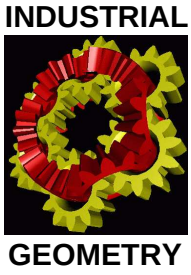
National Research Network S92

Industrial Geometry

http://www.industrial-geometry.at



NRN Report No. 81

Locally Adaptive Total Variation Regularization

Markus Grasmair

February 2009

FWF
Der Wissenschaftsfonds.

universität innsbruck

Locally Adaptive Total Variation Regularization

Markus Grasmair

Department of Mathematics, University of Innsbruck,
Technikerstr. 21a, A-6020 Innsbruck, Austria,
Markus.Grasmair@uibk.ac.at,
WWW home page: http://infmath.uibk.ac.at

Abstract. We introduce a locally adaptive parameter selection method for total variation regularization applied to image denoising. The algorithm iteratively updates the regularization parameter depending on the local smoothness of the outcome of the previous smoothing step. In addition, we propose an anisotropic total variation regularization step for edge enhancement. Test examples demonstrate the capability of our method to deal with varying, unknown noise levels.

1 Introduction

Because of its ability to generate images with piecewise smooth structures that are well separated by pronounced edges, total variation regularization is one of the most widely used techniques for image denoising and related tasks. Since the first proposal by Rudin, Osher, and Fatemi [14] of using the total variation for denoising purposes, that is, the L^1-norm of the gradient, this method has been applied to a wide range of applications in imaging and inverse problems. We refer to [1–3, 5, 12, 13, 15] to name but a few contributions to this field.

Given a noisy function $f \in L^2(\Omega)$ on some open and bounded domain $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}$, the goal of denoising is to find a new function u close to f that retains the important features of f while noise, consisting of fast oscillations, is removed. Noting that edges belong to the most prominent features in images, this task can be achieved by minimizing the total variation functional

$$\mathcal{T}(u;\alpha) := \frac{1}{2}\int_\Omega \big(u(x) - f(x)\big)^2\,dx + \alpha|Du|(\Omega) \tag{1}$$

with respect to $u \in BV(\Omega)$. The regularization parameter $\alpha > 0$ in (1) controls the amount of smoothing that is desired: the larger α, the more the regularized function u_α tends to consist of well separated homogeneous regions. Conversely, a small parameter α implies a function lying close to the input data, but also possibly exhibiting a significant number of oscillations.

The relation between α and u_α, however, exists only on a qualitative level. There is no simple connection between the value of α and the smoothness of u_α, or even between α and the difference $f - u_\alpha$, which is simply the part of the data classified as noise by the functional $\mathcal{T}$. The necessity of taking into

account both the data and the expected noise level is a well established fact in the theory of inverse problems (see for instance [8]). Because for many applications of mathematical imaging, in particular tasks that are to be completely automated, a precise knowledge of the noise is not available, this leads to the conclusion that, in these cases, a-priori parameter choices are not feasible. Instead, one should adapt α until both u_α and the perceived noise $f - u_\alpha$ are satisfactory.

Though better than a fixed a-priori choice, also adaptation of the regularization parameter need not be sufficient for good results. It may happen that the noise on the image f is not identically distributed but varies locally. In this case, it is difficult to find a compromise between oversmoothing in noise-free regions caused by too large a parameter choice, and a still noisy output resulting from a small parameter. Similar effects can be observed, if the structure of the noise-free data itself changes over the image. Then, the regularization parameter should be larger for homogeneous parts of the image than for parts with small details.

The problem of finding a parameter that is suited for the whole image can be circumvented by passing from a global parameter $\alpha > 0$ to a parameter function $\alpha\colon \Omega \to \mathbb{R}_{>0}$. Then, the regularization functional reads as

$$\mathcal{T}(u;\alpha) = \frac{1}{2}\int_\Omega \big(u(x) - f(x)\big)^2\,dx + \int_\Omega \alpha(x)\,d|Du|(x)\ . \tag{2}$$

This functional is well-defined, if α is continuous, and, using direct methods, can readily be shown to attain a minimizer, if α is bounded away from zero.

