega)\right)^m$ such that $|\sigma| \leq 1$ and*

$$\left| \int |\nabla u|^{\tilde{p}} - \int \left(\tilde{p} |\nabla u|^{\tilde{p}-1} \sigma \cdot \nabla u - (\tilde{p}-1) |\nabla u|^{\tilde{p}} \right) \right| < \epsilon. \tag{39}$$

Proof Put $K = \max\{|\nabla u| : u \in \overline{\Omega}\}$. If $|\nabla u| \leq 1$, then $|\nabla u|^p \leq 1$ for any $1 \leq p \leq 2$. If $|\nabla u| > 1$, then $|\nabla u|^p \leq K^2$ for any p between 1 and 2. In either event $|\nabla u|^p \leq K^2 + 1$.

Let $w(x) = \operatorname{sgn}(\nabla u(x))$. Then

$$\int |\nabla u|^{\tilde{p}} = \int |\nabla u|^{\tilde{p}-1} w \cdot \nabla u. \tag{40}$$

Choose σ to be a C_c^1 function such that $\|w - \sigma\|_1 < \dfrac{\epsilon}{2(K^2+1)}$. Then

$$\left| \int |\nabla u|^{\tilde{p}} - \int \left(\tilde{p} |\nabla u|^{\tilde{p}-1} \sigma \cdot \nabla u - (\tilde{p}-1) |\nabla u|^{\tilde{p}} \right) \right| = \left| \int \tilde{p} |\nabla u|^{\tilde{p}-1} (\sigma - w) \cdot \nabla u \right| \tag{41}$$

$$\leq 2(K^2+1) \|w - \sigma\|_1 < \epsilon. \tag{42}$$

□

Lemma 2 *Let $u \in C^1(\overline{\Omega})$. Let $\epsilon > 0$. Then there is $\sigma \in \left(C_c^1(\Omega)\right)^m$ and $v \in C^1(\overline{\Omega})$ such that σ, v satisfy $|\sigma| \leq 1$, $v > 0$ and*

$$\left| \int |\nabla u|^{\tilde{p}} - \int \left(\tilde{p} v^{\tilde{p}-1} \sigma \cdot \nabla u - (\tilde{p}-1) v^{\tilde{p}} \right) \right| < \epsilon. \tag{43}$$

Proof Let K and σ be as in the previous lemma. Since $0 \leq |\nabla u| \leq K$ on $\overline{\Omega}$, there is a sequence (ρ_n) of C^1 functions satisfying $0 \leq \rho_n \leq K$ converging to $|\nabla u|$. Put $v_n = \rho_n + 1/n$; then each v_n is also C^1, (v_n) converges pointwise to $|\nabla u|$, and each $v_n > 0$.

The v_n are uniformly bounded by $K+1$ and converge pointwise to $|\nabla u|$, so $\tilde{p} v_n^{\tilde{p}-1} \sigma \cdot \nabla u - (\tilde{p}-1) v_n$ converges pointwise to $\tilde{p} |\nabla u|^{\tilde{p}-1} \sigma \cdot \nabla u - (\tilde{p}-1) |\nabla u|$ and is uniformly bounded on the bounded set $\overline{\Omega}$. Therefore the integral converges to the integral of the limit, so there is n such that

$$\left| \int \left(\tilde{p} v_n^{\tilde{p}-1} \sigma \cdot \nabla u - (\tilde{p}-1) v_n^{\tilde{p}} \right) - \int \left(\tilde{p} |\nabla u|^{\tilde{p}-1} \sigma \cdot \nabla u - (\tilde{p}-1) |\nabla u|^{\tilde{p}} \right) \right| < \epsilon. \tag{44}$$

Combining this with the result of the previous lemma gives the desired result.

□

4.2 On existence of a minimizer

We will prove that the functional J defined in (21) has a minimizer in either of the two cases of $\tilde{p}(x)$ given in (25) or (26), and in either of the two cases of Q given in (23) or (24).

