

Anisotropic Total Variation Filtering

Markus Grasmair & Frank Lenzen

Department of Mathematics
University of Innsbruck
Technikerstr. 21a
6020 Innsbruck, Austria

e-mail: Markus.Grasmair@uibk.ac.at
Frank.Lenzen@uibk.ac.at

December 9, 2009

Abstract

Total variation regularization and anisotropic filtering have been established as standard methods for image denoising because of their ability to detect and keep prominent edges in the data. Both methods, however, introduce artifacts: In the case of anisotropic filtering, the preservation of edges comes at the cost of the creation of additional structures out of noise; total variation regularization, on the other hand, suffers from the stair-casing effect, which leads to gradual contrast changes in homogeneous objects, especially near curved edges and corners. In order to circumvent these drawbacks, we propose to combine the two regularization techniques. To that end we replace the isotropic TV semi-norm by an anisotropic term that mirrors the directional structure of either the noisy original data or the smoothed image. We provide a detailed existence theory for our regularization method by using the concept of relaxation. The numerical examples concluding the paper show that the proposed introduction of an anisotropy to TV regularization indeed leads to improved denoising: the stair-casing effect is reduced while at the same time the creation of artifacts is suppressed.

MSC: 68U10, 49J45.

1 Introduction

Because of unavoidable inaccuracies inherent in every data acquisition process, it is not possible to recover a precise images of incoming signals. Instead, the recorded data are defective in various manners. Typical problems that may

occur are blurring, optical distortions read-out errors, defects of the electronic systems, etc.

In the present paper, we concentrate on the treatment of data distorted by random noise. A first ansatz is to apply a low-pass filter, for instance convolution with a Gaussian kernel removing the high frequencies of both the noise and the image content. This is equivalent to solving the heat equation

$$\partial_t u + \Delta u = 0, \quad u(0) = u^0 \quad (1)$$

up to some specified time t determining the amount of smoothing, where u^0 denotes the given noisy data. The diffused function $u(t)$ then serves as denoised image.

Linear diffusion indeed efficiently removes noise, but at the same time it destroys many important features. Edges, that is, discontinuities in the intensities, immediately get smeared out and the whole image becomes blurry and diffuse. For these reasons, this method is not suited for most applications of image processing.

In order to enhance the quality of the denoised image, Perona & Malik introduced a modification of the diffusion equation (1) by a non-constant diffusivity [16]. Instead of (1), they proposed to solve the equation

$$\partial_t u + \operatorname{div}\left(\frac{\nabla u}{1 + c|\nabla u|^2}\right) = 0, \quad u(0) = u^0$$

with $c > 0$. The idea behind this modification is to stop the diffusion at edges, which are characterized by steep gradients, while treating regions with nearly constant intensity as in the uniform model. As further improvement and stabilization of the solution, Catté et al.[7] suggested to scale the diffusivity using the gradient of the uniformly smoothed image. Then the diffusion equation reads as

$$\partial_t u + \operatorname{div}\left(\frac{\nabla u}{1 + c|\nabla u_\sigma|^2}\right) = 0, \quad u(0) = u^0, \quad (2)$$

where $u_\sigma = K_\sigma * u$ denotes the convolution of u with a Gaussian kernel of variance σ^2 .

The above method can be further refined by noting that only diffusion across edges should be prohibited, but diffusion along edges in fact encouraged. To that end it is necessary to consider not only the size of the gradient but also its local direction, which can be computed from the smoothed tensor $J_\rho(u) := K_\rho * (\nabla u_\sigma \otimes \nabla u_\sigma)$ for some $\rho > 0$.

The eigenvector v_1 corresponding to the largest eigenvalue of $J_\rho(u)(x)$ points in the prevailing direction of ∇u_σ near x , while the *local coherence*, that is, the squared difference of the eigenvalues, indicates how pronounced this direction is.

This directional knowledge can be incorporated into the diffusion equation by replacing the diffusivity, which for the modified Perona–Malik model (2) has the form $\frac{1}{1+|\nabla u_\sigma|^2}$, by a matrix or tensor $A(u)$ that decreases the diffusion

only in direction of v_1 . One possibility for defining the tensor $A(u)$ is shown in Section 2, further details can be found in [20].

With this modification, we obtain the anisotropic diffusion equation

$$\partial_t u + \operatorname{div}(A(u)\nabla u) = 0, \quad u(0) = u^0. \quad (3)$$

In [20], it has been shown that the diffusion equation (3) has a unique solution u , which satisfies $u(t) \in C^\infty(\Omega)$ for all $t > 0$.

In general, diffusion filtering is closely connected with variational methods [19]. If one applies an implicit Euler scheme for solving the time discretized diffusion equation, one obtains that each step requires the minimization of a variational functional. In the case of anisotropic diffusion, this functional reads as

$$\mathcal{F}(u, u^0) := \frac{1}{2} \int_{\Omega} (u(x) - u^0(x))^2 + \alpha \nabla u(x)^T A(u) \nabla u(x) dx, \quad (4)$$

where $\alpha > 0$ is the step size.

Setting $u_\alpha^0 := u^0$ and defining the functions $u_\alpha^k := \arg \min_u \mathcal{F}(u, u_\alpha^{k-1})$ for $k \in \mathbb{N}$, one can show by using standard results from semi-group theory that, for α sufficiently small, u_α^k is an approximation of the solution of the corresponding diffusion equation at time $k\alpha$ (see [6, 21]).

Parallel to the refinement of diffusion methods, a different variational approach was developed. One reason was that, although the goal of anisotropic diffusion was edge enhancement, all models introduced above have the drawback that the solutions are in fact smooth [7, 20]. Thus it is never possible to recover sharp edges, even if they can be seen clearly in the original noisy data. Also, in regions where no edges are present, the anisotropic model inherits from the isotropic diffusion the problem that speckles appear, because the quadratic regularization term hardly penalizes sufficiently slow oscillations.