Total variation regularization with non-constant regularization parameter has already been studied in several other articles [6, 9–11, 16, 17]. In [16, 17], the choice of α is based on the scale of the features one wants to recover. In [10], at first the uniform problem is solved with an automatically identified optimal regularization parameter α. The result of the first denosing attempt is used for extracting the edges in the image, at which subsequently the regularization parameter is locally increased. Then the minimization problem is solved a second time with the localized parameter $\alpha(x)$.

The approach in [11] uses statistical properties of the residual in order to decide whether the local regularization parameter is suited. The criterion employed there is based on the local variance of the residual: If it is close to the noise level, one can expect that mostly noise has been filtered. It it is higher, then the residual probably contains texture and therefore the regularization parameter has to be decreased. The estimates in [11] are closely related to the inequalities in [10], though the approaches by which they are reached differ considerably. Note moreover that the same idea has already been employed in [6] for one-dimensional total variation regularization.

In this paper, we propose to target some a-priori specified smoothness of the output u_α, which is measured in terms of the oscillations of the direction $\nabla u_\alpha/|\nabla u_\alpha|$ of the gradient of the image. This direction can be determined by passing to a dual formulation, as it essentially equals the rescaled dual variable. This idea of parameter adaptation based on the properties of the dual function is taken from [6].

The main concept of this paper of using a dual variable to provide a guess on the smoothness of the regularized image is introduced in Section 2. For further improving this smoothed image by enhancing the edges, we propose to subsequently apply anisotropic total variation regularization with an anisotropy that is estimated from the same dual variable that has determined the isotropic regularization parameter (see Section 3). A complete description of the algorithm can be found in Section 4. Finally, we apply this method in Section 5 to two test examples that show its suitability for adaptive noise removal.

2 Parameter Adaptation via Dual Variables

Consider the dual formulation of $\mathcal{T}(\cdot;\alpha)$, which consists in solving the constrained minimization problem

$$\begin{aligned} \mathcal{J}(V) &:= \int_\Omega \big(\operatorname{div} V(x) + f(x)\big)^2 \, dx \to \min, \\ |V(x)| &\le \alpha(x) \text{ almost everywhere on } \Omega, \\ V(x)\cdot\nu(x) &= 0 \text{ almost everywhere on } \partial\Omega, \end{aligned} \tag{3}$$

over the space of vector valued essentially bounded functions $L^\infty(\Omega;\mathbb{R}^n)$. In (3), ν denotes the outward normal to the domain Ω, and the equation $V\cdot\nu = 0$ is understood in a distributional sense. Also, the divergence of an essentially bounded function is defined distributionally. To be precise, the functions V and $\operatorname{div} V$ satisfy the equation

$$\int_\Omega \nabla\phi(x)\cdot V(x)\,dx = -\int_\Omega \phi(x)\ \operatorname{div} V(x)\,dx$$

for every $\phi \in C^1(\mathbb{R}^n)$.

Minimization of $\mathcal{T}_\alpha$ is equivalent to solving the dual problem (3) in the sense that a function $V_\alpha \in L^\infty(\Omega;\mathbb{R}^n)$ solves (3), if and only if $u_\alpha := f + \operatorname{div} V_\alpha$ minimizes $\mathcal{T}_\alpha$. We refer to [4], which treats the dual formulation of total variation regularization, and to [7] for a detailed introduction to infinite dimensional convex analysis.

We now examine the dual variable V more closely. Formally, the optimality condition for a minimizer u_α of the functional $\mathcal{T}$ reads as

$$u_\alpha(x) - f(x) \ni \operatorname{div}\left(\alpha(x)\frac{\nabla u_\alpha(x)}{|\nabla u_\alpha(x)|}\right) \qquad \text{for almost every } x\in\Omega\,.$$

Since $u_\alpha - f = \operatorname{div} V_\alpha$, one sees that the dual minimizer V_α introduced above in fact coincides with the direction of the gradient of u_α, multiplied by $\alpha(x)$. In particular, for almost every $x\in\Omega$, we either have that $|V_\alpha(x)| = \alpha(x)$ or the gradient of u_α at x is zero, that is, u_α is approximately constant near x.