For $u \in L^1(\Omega)$, define $p(u) = p_0(|\nabla(G_\delta * u)|^2)$ where the properties of p_0 and G_δ are as at the beginning of Section 4.

Let $H \subset C^1(\Omega)$ be the subset of functions $\tilde{p}$ with $1 \le \tilde{p}(x) \le 2$.

Lemma 3 *The map $u \mapsto |\nabla(G_\delta * u)|^2$ maps $L^1(\Omega)$ to $C^1(\overline{\Omega})$ continuously.*

Proof Note that

$$\big(\partial_j(G_\delta * u_2)\big)^2 - \big(\partial_j(G_\delta * u_1)\big)^2 = \big((\partial_j G_\delta) * (u_2 + u_1)\big)\big((\partial_j G_\delta) * (u_2 - u_1)\big). \quad (45)$$

Therefore

$$\begin{aligned} &\Big\| \big(\partial_j(G_\delta * u_2)\big)^2 - \big(\partial_j(G_\delta * u_1)\big)^2 \Big\|_\infty \\ &\quad \le \big\| (\partial_j G_\delta) * (u_2 + u_1) \big\|_\infty \big\| (\partial_j G_\delta) * (u_2 - u_1) \big\|_\infty \end{aligned} \quad (46)$$

and

$$\begin{aligned} &\Big\| \partial_i \Big[\big(\partial_j(G_\delta * u_2)\big)^2 - \big(\partial_j(G_\delta * u_1)\big)^2 \Big] \Big\|_\infty \\ &\quad \le \big\| (\partial_i \partial_j G_\delta) * (u_2 + u_1) \big\|_\infty \big\| (\partial_j G_\delta) * (u_2 - u_1) \big\|_\infty \\ &\qquad + \big\| (\partial_j G_\delta) * (u_2 + u_1) \big\|_\infty \big\| (\partial_i \partial_j G_\delta) * (u_2 - u_1) \big\|_\infty . \end{aligned} \quad (47)$$

Suppose $u_1, u_2 \in L^1(\Omega)$. Then for any smooth function K, Young's inequality gives

$$\|K * (u_2 \pm u_1)\|_\infty \le \|K\|_\infty \|u_2 \pm u_1\|_1. \quad (48)$$

(The smoothness requirement on K is only to ensure that the convolution is continuous.)

Applying this inequality to (46) and (47),

$$\Big\| \big(\partial_j(G_\delta) * u_2\big)^2 - \big(\partial_j(G_\delta * u_1)\big)^2 \Big\|_{C^1} \le C \|G_\delta\|_{C^2}^2 \|u_2 + u_1\|_1 \|u_2 - u_1\|_1. \quad (49)$$

Hence for each j, $\big(\partial_j(G_\delta * \cdot)\big)^2$ maps $L^1(\Omega)$ to C^1 continuously. The proof is now complete, since $|\nabla(G_\delta * u)|^2$ is the sum over j of these functions. □

Lemma 4 *The map $s \mapsto p_0 \circ s$ maps $C^1(\overline{\Omega}, \mathbb{R}^+)$ to H continuously.*

Proof Let $s_1, s_2 \in C^1(\overline{\Omega}, \mathbb{R}^+)$. Since p_0 is differentiable, the mean value theorem gives

$$\big\| p_0 \circ s_1 - p_0 \circ s_2 \big\|_\infty \le \|p_0'\|_\infty \|s_2 - s_1\|_\infty. \quad (50)$$