As a remedy, it was proposed to consider variational regularization methods in the space $BV(\Omega)$ of functions of bounded variation, which consists of all integrable functions $u : \Omega \rightarrow \mathbb{R}$ whose distributional derivative can be represented by a finite Radon measure Du (see [2, 12]). The most important method is the minimization of the *Rudin–Osher–Fatemi* or *total variation* functional (see [1, 17]), which is defined as

$$\mathcal{F}(u, u^0) := \int_{\Omega} \frac{1}{2} (u(x) - u^0(x))^2 dx + \alpha |Du|(\Omega). \quad (5)$$

Note that for a Sobolev function u we have the identity $|Du|(\Omega) = \int_{\Omega} |\nabla u(x)| dx$.

The linear growth of the regularization term near infinity has the consequence that only the total height of a slope is penalized, but not its steepness. This permits the presence of edges in the minimizer, which can be seen as infinitely steep gradients concentrated on a narrow line. The singularity at zero, on the other hand, encourages zero gradients, as the cost for small oscillations is high compared to a possible gain through the fidelity term $\|u - u^0\|^2$. Thus, the minimizer of the total variation functional typically consists of regions of almost constant intensity, well separated by distinct edges.

Still, total variation regularization has its problems: First, it tends to insert edges where they should not be. Smooth changes of the intensities are broken up and a gradual transition is created instead. In the literature, this issue is called *stair-casing effect*. Second, it leads to a contrast loss near edges that mainly depends on its curvature. Thus, especially corners in objects, but also regions with strongly varying curvature are resolved badly.

In order to combine the benefits of anisotropic filtering and total variation regularization, we propose to apply anisotropic total variation regularization based on the functional

$$\mathcal{F}(u, u^0) = \int_{\Omega} (u(x) - u^0(x))^2 + \alpha (\nabla u^T(x) A(u)(x) \nabla u(x))^{\frac{1}{2}} dx .$$

Similar functionals have already been proposed in the literature for the problem of inpainting (see e.g. [4, 8]). There, the ℓ^1 -norm of the rotated gradient $R(x) \nabla u(x)$ is used as regularization term. Here, we use directional information in order to introduce different weights on the components of $\nabla u(x)$ in a local coordinate system in x .

The precise definition of the anisotropy matrix $A(u)$ is given in Sec. 2. The linear growth of the regularization term implies that $BV(\Omega)$ is the natural space of definition for the functional \mathcal{F} . In particular, it is necessary to define how \mathcal{F} acts on edges, that is, discontinuities in u . Because the regularization term depends in a non-trivial manner on u , standard results from the calculus of variations (see e.g. [5, 11, 13]) do not apply. Still, it is possible to exploit the structure of $A(u)$, which allows to define the required extension of \mathcal{F} to $BV(\Omega)$. A detailed derivation of the extension is provided in Sec. 3 along with the proof of the existence of minimizers.

We describe the numerical implementation of the regularization method in Sec. 4. The discretization is carried out by a finite element method; for the minimization of the discretized functional we use a steepest descent method combined with a fixed point iteration.

The advantages of introducing an anisotropy to total variation regularization are demonstrated in Sec. 5. Numerical experiments show that this ansatz yields cartoon like images similar to standard total variation regularization, but the typical stair-casing effect is suppressed. In addition, the contrast in the images is much better preserved.

2 Anisotropic diffusion/regularization

From now on, we assume that $u^0 \in L^\infty(\Omega)$ is fixed. In order to simplify the notation, we therefore omit the dependence on u^0 in the regularization functional \mathcal{F} . On the other hand, it is convenient to introduce the anisotropy $A(u)$ as an additional argument both for the existence theory in Sec. 3 and the numerical implementation in Sec. 4. Our proposed regularization functional thus reads as

$$\mathcal{F}(u, A(u)) = \int_{\Omega} (u(x) - u^0(x))^2 + \alpha (\nabla u^T(x) A(u)(x) \nabla u(x))^{\frac{1}{2}} dx . \quad (6)$$

The matrix $A(u)$ in (6) is a particular instance of diffusion tensors introduced in [20] for the definition of (linear) anisotropic diffusion. Its construction is carried out in two steps. First we calculate the smoothed structure tensor J_ρ , which reflects the local structure of u . Then we define $A(u)$ by adapting the eigenvalues of J_ρ while leaving the eigenvectors unchanged.

Assume that $\Omega \subset \mathbb{R}^2$ is an open, bounded, and convex domain with sufficiently regular boundary $\partial\Omega$, and that $u \in L^2(\Omega)$. As $A(u)$ is intended to adapt to the local structure of the function u , its local geometry has to be accessed. This is achieved by means of the *structure tensor* J_ρ , which depends on the smoothed mapping $u_\sigma := K_\sigma * u$ obtained by convolving u with the Gaussian kernel of variance $\sigma^2 > 0$,

$$K_\sigma(x) := \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

Now we calculate the structure tensor $J_\rho(u) : \Omega \rightarrow \text{Sym}^2$ from the smooth mapping $\nabla u_\sigma \in C^\infty(\Omega, \mathbb{R}^2)$ by

$$J_\rho(u) := K_\rho * J_0(u) := K_\rho * (\nabla u_\sigma \nabla u_\sigma^T)$$

for some $\rho > 0$. Here, the convolution of J_0 and K_ρ is carried out separately in each component. The smoothing of the data u is performed in order to make the structure tensor insensitive to noise, whereas the smoothing of J_0 provides a local averaging of the information contained in u_σ .

At each $x \in \Omega$, the structure tensor $J_\rho(u)(x)$ is a positive semi-definite, symmetric 2×2 -matrix. Consequently, its eigenvalues are real and non-negative. Denote now by $\lambda_1(x) \geq \lambda_2(x) \geq 0$ the ordered eigenvalues of $J_\rho(u)(x)$. Then we find that $\lambda_1(x) \approx \lambda_2(x) \approx 0$ in regions where the function u is almost constant, $\lambda_1(x) \gg \lambda_2(x) = 0$ at edges, i.e., at the border between these regions, and $\lambda_1(x) \geq \lambda_2(x) \gg 0$ at corners, i.e., at junctions of edges. Along edges, the eigenvector corresponding to the eigenvalue $\lambda_2(x)$ is parallel to the isolevels of u_σ .