Even more, the local behaviour of V_α is strongly related to a certain kind of regularity of the regularized function u_α: Large variations of V_α/α on the unit

sphere imply equally large variations of the direction of the gradient of u_α. In other words, variations of V_α/α imply small oscillations of u_α. The method we propose in the following takes advantage of these properties of V_α and u_α and exploits their relation.

Let $r > 0$ be some fixed parameter. We define the r-local mean of a vector valued, essentially bounded function $W \in L^\infty(\Omega; \mathbb{R}^n)$ at $x \in \Omega$ by

$$M_r(x; W)(x) := \fint_{B_r(x)\cap\Omega} W(y)\, dy := \frac{1}{\mathcal{L}^n\big(B_r(x) \cap \Omega\big)} \int_{B_r(x)\cap\Omega} W(y)\, dy\,.$$

Here, $\mathcal{L}^n$ denotes the n-dimensional Lebesgue measure. In addition, we define the r-local variation of W by

$$\Sigma_r(x; W)(x) := \big|W(x) - M_r(x; W)\big|\,. \tag{4}$$

The definition of Σ_r directly implies that

$$\big|\Sigma_r(x; W)\big| \le 2\,\operatorname{ess\,sup}\big\{|W(y)| : y \in B_r(x) \cap \Omega\big\}$$

for almost every $x \in \Omega$.

Applying the above inequality to the scaled solution

$$W_\alpha(x) := V_\alpha(x)/\alpha(x)$$

of (3), one immediately sees that

$$0 \le \big|\Sigma_r(x; W_\alpha)\big| \le 2\max\big\{|V_\alpha(y)|/\alpha(y) : y \in B_r(x) \cap \Omega\big\} \le 2\,.$$

Moreover, the actual size of the value $\Sigma_r(x; W_\alpha)$ provides an indication of the oscillation of the function u_α near x: If $\Sigma_r(x; W_\alpha)$ is close to zero, then the gradient ∇u_α points in roughly the same direction on the whole set $B_r(x)$. Conversely, a value above one implies that the orientation of $\nabla u_\alpha(x)$ vastly differs from the majority of directions present in $B_r(x)$. See Figure 1 for an example of a smoothed image with corresponding local variation of the dual variable V_α.

In this manner, the function $\Sigma_r(x; W_\alpha)$ can serve as a local criterion for the smoothness of the regularized function u_α. If the finally desired smoothness is not yet reached, that is, if $\Sigma_r(x; W_\alpha)$ is too large, it is necessary to increase the local regularization parameter $\alpha(x)$. Conversely, if the function u_α appears too smooth, that is, $\Sigma_r(x; W_\alpha)$ is close to zero, then $\alpha(x)$ is decreased and a new tentative solution u_α is computed. This process of computing $\Sigma_r(x; W_\alpha)$ and updating α is repeated until the update of u_α becomes small enough.

In order to reach a uniform smoothness of the regularized image u_α over its whole domain, we propose to prescribe some target smoothness $0 < \theta < 1$. Then one can compute a suitable update $\tilde{\alpha}$ of α setting

$$\tilde{\alpha}(x) = \alpha(x)\big(\theta + \Sigma_r(x; W_\alpha)/2\big)^s \tag{5}$$

Fig. 1. Smoothed image (left) and corresponding function Σ_r (right). Bright pixel values indicate a higher value of Σ_r.

for some parameter $s > 0$ determining the size of the update. Iteration of this update will lead to a uniform smoothness $\Sigma_r(x; W_\alpha) \approx 2(1-\theta)$. The choice of the target smoothness should reflect the properties of the image one wants to recover: A large parameter ($\theta \geq 0.7$) means that only the structures about the size of r are of interest. Small values ($\theta \approx 0.55$) put more emphasis on the structures of size smaller than r (see also Figure 4).

In order to avoid too rapid changes of the parameter $\alpha(x)$, it is necessary smooth the update $\tilde{\alpha}$ computed by means of (5). Also from a theoretical point of view, this smoothing procedure is required for obtaining a continuous regularization function α. We propose to simply replace the update $\tilde{\alpha}(x)$ by its local mean value $M_r(x; \tilde{\alpha})$. In this way, the average smoothness in the balls $B_r(x)$ will be almost independent of x.