Hence $s \mapsto p_0 \circ s$ maps $C(\overline{\Omega}, \mathbb{R}^+)$ to $C(\overline{\Omega})$ continuously. We also have

$$|\partial_i(p \circ s_2) - \partial_i(p \circ s_1)| = |(p_0' \circ s_2)\partial_i s_2 - (p_0' \circ s_1)\partial_i s_1| \tag{51}$$

$$\leq |p_0' \circ s_2|\,|\partial_i s_2 - \partial_i s_1| + |p_0' \circ s_2 - p_0' \circ s_1||\partial_i s_1| \tag{52}$$

$$\leq |p_0' \circ s_2|\,|\partial_i s_2 - \partial_i s_1| + \|p_0''\|_\infty |s_2 - s_1||\partial_i s_1| \tag{53}$$

$$\leq \|p_0'\|_\infty \|s_2 - s_1\|_{C^1} + \|p_0''\|_\infty \|s_2 - s_1\|_\infty \|s_1\|_{C^1}. \tag{54}$$

Hence $s \mapsto p_0 \circ s$ maps $C^1(\overline{\Omega}, \mathbb{R}^+)$ to $C^1(\overline{\Omega})$ continuously; finally, $p_0 \circ s \in H$ because $1 \leq p_0(x) \leq 2$. □

Lemma 5 *Let $\nu \in C^1(\overline{\Omega})$ such that $\nu > 0$. Then the map $p \mapsto \nu^{p-1}$ maps H to $C^1(\Omega)$ continuously, and H to $W^{1,2}$ continuously.*

Proof Let $K = \max |\log \nu| < \infty$. Then $\nu^{p-1} \leq e^K$ for any $p \in H$. Therefore for any p_1, p_2,

$$|\nu^{p_2-1} - \nu^{p_1-1}| = |\nu^{p_1-1}||\nu^{p_2-p_1} - 1| \tag{55}$$

$$\leq e^K |\exp((p_2 - p_1)\log \nu) - 1| \tag{56}$$

$$\leq \tfrac{1}{K} e^K \left(e^K - 1\right) \|p_2 - p_1\|_\infty, \tag{57}$$

since $|e^x - 1| \leq (e^a - 1)|x|/a$ for $-a \leq x \leq a$, and $|(p_2 - p_1)\log \nu| \leq K$.

Thus

$$\|\nu^{p_2-1} - \nu^{p_1-1}\|_\infty \leq C_1 \|p_2 - p_1\|_\infty, \tag{58}$$

$$\|\nu^{p_2-1} - \nu^{p_1-1}\|_2 \leq C_1 \mathrm{Vol}(\Omega) \|p_2 - p_1\|_\infty, \tag{59}$$

where C_1 depends only on K.

The derivatives satisfy

$$|\partial_i \nu^{p_2-1} - \partial_i \nu^{p_1-1}| \tag{60}$$

$$= |\partial_i((p_2 - 1)\log \nu)\nu^{p_2-1} - \partial_i((p_1 - 1)\log \nu)\nu^{p_1-1}| \tag{61}$$

$$\leq |\partial_i((p_2-1)\log \nu) - \partial_i((p_1-1)\log \nu)||\nu^{p_2-1}| + |\partial_i((p_1-1)\log \nu)||\nu^{p_2-1} - \nu^{p_1-1}| \tag{62}$$

$$\leq e^K |\partial_i((p_2-1)\log \nu) - \partial_i((p_1-1)\log \nu)| + C_1 |\partial_i((p_1-1)\log \nu)| \cdot \|p_2 - p_1\|_\infty. \tag{63}$$

For the first term,

$$|\partial_i((p_2-1)\log \nu) - \partial_i((p_1-1)\log \nu)| \leq |\partial_i(p_2 - p_1)\log \nu| + \left|(p_2 - p_1)\frac{\partial_i \nu}{\nu}\right| \tag{64}$$

$$\leq \|p_2 - p_1\|_{C^1} K + \|p_2 - p_1\|_\infty \|\nu\|_{C^1} / \min \nu. \tag{65}$$