We now define the diffusion tensor $A(u)(x)$. For notational convenience, we omit in the following presentation all dependencies on the function u . We now consider an eigenvalue decomposition $J_\rho(x) = U^T(x)\Sigma(x)U(x)$, where $U(x) \in \mathbb{R}^{2 \times 2}$ is an orthogonal matrix and $\Sigma(x) := \text{diag}(\lambda_1(x), \lambda_2(x)) \in \mathbb{R}^{2 \times 2}$ is the diagonal matrix of the ordered eigenvalues. The diffusion tensor is defined by

$$\begin{aligned} A(u) : \Omega &\rightarrow \text{Sym}^2 \\ x &\mapsto U^T(x) \text{diag}(\gamma(\lambda_1(x) - \lambda_2(x)), 1) U(x) \end{aligned} \quad (7)$$

where

$$\gamma(s) := \frac{1}{1 + \frac{s^2}{\gamma_0^2}} \quad (8)$$

for some $\gamma_0 > 0$ (cf. [16, 7, 20]).

There are two main possibilities for the application of the diffusion tensor in the regularization functional \mathcal{F} . The first, main variant, which is also stated

in (6), defines the anisotropy by means of the structural information of the minimizer of \mathcal{F} . Though leading to a good estimation of the edges present in the data, the nonlinear dependence of A on the image u poses severe difficulties to the actual computation of the solution. In particular, it is possible that local or even multiple global minimizers exist, because the functional $u \mapsto \mathcal{F}(u, A(u))$ fails to be convex.

As an alternative, it is possible to base the definition of the diffusion solely on the structural properties of the given data u^0 . That is, one minimizes the functional $u \mapsto \mathcal{F}(u, A(u^0))$. This strategy has the drawback that the diffusion tensor $A(u^0)$ is affected by the noise contained in the original data. On the other hand, the computations become considerably easier, as the regularization functional now is convex. Also, its theoretical treatment is simplified—most problems that arise during the existence proof in the following section are rooted in the non-locality of the integrand in (6) that arises if A depends on u .

3 Existence Theory

In the case of linear anisotropic regularization, where the regularization term is quadratic, the existence of minimizers can be proven easily by applying direct methods (see for instance [9, 14]). The convexity of the integrand with respect to the gradient variable implies the weak lower semi-continuity of the functional \mathcal{F} in $W^{1,2}(\Omega)$. Then the fact that $\mathcal{F}(u) \geq c\|u\|_{W^{1,2}}^2$ for some $c > 0$ proves that a minimizer exists.

The situation is more complicated for the non-quadratic functional proposed in this paper. Due to the linear growth of the integrand with respect to the gradient variable, $BV(\Omega)$ is the natural space in which to search for a minimizer. As the functional in (6) is only given for $u \in W^{1,1}(\Omega)$, it is therefore necessary to find a suitable extension for $u \in BV(\Omega) \setminus W^{1,1}(\Omega)$. In particular, one has to define how the integrand acts on jump discontinuities of u .

This extension to functions of bounded variation is generally carried out by *relaxation*. First, one sets $\mathcal{F}(u) := +\infty$ for $u \in BV(\Omega) \setminus W^{1,1}(\Omega)$. Then one defines the *relaxed functional* \mathcal{F}_R as the largest weakly* lower semi-continuous functional on $BV(\Omega)$ that is smaller or equal than \mathcal{F} . If the original functional satisfies a certain growth condition, then the relaxed functional \mathcal{F}_R has a minimizer \hat{u} . If moreover $\hat{u} \in W^{1,1}(\Omega)$, then it is in fact a minimizer of \mathcal{F} .

We now treat the relaxation of the functional \mathcal{F} defined in Section 2.

Since the mapping $u \mapsto \mathcal{S}(u) := \frac{1}{2}\|u - u^0\|^2$ is weakly* continuous on $BV(\Omega)$, it easily follows that

$$\mathcal{F}_R(u) = (\mathcal{S} + \alpha\mathcal{R})_R(u) = \mathcal{S}(u) + \alpha\mathcal{R}_R(u)$$

for every $u \in BV(\Omega)$. Therefore, it is sufficient to compute the relaxation $\mathcal{R}_R(u)$ of the regularization term $\mathcal{R}(u)$.

In order to be able to also deal with other integrands proposed for anisotropic regularization, we consider a generalized setting. We assume that $\mathcal{R}: BV(\Omega) \rightarrow$

$[0, +\infty]$ has the form

$$\mathcal{R}(u) := \begin{cases} \int_{\Omega} g(G(u)(x), \nabla u(x)) dx, & \text{if } u \in W^{1,1}(\Omega), \\ +\infty, & \text{if } u \in BV(\Omega) \setminus W^{1,1}(\Omega), \end{cases}$$

where $G: L^1(\Omega) \rightarrow C(\bar{\Omega}; V)$ maps a summable function u to a continuous function $G(u)$ with values in some normed vector space V , and the integrand g maps $V \times \mathbb{R}^n$ to $[0, +\infty)$.

We assume that G and g satisfy the following conditions:

1. The functional $G: BV(\Omega) \rightarrow C(\bar{\Omega}; V)$ is continuous and bounded with respect to the L^1 -norm on $BV(\Omega)$ and the maximum norm on $C(\bar{\Omega}; V)$.
2. The integrand $g: V \times \mathbb{R}^n \rightarrow [0, +\infty)$ is continuous and non-negative.
3. For every $A \in V$, the function $\xi \mapsto g(A, \xi)$ is convex and positively homogeneous, that is, $g(A, \lambda\xi) = \lambda g(A, \xi)$ for all $\lambda \geq 0$.
4. There exists a non-decreasing, continuous function $C: [0, +\infty) \rightarrow [1, +\infty)$ such that

$$C(|A|)^{-1} |\xi| \leq g(A, \xi) \leq C(|A|) (1 + |\xi|) \quad (9)$$

for all $(A, \xi) \in V \times \mathbb{R}^n$.