3 Edge Enhancement by Anisotropy

Having determined the size of the local regularization parameter $\alpha(x)$ by means of the scaled dual variable W_α, it is in addition possible to use the distribution of the values of W_α on the unit sphere for sharpening edges and, in particular, thin ridges, which usually tend to get oversmoothed. To that end, instead of applying isotropic regularization, we introduce an anisotropy the direction of which is determined by the local covariance of W_α.

For $R > 0$ we define the $\mathbb{R}^{n\times n}$-valued function $\mathrm{Cov}_R(x; W)$, the covariance of W on $B_R(x) \cap \Omega$, by defining its (i,j)-th component as

$$\mathrm{Cov}_R^{(i,j)}(x; W) := \fint_{B_R(x)\cap\Omega} \big(W^{(i)}(y) - M_R^{(i)}(x; W)\big)\big(W^{(j)}(y) - M_R^{(j)}(x; W)\big)\, dy\,. \tag{6}$$

Again using the property that W_α is proportional to ∇u_α, one sees that the principal component of $\mathrm{Cov}_R(x; W_\alpha)$ indicates, up to sign, the prevailing direction of ∇u_α near x.

This dominant direction can be pronounced further by replacing the isotropic bound $|V_\alpha(x)| \le \alpha(x)$ in (3) by an anisotropic one defined by $\mathrm{Cov}_R(x; W_\alpha)$. This is achieved by minimizing $\mathcal{J}(V)$ respecting the constraints $V \cdot \nu = 0$ on $\partial\Omega$ and

$$c(x)\, V(x)^t \,\mathrm{Cov}_R(x; W_\alpha) V(x) \le 1 \qquad \text{on } \Omega\,. \tag{7}$$

Here, the scalar valued function $c\colon \Omega \to \mathbb{R}_{>0}$ has to be chosen in such a way that a similar amount of smoothing is reached as for isotropic regularization with parameter $\alpha(x)$.

For determining a suitable size for c, note that the amount of smoothing induced by the bound (7) can be estimated by the determinant of the matrix $c(x)\,\mathrm{Cov}_R(x; W_\alpha)$, which, for consistency with the constraint $|V(x)| \le \alpha(x)$, should equal $\alpha(x)^{-2n}$. Thus one obtains for the function c the value

$$c(x) = \alpha(x)^{-2} \det\big(\mathrm{Cov}_R(x; W_\alpha)\big)^{-1/n}\,.$$

We therefore propose an edge enhancement via solving the minimization problem

$$\begin{aligned} \mathcal{J}(V) &= \int_\Omega \big(\operatorname{div} V(x) + f(x)\big)^2\, dx \to \min\,, \\ V(x)^t A(x) V(x) &\le 1 \text{ almost everywhere on } \Omega\,, \\ V(x)\cdot\nu(x) &= 0 \text{ almost everywhere on } \partial\Omega\,. \end{aligned} \tag{8}$$

Here

$$A(x) = \alpha(x)^{-2} \det\big(\mathrm{Cov}_R(x; W_\alpha)\big)^{-1/n} \mathrm{Cov}_R(x; W_\alpha)\,,$$

and $W_\alpha = V_\alpha/\alpha$, where V_α is the solution of (3). Denoting the solution of (8) by V_A and defining $u_A := f + \operatorname{div} V_A$, we obtain an enhanced version of the isotropic total variation minimizer u_α.

4 Summary of the Algorithm

We now summarize the method developed in the previous sections for adaptive denoising of a noisy image $f \in L^2(\Omega)$.

Algorithm 1. Set $k = 1$, choose some initial regularization function $\alpha_1\colon \Omega \to \mathbb{R}_{>0}$, a smoothness parameter $0 < \theta < 1$, some $r > 0$, $R > 0$, $s > 0$, and $\varepsilon > 0$.