For the second term,

$$|\partial_i((p_1-1)\log v)| = \left|(p_1-1)\frac{\partial_i v}{v} + (\partial_i p_1)\log v\right| \tag{66}$$

$$\leq \frac{\|v\|_{C^1}}{\min v} + K\|p_1\|_{C^1}. \tag{67}$$

Thus

$$|\partial_i v^{p_2-1} - \partial_i v^{p_1-1}| \leq C_2\|p_2-p_1\|_{C^1} + C_3\|p_1\|_{C^1}\|p_2-p_1\|_\infty \tag{68}$$

where C_2 and C_3 depend only on v. We now have

$$\|v^{p_2-1} - v^{p_1-1}\|_{C^1} \leq (C_1+C_2+C_3\|p_1\|_{C^1})\|p_2-p_1\|_{C^1} \tag{69}$$

$$\|v^{p_2-1} - v^{p_1-1}\|_{W^{1,2}} \leq \mathrm{Vol}(\Omega)(C_1+C_2+C_3\|p_1\|_{C^1})\|p_2-p_1\|_\infty \tag{70}$$

Thus we see that $p \mapsto v^{p-1}$ is continuous in the spaces named. □

Lemma 6 *Let $v : \Omega \to [c,\infty)$ with $c > 0$, and $\sigma : \Omega \to \mathbb{R}^n$ be C^1 functions, and suppose that σ has compact support. Then the map $p \mapsto pv^{p-1}\sigma$ maps H to C^1_c continuously, and to $W^{1,2}$ continuously.*

Proof The Leibniz rule implies that for any f, g, $\|fg\|_{C^1} \leq 2\|f\|_{C^1}\|g\|_{C^1}$, and hence that $\|f_2g_2 - f_1g_1\|_{C^1} \leq 2\|f_2-f_1\|_{C^1}\|g_2\|_{C^1} + 2\|f_1\|_{C^1}\|g_2-g_1\|_{C^1}$. Therefore the product of continuous maps into C^1 is a continuous map into C^1.

Similarly, if $f \in C^1$ and $g \in W^{1,2}$, then $fg \in W^{1,2}$; and if $f_1, f_2 \in C^1$ and g_1, g_2 are in $W^{1,2}$, then

$$\|f_2g_2 - f_1g_1\|_2 \leq \|f_2-f_1\|_\infty\|g_2\|_2 + \|f_1\|_\infty\|g_2-g_1\|_2. \tag{71}$$

Also,

$$\begin{aligned}\|\partial_i(f_2g_2 - f_1g_1)\|_2 &\leq \|\partial_i(f_2-f_1)\|_\infty\|g_2\|_2 + \|f_2-f_1\|_\infty\|\partial_i g_2\|_2 \\ &\quad + \|\partial_i f_1\|_\infty\|g_2-g_1\|_2 + \|f_1\|_\infty\|\partial_i(g_2-g_1)\|_2 \\ &\leq 2\big(\|f_2-f_1\|_{C^1}\|g_2\|_{W^{1,2}} + \|f_1\|_{C^1}\|g_2-g_1\|_{W^{1,2}}\big). \end{aligned}\tag{72}$$

Consequently, if f, g are continuous maps from some function space into C^1 and $W^{1,2}$, then fg is a continuous map into $W^{1,2}$. The result now follows because $p \mapsto p\sigma$ is a continuous map into C^1, and $p \mapsto v^{p-1}$ is a continuous map into $W^{1,2}$. □

Lemma 7 *Let v and σ be as in the previous lemma. Then the map $p \mapsto (p-1)v^p$ maps H to $L^1(\Omega)$ continuously.*

Proof Let $K = \max |\log \nu| < \infty$. Fix $p_1 \in H$ and consider $p_2 \in H$ close to p_1. Then the procedure used in (55)–(57) gives us

$$|\nu^{p_2} - \nu^{p_1}| \leq \frac{1}{K} e^{2K} \left(e^K - 1\right) \|p_2 - p_1\|_\infty, \tag{73}$$

where in this instance we have $|\nu^p| \leq (e^K)^p \leq e^{2K}$. Therefore

$$\|\nu^{p_2} - \nu^{p_1}\|_1 \leq \mathrm{Vol}(\Omega) \frac{1}{K} e^{2K} \left(e^K - 1\right) \|p_2 - p_1\|_\infty, \tag{74}$$

which approaches zero as $p_2 \to p_1$ in C^1. □

We can now prove the following theorems:

Theorem 8 *Let $\delta > 0$, $d \in L^1(\Omega)$. Let $p_0 : [0, \infty) \to [1, 2]$ be a nonincreasing C^2 function such that $p_0(M) = 1$ for some M. Define $p(d) = p_0 |\nabla(G_\delta * d)|^2$. Then F_d defined by*

$$F_d(u) = \sup_{(\nu,\sigma)\in E} \int_\Omega \left(\nabla \cdot (p(d)\nu^{p(d)-1}\sigma)u - (p(d) - 1)\nu^{p(d)}\right), \tag{75}$$

where

$$E = \left\{(\nu, \sigma) : \nu \in C_c^1(\Omega), \nu > 0, \sigma \in \left(C_c^1(\Omega)\right)^m), |\sigma| \leq 1\right\}, \tag{76}$$

is a lower semicontinuous function on $L^1(\Omega)$.

Proof Let $u, v \in L^1(\Omega)$. Fix $(\nu, \sigma) \in E$. Then $p(d) \in H$, so $\nabla \cdot (p(d)\nu^{p(d)-1}\sigma)$ is continuous by Lemmas 3, 4, and 6. Therefore

$$\begin{aligned} &\left| \int_\Omega \left(\nabla \cdot \left(p(d)\nu^{p(d)-1}\sigma\right)u - (p(d) - 1)\nu^{p(d)}\right)\right. \\ &\qquad \left. - \int_\Omega \left(\nabla \cdot \left(p(d)\nu^{p(d)-1}\sigma\right)v - (p(d) - 1)\nu^{p(d)}\right)\right| \qquad (77) \\ &\leq \|\nabla \cdot \left(p(d)\nu^{p(d)-1}\sigma\right)\|_\infty \|u - v\|_1. \qquad (78) \end{aligned}$$

Thus, the map

$$u \mapsto \int_\Omega \left(\nabla \cdot \left(p(d)\nu^{p(d)-1}\sigma\right)u - (p(d) - 1)\nu^{p(d)}\right) \tag{79}$$

is continuous on L^1. Since $F_d(u)$ is the supremum of a family of continuous maps, it is lower semicontinuous on L^1. □

Theorem 9 *Let δ, p_0, and E be as in Theorem 8; for $u \in L^1(\Omega)$, define $p(u) = p_0 |\nabla(G_\delta * u)|^2$. Then the function F given by*

$$F(u) = \sup_{(\nu,\sigma)\in E} \int_\Omega \left(\nabla \cdot \left(p(u)\nu^{p(u)-1}\sigma\right)u - (p(u) - 1)\nu^{p(u)}\right) \tag{80}$$

is a lower semicontinuous function on $L^1(\Omega)$.

Proof The outline of the proof is the same as that of Theorem 8, but there are additional details. Let $u \in L^1$. Fix $(v, \sigma) \in E$. Then $\nabla \cdot \left(p(u)v^{p(u)-1}\right)$ is continuous by Lemmas 3, 4, and 6. Given $\epsilon > 0$ we can choose v L^1-close enough to u that

$$\left\|\nabla \cdot \left(p(u)v^{p(u)-1}\sigma\right) - \nabla \cdot \left(p(v)v^{p(v)-1}\sigma\right)\right\|_\infty < \epsilon \tag{81}$$