Remark 1. The functional \mathcal{F} of Section 2 satisfies above conditions with the choices $V = \text{Sym}^2$ normed with the spectral norm $|A| = \|A\|_2$,

$$G(u) = J_\rho(u), \quad \text{and} \quad g(A, \xi) := \sqrt{\xi^T \Gamma(A) \xi},$$

where $\Gamma(A) = U^T \text{diag}(\gamma(\lambda_1 - \lambda_2), 1)U$, if $A = U^T \text{diag}(\lambda_1, \lambda_2)U$ is a singular value decomposition of A with $\lambda_1 \geq \lambda_2$ and $\gamma(s) = 1/(1 + s^2/\gamma_0^2)$ is as in (8).

- Condition 1 is satisfied, because J_ρ only consists of multiplications and convolutions with smooth functions.
- Since the singular values of a symmetric matrix depend continuously on the entries of the matrix, Condition 2 is satisfied.
- Condition 3 obviously holds.
- Finally, $\|\Gamma(A)\|_2 = 1$ and

$$\|\Gamma(A)^{-1}\|_2^2 = 1 + \frac{(\lambda_1 - \lambda_2)^2}{\gamma_0^2} \leq 1 + \frac{\lambda_1^2}{\gamma_0^2} = 1 + \frac{\|A\|_2^2}{\gamma_0^2},$$

which implies that

$$\frac{\gamma_0^2}{\gamma_0^2 + \|A\|_2^2} |\xi| \leq \sqrt{\xi^T \Gamma(A) \xi} \leq |\xi|,$$

and thus Condition 4 holds.

This shows the assertion. ■

Theorem 2. *Let the Conditions 1-4 be satisfied. Then*

$$\mathcal{R}_R(u) = \int_{\Omega} g\left(G(u)(x), \frac{dDu}{d|Du|}(x)\right) d|Du| \quad (10)$$

for every $u \in BV(\Omega)$. Here, the relaxation \mathcal{R}_R of \mathcal{R} is defined as

$$\mathcal{R}_R := \sup\left\{\mathcal{T} : BV(\Omega) \rightarrow [0, +\infty] : \mathcal{T} \leq \mathcal{R} \text{ is weakly* lower semi-continuous}\right\}.$$

Theorem 3. *Let the Conditions 1-4 be satisfied. For every $u^0 \in L^2(\Omega)$ and $\alpha > 0$, the mapping $u \mapsto \mathcal{F}_R(u) = \frac{1}{2}\|u - u^0\|^2 + \alpha\mathcal{R}_R(u)$ attains its minimum in $BV(\Omega)$.*

If no dependence of the integrand g on $G(u)$ were present, Theorem 2 would directly follow from well-known results concerning relaxation of local integral functionals on $BV(\Omega)$. The non-locality introduced by the term $G(u)$, however, poses severe difficulties. In the present case, these difficulties can be met, as, locally around $u \in BV(\Omega)$, the function G is almost constant.

We now introduce some notation and auxiliary Lemmas 5-7 for the proof of Theorem 2.

Definition 4. Let $u \in BV(\Omega)$. A sequence $(u_k)_{k \in \mathbb{N}} \subset W^{1,1}(\Omega)$ is called *admissible* for u , if $\|u_k - u\|_1 \rightarrow 0$ and

$$\sup_{k \in \mathbb{N}} \mathcal{R}(u_k) < +\infty. \quad \blacksquare$$

Lemma 5. *A sequence $(u_k)_{k \in \mathbb{N}} \subset W^{1,1}(\Omega)$ weakly* converges to $u \in BV(\Omega)$ if and only if it is admissible. Moreover, for every u there exists at least one admissible sequence $(u_k)_{k \in \mathbb{N}}$ satisfying*

$$\liminf_{k \rightarrow \infty} \mathcal{R}(u_k) \leq C(\|G(u)\|_{\infty}) (1 + |Du|(\Omega)). \quad (11)$$

Proof. Let $(u_k)_{k \in \mathbb{N}} \subset W^{1,1}(\Omega)$ weakly* converge to u . Then in particular $\|u_k - u\|_1 \rightarrow 0$. Therefore, the continuity of G implies that $\|G(u_k) - G(u)\|_{\infty} \rightarrow 0$. In particular, there exists $t > 0$ such that $\|G(u_k)\|_{\infty} \leq t$ for all $k \in \mathbb{N}$. Now (9) implies that

$$\begin{aligned} \mathcal{R}(u_k) &= \int_{\Omega} g(G(u_k)(x), \nabla u_k(x)) dx \\ &\leq C(t) \int_{\Omega} (1 + |\nabla u_k(x)|) dx = C(t)(\mathcal{L}^2(\Omega) + \|\nabla u_k\|_1). \end{aligned}$$

Since $(u_k)_{k \in \mathbb{N}}$ is weakly* convergent, it follows that $\sup_k \|\nabla u_k\|_1 < \infty$, which in turn implies that $\sup_k \mathcal{R}(u_k) < \infty$. The converse implication can be shown in a similar manner.

In order to show the existence of an admissible sequence $(u_k)_{k \in \mathbb{N}}$ for u satisfying (11) recall that there exists a sequence $(u_k)_{k \in \mathbb{N}}$ with $\|u_k - u\|_1 \rightarrow 0$ and $\lim_k \|\nabla u_k\|_1 = |Du|(\Omega)$. In particular we have that $\|G(u_k)\|_\infty \rightarrow \|G(u)\|_\infty$. The continuity of C therefore implies that also $C(\|G(u_k)\|_\infty) \rightarrow C(\|G(u)\|_\infty)$. The inequality (11) now follows again from (9). \square

Lemma 6. *There is a non-increasing, positive function $s: [0, +\infty) \rightarrow (0, +\infty)$ such that*

$$\mathcal{R}(u) \geq s(\|u\|_1) \|\nabla u\|_1 \quad (12)$$

for every $u \in W^{1,1}(\Omega)$.