1. Compute
$$V_k := \arg\min\Big\{\mathcal{J}(V) : |V(x)| \le \alpha_k(x) \text{ on } \Omega,\ V\cdot\nu = 0 \text{ on } \partial\Omega\Big\}\,.$$
2. Define $W_k := V_k/\alpha_k$ and compute $\Sigma_r(x; W_k)$ (see (4)).
3. If $\|V_k - V_{k-1}\| < \varepsilon$ go to 5.
4. Compute
$$\hat\alpha_{k+1}(x) := \alpha_k(x)\big(\theta + \Sigma_r(x; W_\alpha)/2\big)^s$$
and
$$\alpha_{k+1}(x) := M_r(x; \hat\alpha_{k+1})\,,$$
increase k by one, and go to 1.

5. Compute $\mathrm{Cov}_R(x; W_k)$ (see (6)) and

$$A(x) := \alpha(x)^{-2} \det\big(\mathrm{Cov}_R(x; W_k)\big)^{-1/n} \mathrm{Cov}_R(x; W_k) \,.$$

6. Compute

$$V_A := \arg\min\Big\{ \mathcal{J}(V) : V(x)^t A(x) V(x) \le 1 \text{ on } \Omega,\ V \cdot \nu = 0 \text{ on } \partial\Omega \Big\} \,.$$

Define the regularized function $u_A := f + \operatorname{div} V_A$.

In steps 1–4, only the regularization function α is determined. For this, it is not necessary to compute the minimizers of $\mathcal{J}$ precisely. Instead, a reasonable approximation of a minimizer is sufficient to provide a good update of α, at least during the first iterations. In particular if an iterative method is used for the minimization of $\mathcal{J}$, the computation time can be improved by stopping the iteration well before convergence is reached.

In the numerical examples below, the functions V_k and V_A were computed by alternating between gradient descent steps for the minimization of $\mathcal{J}$ and projections of V on the sets $\big\{V : |V(x)| \le \alpha_k(x)\big\}$ and $\big\{V : V(x)^t A(x) V(x) \le 1\big\}$, respectively. The function V_{k-1} was used as initial guess for the computation of V_k.

5 Examples

The algorithm presented in Section 4 is tested by means of two images. The first, synthetic image shows a collection of ellipses and rectangles of different size and intensity (see Figure 2, upper left). These clean data were distorted by normally distributed random noise. In order to illustrate the capability of the algorithm for dealing with varying noise level, the standard deviation of the random noise was chosen to increase towards the right bottom of the image from about 10% of the maximal intensities to 150% (see Figure 2, lower left).

The original image only consisting of simple geometric forms without any texture, it should be perfectly suited for total variation regularization. The changing noise level within the distorted data, however, makes a uniform parameter choice almost impossible: If the regularization parameter is chosen too small, then the noise on the right hand side of the data is barely removed. In particular, the right hand edges of the lower ellipses can hardly be recovered. On the other hand, a too large regularization parameter leads to the disappearance of the small circle at the left hand side of the image (see Figure 2, middle column). Only a very small range of parameters removes the noise reasonably well while still preserving the small scale structure—and even then the contrast deteriorates.

Figure 2, upper right, shows the smoothed image obtained with Algorithm 1. Since the original image is very smooth, the smoothness parameter was chosen rather large as $\theta = 0.85$. The variance Σ_r was evaluated on balls with a radius of 3 pixels, the complete image measuring 256×256 pixels. The lower right image in Figure 2 shows the distribution of the finally chosen regularization function α.