by Lemmas 3, 4, and 6, and

$$\left\|(p(u)-1)v^{p(u)} - (p(v)-1)v^{p(v)}\right\|_1 < \epsilon \tag{82}$$

by Lemmas 3, 4, and 7. Then

$$\begin{aligned}
&\left|\int_\Omega \left(\nabla \cdot \left(p(u)v^{p(u)-1}\sigma\right)u - (p(u)-1)v^{p(u)}\right)\right. \\
&\quad \left. - \int_\Omega \left(\nabla \cdot \left(p(v)v^{p(v)-1}\sigma\right)v - (p(v)-1)v^{p(v)}\right)\right| \\
&\quad \le \left\|\nabla \cdot \left(p(u)v^{p(u)-1}\sigma\right)\right\|_\infty \|u - v\|_1 \\
&\qquad + \left\|\nabla \cdot \left(p(u)v^{p(u)-1}\sigma\right) - \nabla \cdot \left(p(v)v^{p(v)-1}\sigma\right)\right\|_\infty \|v\|_1 \\
&\qquad + \left\|(p(u)-1)v^{p(u)-1} - (p(v)-1)v^{p(v)-1}\right\|_1 \\
&\quad \le K(u)\|u - v\|_1 + \epsilon\left(\|u - v\|_1 + \|u\|_1\right) + \epsilon,
\end{aligned} \tag{83}$$

where $K(u)$ and $\|u\|_1$ are bounded quantities for any fixed $u \in L^1$. The continuity of

$$u \mapsto \int \left(\nabla \cdot \left(p(u)v^{p(u)-1}\sigma\right)u - (p(u)-1)v^{p(u)}\right) \tag{84}$$

for fixed v, σ now follows, and from that fact we get the lower semicontinuity of F. □

Theorem 10 *Let $d \in L^1(\Omega)$, and let F_d and F be as in Theorems* 8 *and* 9. *For $u \in L^1(\Omega)$ let $J_d(u) = F_d(u) + \|u - d\|_1$ and $J(u) = F(u) + \|u - d\|_1$. Then J_d and J have minimizers in $L^1(\Omega)$. The minimizers need not be unique.*

Proof By (35), J is bounded below. Let (u_n) be a minimizing sequence for J. Also by (35),

$$J(u_n) = \int_\Omega |\nabla u_n|^{p(u)} + \|u_n - d\|_1 \ge TV(u_n) - \frac{1}{4}\mathrm{Vol}(\Omega) + \|u_n\|_1 - \|d\|_1. \tag{85}$$

The tail of the sequence $J(u_n)$ is bounded; therefore, so are the tails of the sequences $TV(u_n)$ and $\|u_n\|_1$. Hence the minimizing sequence is bounded in BV. Since BV is precompact in L^1 ([1], Theorem 3.23), there is a minimizing sequence that converges in L^1. By Theorem 9, F, and hence J, is lower semicontinuous on L^1. Therefore the limit of the convergent minimizing sequence is a minimizer.

The proof for J_d is exactly the same.

The example we will use to show nonuniqueness is taken from [4]. For $r > 0$, let $d = \chi_{B(0,r)}$, and consider $u = cd$ for $c \in [0, 1]$. For a chosen M, if δ is sufficiently small, we will have both $p(d)(x) = 1$ and $p(u)(x) = 1$ for $|x|$ near r. Since $\nabla u(x) = 0$ for $|x| \neq r$, we have that $F_d(u) = F(u) = TV(u) = 2\pi rc$. Then $J_d(u) = J(u) = 2\pi rc + \lambda\pi r^2(1 - c)$. If $\lambda = 2/r$, we have that $J_d(u) = J(u) = 2\pi r$ for all c. As explained in [4], any minimizer u must be of this form, so we have that u is a minimizer for all c. □

We also want to show that there exist minimizers as functionals on $L^2(\Omega)$. However, the proof above does not work for L^2, because BV is not precompact in L^2. However, L^2 is reflexive, so we may use the boundedness of the minimizing sequence to obtain a weakly convergent subsequence.