Proof. Define

$$s(K) := \inf \left\{ C(\|G(u)\|_\infty)^{-1} : \|u\|_1 \leq K \right\}.$$

By assumption, the functional G is bounded and C is locally bounded, which implies that $s(K) > 0$ for all K . The estimate (12) now follows from (9). \square

Lemma 7. *The functional \mathcal{R}_R is weakly* lower semi-continuous and*

$$\mathcal{R}_R(u) = \inf \left\{ \liminf_{k \rightarrow \infty} \mathcal{R}(u_k) : (u_k)_{k \in \mathbb{N}} \text{ is admissible for } u \right\} \quad (13)$$

for every $u \in BV(\Omega)$.

Proof. Equation (13) is a consequence of Lemma 5. The proof of the weak* lower semi-continuity of \mathcal{R}_R is along the lines of [15, Le. 4.3] or [18, Le. 5.4, Thm. 5.8]. The only difference is that the estimate $\mathcal{R}(u) \geq c\|\nabla u\|_1$ is replaced by the weaker inequality (12). \square

For the proof of Theorem 2, it is convenient to separate the dependence of \mathcal{R} on $G(u)$ and ∇u . This is achieved by introducing a functional $\mathcal{S}: C(\bar{\Omega}; V) \times BV(\Omega) \rightarrow [0, +\infty)$ setting

$$\mathcal{S}(v, u) := \int_{\Omega} g\left(v(x), \frac{dDu}{d|Du|}(x)\right) d|Du|.$$

Lemma 8. *For every $K > 0$ there exists a function $\omega_K: [0, +\infty) \rightarrow [0, +\infty)$ with $\lim_{t \rightarrow 0^+} \omega_K(t) = 0$ such that*

$$|\mathcal{S}(v, u) - \mathcal{S}(\hat{v}, u)| \leq \omega_K(\|v - \hat{v}\|_\infty) |Du|(\Omega)$$

whenever $u \in BV(\Omega)$ and $v, \hat{v} \in C(\bar{\Omega}; V)$ satisfy $\|v\|_\infty, \|\hat{v}\|_\infty \leq K$.

Proof. The continuous function g is uniformly continuous on the compact set $\bar{B}_K(0) \times S^1$. Therefore there exists a modulus of continuity $\omega_K: [0, +\infty) \rightarrow [0, +\infty)$ satisfying $\lim_{t \rightarrow 0^+} \omega_K(t) = 0$ such that

$$|g(w, s) - g(\hat{w}, \hat{s})| \leq \omega_K(|w - \hat{w}| + |s - \hat{s}|)$$

for every $w, \hat{w} \in V$ with $|w|, |\hat{w}| \leq K$, and $s, \hat{s} \in S^1$.

Then, for $u \in BV(\Omega)$ and $v, \hat{v} \in C(\bar{\Omega})$ with $\|v\|_\infty, \|\hat{v}\|_\infty \leq K$, we have

$$\begin{aligned} |\mathcal{S}(v, u) - \mathcal{S}(\hat{v}, u)| &= \left| \int_{\Omega} \left[g\left(v(x), \frac{\nabla u(x)}{|\nabla u(x)|}\right) - g\left(\hat{v}(x), \frac{\nabla u(x)}{|\nabla u(x)|}\right) \right] d|Du| \right| \\ &\leq |Du|(\Omega) \sup_{x \in \Omega} \left| g\left(v(x), \frac{\nabla u(x)}{|\nabla u(x)|}\right) - g\left(\hat{v}(x), \frac{\nabla u(x)}{|\nabla u(x)|}\right) \right| \\ &\leq |Du|(\Omega) \sup_{s \in S^1} \sup_{x \in \Omega} |g(v(x), s) - g(\hat{v}(x), s)| \\ &\leq |Du|(\Omega) \omega_K (\|v - \hat{v}\|_\infty), \end{aligned}$$

which shows the assertion. \square

Lemma 9. *Let $u \in BV(\Omega)$ and let $(u_k)_{k \in \mathbb{N}}$ be an admissible sequence for u . Then*

$$\liminf_{k \rightarrow \infty} \mathcal{R}(u_k) = \liminf_{k \rightarrow \infty} \mathcal{S}(G(u_k), u_k) = \liminf_{k \rightarrow \infty} \mathcal{S}(G(u), u_k).$$

In particular, we have in view of Lemma 7 that

$$\mathcal{R}_R(u) = \inf \left\{ \liminf_{k \rightarrow \infty} \mathcal{S}(G(u), u_k) : (u_k)_{k \in \mathbb{N}} \text{ is admissible for } u \right\}.$$

Proof. Since $(u_k)_{k \in \mathbb{N}}$ is an admissible sequence for u , it follows that $\|u_k - u\|_1 \rightarrow 0$ and $L := \sup_{k \in \mathbb{N}} |Du_k|(\Omega) < +\infty$ (see Lemma 5). In particular, the L^1 -convergence of $(u_k)_{k \in \mathbb{N}}$ implies that $K := \sup_{k \in \mathbb{N}} \|G(u_k)\|_\infty < +\infty$ and that $\|G(u_k) - G(u)\|_\infty \rightarrow 0$.

Let $\omega_K: [0, +\infty) \rightarrow [0, +\infty)$ be as in Lemma 8. Then

$$\begin{aligned} |\mathcal{R}(u_k) - \mathcal{S}(G(u), u_k)| &= |\mathcal{S}(G(u_k), u_k) - \mathcal{S}(G(u), u_k)| \\ &\leq \omega_K (\|G(u_k) - G(u)\|_\infty) |Du_k|(\Omega) \leq L \omega_K (\|G(u_k) - G(u)\|_\infty) \rightarrow 0. \quad \square \end{aligned}$$

Using the above result, we can now conclude the proof of Theorem 2:

Proof (of Thm. 2). Let $u \in BV(\Omega)$ be fixed. Define the auxiliary localized functional $\tilde{\mathcal{R}}: BV(\Omega) \rightarrow [0, +\infty)$ setting

$$\tilde{\mathcal{R}}(v) := \begin{cases} \mathcal{S}(G(u), v), & \text{if } v \in W^{1,1}(\Omega), \\ +\infty, & \text{if } v \in BV(\Omega) \setminus W^{1,1}(\Omega). \end{cases}$$