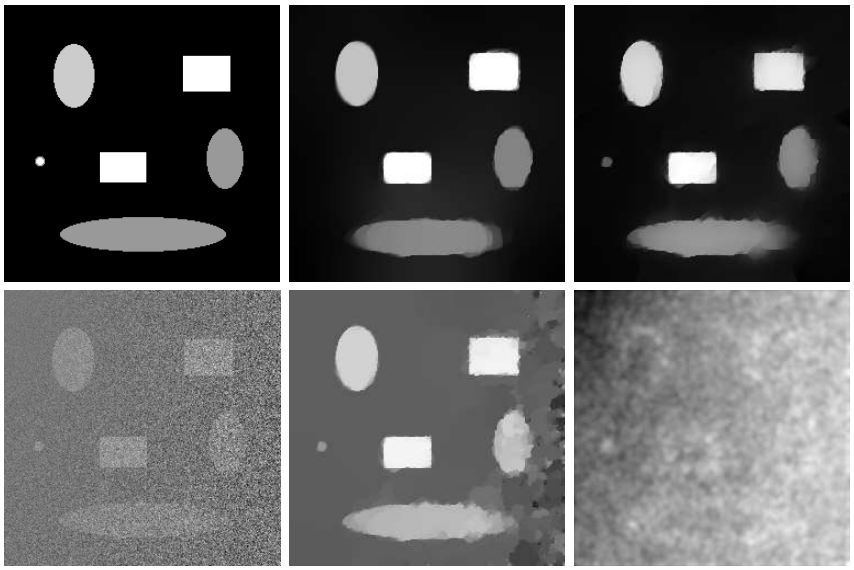


Fig. 2. Left column: original and noisy image; the noise level increases to the right bottom of the image. Middle column: denoising without parameter adaptation; either small details are lost or the smoothing effects are partially insufficient. Right column, upper row: denoised image for smoothness parameter $\theta = 0.85$. Right column, lower row: logarithm of the finally chosen regularization function α; the minima and maxima of α differ by a factor of 12.

As expected, it increases to the right bottom, where more noise is present. Over the whole image, the maxima and minima of α differ by a factor of 12.

One can see in the resulting image that the noise is efficiently removed. Also, the shape of the two lower ellipses is reconstructed in a reasonable way, considering that rather more noise than signal is present in these regions. Moreover, the small circle on the left is clearly visible, though some contrast was lost.

As a second test example, we consider the photographer image. In a first experiment we add different levels of random noise (see Figure 3). The outcome of the adaptive Algorithm 1 (right column) is compared with the solution of standard total variation regularization with constant parameter choice independent of the noise level (middle column). The smoothness parameter for the adaptive algorithm was chosen as $\theta = 0.60$; the regularization parameter for the standard algorithm was selected in such a way that the results for moderate noise level (third row) are comparable.

The results show that, as expected, a constant regularization parameter only yields good results for a very specific noise level. For stronger noise, almost no smoothing is obtained, whereas the image is oversmoothed in case it is already quite clean. In contrast, the adaptive algorithm yields comparable results for different noise levels, and is also able to treat noise-free images (first row).

Fig. 3. Left column: image with Gaussian noise; the noise level increases with each row ($\sigma = 0$, 30, 50, 100). Middle column: total variation denoised image with constant parameter choice. The regularization parameter is kept the same for all images. Right column: denoised images with adaptive parameter choice for a smoothness parameter $\theta = 0.60$.

In order to illustrate the effect of the smoothness parameter, we apply Algorithm 1 to the noise-free photographer image and vary θ (see Figure 4). For a value of $\theta = 0.55$ mainly the grass and details of the camera are smoothed. As

Fig. 4. Influence of the smoothness parameter θ. First row: original image and smoothed images with $\theta = 0.55$ and $\theta = 0.60$. Second row: smoothed images with $\theta = 0.65$, $\theta = 0.70$, and $\theta = 0.80$.

θ increases, more and more details are lost until only the large scale structures in the image remain. Thus, the smoothness parameter works in some sense like the regularization parameter of standard total variation regularization.

There is, however, a notable difference. In the standard method, the time when structures in the image disappear depends on their scale, which is basically the ratio between contrast, that is, the difference of the intensities of the structure and the background, and the perimeter of the structure. As opposed to this, the model presented here puts much less emphasis on the contrast. Low contrast but distinct parts of the image tend to disappear much later than with uniform regularization. Compare for instance the rightmost building in the images of Figure 4 with the outcome of the standard method (Figure 3, first row, middle image).

Finally, Table 1 compares the performance of our algorithm with uniform total variation regularization and the adaptive method from [11]. The regularization parameter for the comparison was chosen in such a way that the norm of the residual equals the norm of the noise. At small noise levels, the texture enhancing method [11] and even uniform regularization perform better. On the other hand, our algorithm provides good results if much noise is present. Note moreover that the here proposed method does not require a guess on the noise level, whereas the other methods do.