Theorem 11 *Let $d \in L^2(\Omega)$ be given; define E, $p(d)$ as in Theorem 8. Then*

$$F_d(u) = \sup_{(v,\sigma)\in E} \int_\Omega \left(\nabla \cdot \left(p(d)v^{p(d)-1}\sigma\right)u - (p(d) - 1)v^{p(d)}\right) \tag{86}$$

is weakly lower semicontinuous on $L^2(\Omega)$, and therefore

$$J_d(u) = F_d(u) + \|u - d\|_2^2 \tag{87}$$

has a minimizer. The minimizer is unique.

Proof Suppose $u_n \rightharpoonup u$ weakly in L^2. Fix $(v, \sigma) \in E$. Since $\nabla \cdot \left(p(d)v^{p(d)-1}\sigma\right)$ is in L^2,

$$\int \nabla \cdot \left(p(d)v^{p(d)-1}\sigma\right)u_n \to \int \nabla \cdot \left(p(d)v^{p(d)-1}\sigma\right)u. \tag{88}$$

Also, $(p(d) - 1)v^{p(d)} \in L^1$, so

$$\int \left(\nabla \cdot \left(p(d)v^{p(d)-1}\sigma\right)u - (p(d)-1)v^{p(d)}\right) = \lim \int \left(\nabla \cdot \left(p(d)v^{p(d)-1}\sigma\right)u_n - (p(d)-1)v^{p(d)}\right) \leq \underline{\lim}\, F_d(u_n). \tag{89}$$

Weak lower semicontinuity now follows from taking suprema over v and σ.

Let u_n be a minimizing sequence for $J_d(u)$. Then the tail of u_n is bounded in L^2, so u_n has a subsequence that weakly converges, say to u_0. Since F_d and $\|\cdot - d\|_2^2$ are weakly lower semicontinuous, J_d is weakly lower semicontinuous as well. Therefore u_0 is a minimzer of J_d.

To show uniqueness, observe from (86) that $F_d(u)$ is a supremum of affine functions of u. Thus F_d is convex. Since $u \mapsto \|u - d|_2^2$ is strictly convex, J_d is strictly convex. Therefore its minimizer is unique. □

Theorem 12 *Define E, p(u)as in Theorem 9. The functional F defined by*

$$F(u) = \sup_{(v,\sigma)\in E} \int_\Omega \left(\nabla \cdot \left(p(u)v^{p(u)-1}\sigma\right)u - (p(u)-1)v^{p(u)}\right) \tag{90}$$

is weakly lower semicontinuous on $L^2(\Omega)$, and therefore

$$J(u) = F(u) + \|u-d\|_2^2 \tag{91}$$

has a minimizer.

Proof Let $u_n \rightharpoonup u$ weakly in $L^2(\Omega)$. The boundary of Ω is a set of measure zero, so we can treat u_n, u as elements of $L^2(\overline{\Omega})$. Since $\|u_n\|_2$ is bounded, say by M, we have the estimates

$$\|\partial_i G_\delta * u_n\|_\infty \le \|\partial_i G_\delta\|_2 \|u_n\|_2 \le M\|G_\delta\|_{W^{1,2}} \tag{92}$$

so that $\{\partial_i G_\delta * u_n\}$ is uniformly bounded; and

$$\|\partial_j\partial_i G_\delta * u_n\|_\infty \le \|\partial_j\partial_i G_\delta\|_2 \|u_n\|_2 \le M\|G_\delta\|_{W^{2,2}} \tag{93}$$

so that $\{\partial_i G_\delta * u_n\}$ is equicontinuous.

Now for each $x \in \overline{\Omega}$, the functional $f \mapsto (\partial_i G_\delta * f)(x)$ is a bounded linear functional on L^2, so that the weak convergence of u_n gives that $\partial_i G_\delta * u_n \to \partial_i G_\delta * u$ pointwise.

Since the sequence $(\partial_i G_\delta * u_n)$ is equicontinuous, the convergence is in fact uniform. The same argument applied to the second and third derivatives of G_δ shows that the convergence is in C^1.