Denote by $\tilde{\mathcal{R}}_R$ the relaxation of $\tilde{\mathcal{R}}$. From standard results in relaxation of functionals on $BV(\Omega)$ (see for instance [3, 10, 13]), it follows that

$$\tilde{\mathcal{R}}_R(v) = \int_{\Omega} g\left(G(u)(x), \frac{dDv}{d|Dv|}(x)\right) d|Dv|$$

for every $v \in BV(\Omega)$. Lemma 9 implies that

$$\tilde{\mathcal{R}}_R(u) = \inf \left\{ \liminf_{k \rightarrow \infty} \mathcal{S}(G(u), u_k) : (u_k)_{k \in \mathbb{N}} \text{ is admissible for } u \right\} = \mathcal{R}_R(u).$$

This proves the assertion. \square

Proof (of Thm. 3). Let $(u_k)_{k \in \mathbb{N}} \subset BV(\Omega)$ be a minimizing sequence for \mathcal{F}_R . The definition of \mathcal{F}_R implies that u_k can be chosen to be an element of $W^{1,1}(\Omega)$ for every $k \in \mathbb{N}$. Since $(u_k)_{k \in \mathbb{N}}$ is a minimizing sequence, it follows that the sequence $(\|u_k\|_2)_{k \in \mathbb{N}}$ is bounded. Because $\|u_k\|_1 \leq \sqrt{\mathcal{L}^2(\Omega)}\|u_k\|_2$, there exists $K > 0$ such that $\|u_k\|_1 \leq K$ for all $k \in \mathbb{N}$. Lemma 6 therefore implies that $\mathcal{R}_R(u_k) \geq s(K)\|\nabla u_k\|_1$ for every $k \in \mathbb{N}$. Thus, also the sequence $(\|\nabla u_k\|_1)_{k \in \mathbb{N}}$ is bounded. Consequently, the sequence $(u_k)_{k \in \mathbb{N}}$ admits a subsequence, again denoted $(u_k)_{k \in \mathbb{N}}$, weakly* converging to some $u \in BV(\Omega)$. Now the weak* lower semi-continuity of \mathcal{F}_R implies that $\mathcal{F}_R(u) \leq \liminf_k \mathcal{F}_R(u_k)$, and therefore u minimizes \mathcal{F}_R . \square

4 Numerics

After the general considerations of the previous section, we now turn back to the functional

$$\mathcal{F}(u, A(u)) := \int_{\Omega} f(u(x) - u^0(x), \nabla u(x), A(u)(x)) dx, \quad (14)$$

where the integrand equals

$$f(s, \xi, A) := s^2 + \alpha (\xi^T A \xi)^{\frac{1}{2}}.$$

We describe the numerical minimization of (14) with respect to u .

In order to simplify computations, we assume that $\Omega \subset \mathbb{R}^2$ is rectangular. The anisotropy $A(u): \Omega \rightarrow \mathbb{R}^{2 \times 2}$ is as defined in (7). Recall that for every $x \in \Omega$ the matrix $A(u)(x)$ is symmetric and positive definite.

We implement a finite element method for minimizing \mathcal{F} . To that end, we consider on $\bar{\Omega}$ a uniform grid of grid size h , nodes x_i , $i = 1, \dots, N$, and grid cells or *elements* Q_j , $j = 1, \dots, M$. The corresponding finite element space V_h consists of continuous functions that are bilinear on each element Q_j . A function $u \in V_h$ is identified with the vector of its values $\mathbf{u} = (u_1, \dots, u_N)$ at the nodes x_1, \dots, x_N .

Now let $\mathcal{N}(Q_j)$ denote the index set of the vertices of Q_j , and denote by x_j^c the center of Q_j . On each element Q_j we apply the quadrature rule

$$\begin{aligned} & \int_{Q_j} f(u(x) - u^0(x), \nabla u(x), A(u)(x)) dx \\ & \approx \frac{h^2}{4} \sum_{i \in \mathcal{N}(Q_j)} f(u_i - u_i^0, \nabla u|_{Q_j}(x_i), A(u)(x_j^c)). \end{aligned} \quad (15)$$

Note that ∇u is not continuous on $\bar{\Omega}$. Thus in general $\nabla u|_{Q_i}(x) \neq \nabla u|_{Q_j}(x)$ if $i \neq j$ and $x \in Q_i \cap Q_j$. Summing up both sides of (15) over all elements Q_j , we arrive at a discretization $\tilde{\mathcal{F}}_h$ of \mathcal{F} , defined by

$$\tilde{\mathcal{F}}_h(\mathbf{u}, A(u)) := \frac{h^2}{4} \sum_{j=1}^M \sum_{i \in \mathcal{N}(Q_j)} f(u_i - u_i^0, \nabla u|_{Q_j}(x_i), A(u)(x_j^c)). \quad (16)$$

In (16), $\nabla u|_{Q_j}$ is evaluated at the vertices of Q_j . Since the ansatz functions are chosen to be bilinear, the gradients can therefore be expressed by difference quotients of \mathbf{u} . Thus $A(u)$ remains the only term in the right hand side of (16) that still depends on the continuous function u . In the following, we describe how to compute a discrete approximation.

The first step in the calculation of $A(u)$ consists the convolution of u with a Gaussian kernel K_σ . In view of the relationship between Gaussian convolution and linear diffusion, we compute u_σ by numerically solving a linear diffusion equation. We define $u_\sigma := \tilde{u}(T)$ with $T = \sigma^2$, where \tilde{u} solves

$$\partial_t \tilde{u} - \Delta \tilde{u} = 0, \quad \tilde{u}(0) = u,$$

with Neumann boundary conditions.

As a next step, $J_0 = \nabla u_\sigma^T \nabla u_\sigma$ is calculated at the mid points x_j^c of the grid. Similarly as for u_σ , we obtain the smoothed function J_ρ by linear diffusion on the staggered grid given by the points x_j^c , $j = 1, \dots, M$.