Table 1. Comparison between standard TV regularization, the method proposed in [11], and our method for different smoothness parameters. The table provides signal to noise ratios for the photographer image with various levels of Gaussian noise added ($\sigma = 20, 30, 40, 50$).

original	uniform TV	adaptive ([11])	$\theta = 0.55$	$\theta = 0.60$	$\theta = 0.65$
9.47	14.63	15.63	15.30	14.73	13.71
6.31	12.13	13.35	13.28	13.46	12.84
3.86	10.05	11.59	11.47	12.45	12.18
1.93	8.38	10.11	10.00	11.54	11.51

6 Conclusion

We have introduced an algorithm for the local adaptation of the regularization parameter in total variation regularization applied to the task of image denoising. The main idea of the method is to base the parameter choice on the smoothness of the output image, which is measured in terms of the variation of the direction of its gradient. This variation can be obtained when employing a dual method for the actual minimization of the total variation regularization functional. Starting from an initial guess of the regularization function, the proposed algorithm consecutively computes the corresponding minimizer of the total variation functional and updates the regularization function depending on the smoothness of the update. The iteration stops when the update is sufficiently small.

As a post-processing step, we propose to apply an anisotropic regularization method intended to sharpen edges. Again, the regularization is determined by the dual variable. This anisotropic regularization step reduces the contrast loss due to isotropic smoothing and, in particular, is suited for the enhancement of ridges.

The examples presented in Section 5 indicate the suitability of the proposed method for denoising images with unknown, varying noise levels. In particular, they show its ability to provide an estimate for the amount of smoothing required to obtain a certain smoothness of the output.

Acknowledgement

This work has been supported by the Austrian Science Fund (FWF) within the national research network Industrial Geometry, project 9203-N12.

References

1. R. Acar and C. R. Vogel. Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Probl.*, 10(6):1217–1229, 1994.
2. G. Aubert and P. Kornprobst. *Mathematical problems in image processing*, volume 147 of *Applied Mathematical Sciences*. Springer, New York, second edition, 2006. Partial differential equations and the calculus of variations, With a foreword by Olivier Faugeras.

3. M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Probl.*, 20(5):1411–1421, 2004.
4. A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vision*, 20(1–2):89–97, 2004.
5. A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numer. Math.*, 76(2):167–188, 1997.
6. P. L. Davies and A. Kovac. Local extremes, runs, strings and multiresolution. *Ann. Statist.*, 29(1):1–65, 2001.
7. I. Ekeland and R. Temam. *Convex Analysis and Variational Problems*. North-Holland, Amsterdam, 1976.
8. H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.
9. I. A. Frigaard, G. Ngwa, and O. Scherzer. On effective stopping time selection for visco-plastic nonlinear BV diffusion filters used in image denoising. *SIAM J. Appl. Math.*, 63(6):1911–1934 (electronic), 2003.
10. I. A. Frigaard and O. Scherzer. Herschel–Bulkley diffusion filtering: non-Newtonian fluid mechanics in image processing. *Z. Angew. Math. Mech.*, 86(6):474–494, 2006.
11. G. Gilboa, N. Sochen, and Y.Y. Zeevi. Variational denoising of partly-textured images by spatially varying constraints. *IEEE Trans. Image Process.*, 15(8):2281–2289, 2006.
12. K. Ito and K. Kunisch. Augmented Lagrangian methods for nonsmooth, convex optimization in Hilbert spaces. *Nonlinear Anal.*, 41A:591–616, 2000.
13. M. Z. Nashed and O. Scherzer. Least squares and bounded variation regularization with nondifferentiable functional. *Numer. Funct. Anal. Optim.*, 19(7&8):873–901, 1998.
14. L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1–4):259–268, 1992.
15. O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, New York, 2008.
16. D.M. Strong. Adaptive Total Variation Minimizing Image Restoration. CAM Report 97-38, University of California, Los Angeles, 1997.
17. D.M. Strong, J.-F. Aujol, and T.F. Chan. Scale recognition, regularization parameter selection, and Meyer's G norm in total variation regularization. *Multiscale Model. Simul.*, 5(1):273–303 (electronic), 2006.