Lemma 4 now shows that $p(u_n)$ converges to $p(u)$ in H. Fix $(v,\sigma) \in E$. Then Lemma 6 shows that $\nabla \cdot \left(p(u_n)v^{p(u_n)-1}\sigma\right)$ converges strongly in L^2 to $\nabla \cdot \left(p(u)v^{p(u)-1}\sigma\right)$, and Lemma 7 gives that $(p(u_n)-1)v^{p(u_n)}$ converges strongly in L^1 to $(p(u)-1)v^{p(u)}$. Define $q_n = \nabla \cdot \left(p(u_n)v^{p(u_n)-1}\sigma\right)$ and $q = \nabla \cdot \left(p(u)v^{p(u)-1}\sigma\right)$. Then $q_n \to q$ in L^2. Hence

$$\left|\int q_n u_n - \int qu\right| \le \|q_n - q\|_2 \|u_n\|_2 + \left|\int q(u_n - u)\right| \to 0, \tag{94}$$

due to the weak convergence of (u_n) to u. Then if we let

$$f_{v,\sigma}(v) = \int \left(\nabla \cdot \left(p(v)v^{p(v)-1}\sigma\right)v - (p(v)-1)v^{p(v)}\right), \tag{95}$$

we have

$$f_{v,\sigma}(u) = \lim f_{v,\sigma}(u_n) \le \underline{\lim}\, F(u_n). \tag{96}$$

Taking the supremum over $(v,\sigma) \in E$, we see that F is weakly lower semicontinuous.

The weak lower semicontinuity of J and the existence of a minimizer follows in exactly the same way as in Theorem 11.

5 Conclusions

We have discussed trade-offs of spatial scales, noise intensity, and denoising strength in terms of choosing a interpolate between total-variation denoising and isotropic diffusion denoising, motivated by earlier work of Mulet, Chan and Wong [2]. We have proven here existence theorems for such methods, and we have shown appropriate theorems regarding uniqueness of such minimizers. While we have presented numerical algorithms, and examples for our methods here, this remains an area where further research should be able to continue to improve the practical implementation of such methods.

Acknowledgements The work of the third and fifth authors was funded by the Department of Energy under contract W-7405-ENG-36. The work of the fourth author was supported by NSF grant DMS-0410085.

References

1. Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems. Oxford University Press (2000)
2. Blomgren, P., Chan, T.F., Mulet, P., Vese, L., Wan, W.L.: Variational PDE models and methods for image processing. In: Numerical Analysis 1999 (Dundee), pp. 43–67. Chapman & Hall, Boca Raton, FL (2000)
3. Chambolle, A.: An algorithm for total variation minimization and applications. J. Math. Imaging Vision **20**, 89–97 (2004)
4. Chan, T.F., Esedoğlu, S.: Aspects of total variation regularized L^1 function approximation. SIAM J. Appl. Math. **65**(5), 1817–1837 (2005)
5. Chen, Y., Levine, S., Rao, M.: Variable exponent, linear growth functionals in image restoration. SIAM J. Appl. Math. **66**, 1383–1406 (2006)
6. Dacarogna, B.: Introduction to the Calculus of Variations. Imperial College Press, London (2004)
7. Donoho, D.L., Johnstone, I.M.: Ideal spatial adaptation by wavelet shrinkage. Biometrika **81**, 425–455 (1994)
8. Ekeland, I., Témam, R.: Convex Analysis and Variational Problems. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (1999)
9. Evans, L.C., Gariepy, R.F.: Measure Theory and Fine Properties of Functions. CRC Press (1992)
10. Folland, G.B.: Real Analysis, 2nd edn. John Wiley & Sons Inc., New York (1999)
11. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. **6**, 721–741 (1984)
12. Levine, S., Chen, Y., Stanich, J.: Image restoration via nonstandard diffusion. Technical Report 04-01, Department of Mathematics and Computer Science, Duquesne University, Pittsburgh (2004)
13. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D **60**, 259–268 (1992)