Finally, we calculate for each j the approximation $A_j(\mathbf{u}) \approx A(u)(x_j^c)$ from the eigenvalue decomposition of $J_\rho(u)(x_j^c)$. This is done by substituting the eigenvalues $\lambda_1 \geq \lambda_2$ of $J_\rho(u)(x_j^c)$ by $\gamma(\lambda_1 - \lambda_2)$ and 1. Here, $\gamma(t) = 1/(1+t^2/\gamma_0^2)$ is as in (8).

Inserting $\mathbf{A}(\mathbf{u})$ in (16), we end up with the minimization of

$$\mathcal{F}_h(\mathbf{u}, \mathbf{A}(\mathbf{u})) := \frac{h^2}{4} \sum_{j=1}^M \sum_{i \in \mathcal{N}(Q_j)} f(u_i - u_i^0, \nabla u|_{Q_j}(x_i), A_j(\mathbf{u})). \quad (17)$$

The challenging part in the minimization of \mathcal{F}_h is the complex dependence of the anisotropy \mathbf{A} on \mathbf{u} . We have chosen to apply a fixed point iteration with respect to the second argument of \mathcal{F}_h . Starting with a discretization \mathbf{u}^0 of u^0 , we iteratively compute

$$\mathbf{u}^k := \arg \min_{\mathbf{u}} \mathcal{F}_h(\mathbf{u}, \mathbf{A}(\mathbf{u}^{k-1}))$$

until the difference between two subsequent iterates \mathbf{u}^k and \mathbf{u}^{k-1} becomes small enough.

The minimization of $\mathcal{F}_h(\cdot, \mathbf{A}(\mathbf{u}^{k-1}))$ is carried out by a steepest descent method. Since the integrand f is not differentiable, we replace it by the approximation

$$f_\varepsilon(s, \xi, A) := s^2 + \alpha \sqrt{\xi^T A \xi + \varepsilon^2}$$

for some small $\varepsilon > 0$. This introduction of the parameter $\varepsilon > 0$ is standard in total variation regularization.

As mentioned in Section 2, one can also exploit the structural properties of the original data and replace $A(u)$ by $A(u^0)$ in the functional \mathcal{F} . In this situation, the numerical approach is the same, except that no fixed point iteration is required.

5 Numerical Results

In the following, we present numerical results that illustrate the properties of non-quadratic anisotropic regularization. For each of the results we used grid size $h = 1$. The intensity values were scaled to the interval $[0, 1]$ before filtering. The parameters α , σ , ρ , ε , γ_0 , and the number of fixed point iterations, which were chosen individually for each example, are provided with the results.

As a first example, we compare the proposed anisotropic TV regularization with anisotropic quadratic regularization (see (4)) and standard TV regularization (see (5)).

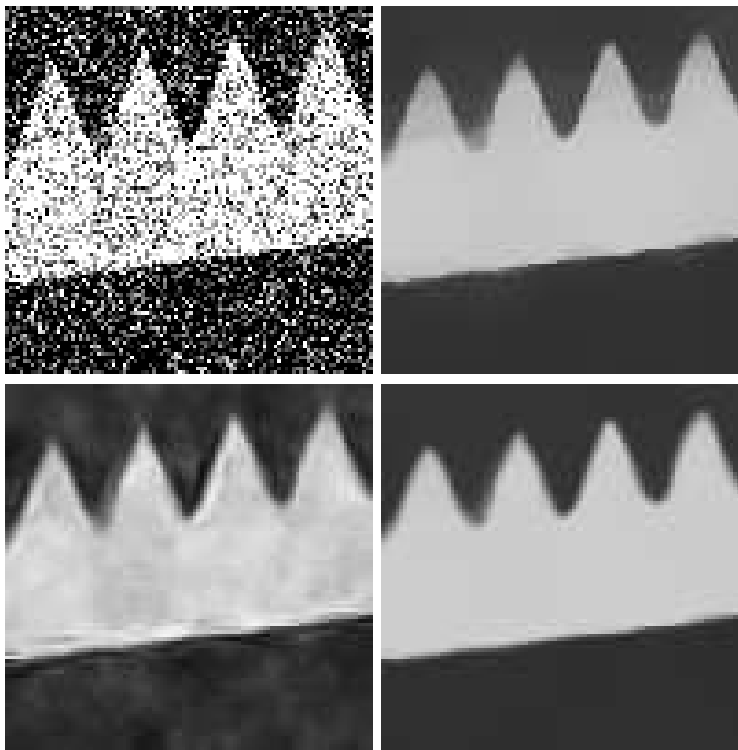


Figure 1: Top left: test image with noise. Top right: image filtered with TV regularization using $\alpha = 1$. Bottom left: image filtered with anisotropic quadratic regularization using $\alpha = 5$. Bottom right: image filtered with anisotropic non-quadratic regularization using $\alpha = 1$.

Our first test image Fig. 1, top left, consists of three separated regions, which are distorted by noise. We run each of the three regularization methods with the parameters $\sigma = 1$, $\rho = 5$, $\varepsilon = 0.001$, $\gamma_0 = 0.001$; for the minimization of \mathcal{F}_h , ten fixed point iterations have been used. The regularization parameter α was adapted to each method so as to obtain results with comparable noise removal and sharpness.

The result of TV regularization with the parameter choice $\alpha = 2$ is shown in Fig. 1, top right. The noise is removed in an adequate manner, but the well-known stair-casing effect is clearly visible: the contrast of the indentations is reduced and the peaks are rounded. Conversely, anisotropic quadratic regularization (see Fig. 1, bottom left, with $\alpha = 5$) preserves the contrast by enhancing the edges, but this edge enhancement comes at the cost of an insufficient noise removal.

As the result of non-quadratic regularization with $\alpha = 1$ (Fig. 1 bottom right) shows, non-quadratic anisotropic regularization combines the benefits of the other two methods: On the one hand, our method is able to restore the peaks in a regular manner, which is achieved by the anisotropy of the regularization. On the other hand, the non-quadratic regularization term encourages piecewise constant results, so that less residual structures remain from noise than in the case of quadratic regularization. We remark, however, that anisotropic quadratic diffusion yields the best reconstruction of the bright peaks.

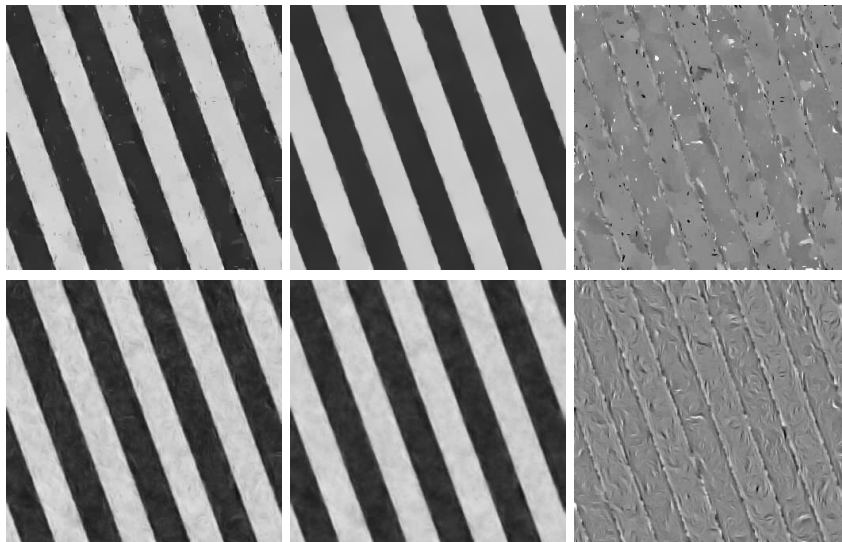


Figure 2: Results after 1 (left column) and 10 iterations (middle column) and the difference between these results (right column) for non-quadratic (top row) and quadratic anisotropic regularization (bottom row), respectively.

In the functional for anisotropic regularization, the anisotropy matrix A depends on the minimizer u . As mentioned in Sect. 2, one may also compute the anisotropy from the data u_0 . We now demonstrate the benefits of the former approach by means of a test image consisting of diagonal stripes with added noise.

Fig. 2 illustrates, how the resulting image improves with the number of fixed point iterations. The results of anisotropic quadratic and non-quadratic

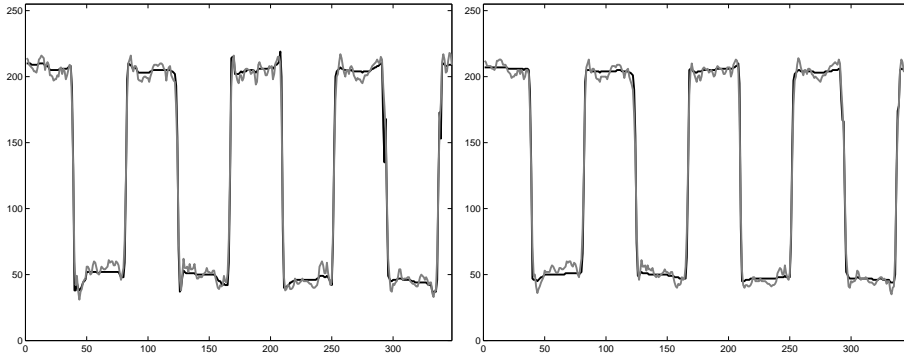


Figure 3: Horizontal cross-section of the results shown in Figure 2 after 1 iteration (left) and 10 iterations (right).

diffusion after one iteration (i.e. A depending only on u^0) are depicted in the left column of Fig. 2; the results after ten iterations in the middle column. Here the parameters $\sigma = 1$, $\rho = 20$, $\varepsilon = 0.001$, $\gamma_0 = 0.001$ were used. The regularization parameter was chosen as $\alpha = 0.5$ in the case of non-quadratic regularization, and $\alpha = 10$ for quadratic regularization.

Numerical tests show that the difference between two subsequent iterates steadily decreases, until after 10 iteration steps no visible changes occur anymore. Therefore no further iterations are required.

After the first iteration, the image still contains some noise, because the anisotropy in the regularization term is not yet adapted to the structure of the solution. As the iteration process continues, the structure is better captured by the anisotropy and artifacts stemming from noise vanish. The artifacts that are removed by the iteration process are highlighted in the difference images in Fig. 2, right column. Moreover, the reduction of artifacts can be observed particularly well in the horizontal cross sections of the results, see Fig. 3.

Finally we investigate the impact of the regularization parameter α on the filtering process. Fig. 4, top left, shows the camera man test image. In the same figure, we present the result of applying anisotropic non-quadratic regularization with the regularization parameters $\alpha = 0.1$, $\alpha = 0.5$ and $\alpha = 1$ and additional parameters $\sigma = 1$, $\rho = 5$, $\varepsilon = 0.001$, $\gamma_0 = 0.001$.

We notice that with increasing parameter α more and more objects of increasing scale vanish. Due to the non-quadratic regularization, regions of constant intensity emerge in the filtered image, while the edges between these regions are kept sharp.

A comparison of non-quadratic anisotropic (see Fig. 4, bottom left) with isotropic TV regularization (see Fig. 5, left) shows that the former method produces more regular edges and preserves contrast. The contrast is also well preserved by quadratic anisotropic regularization (see Fig. 5), but texture and edges appear blurry. On the contrary, the proposed method yields a piecewise



Figure 4: Original image (top left) and iteratively filtered images using $p = 1$ and $\alpha = 0.1$ (top right), $\alpha = 0.5$ (bottom left) and $\alpha = 1$ (bottom right), respectively.

constant, cartoon like solution with well pronounced edges.

Acknowledgments

The work of M.G. 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.



Figure 5: Results of isotropic TV regularization with $\alpha = 0.5$ (left) and quadratic anisotropic regularization with $\alpha = 10$ (right).

- [2] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 2000.
- [3] P. Aviles and Y. Giga. Variational integrals on mappings of bounded variation and their lower semicontinuity. *Arch. Ration. Mech. Anal.*, 115(3):201–255, 1991.
- [4] B. Berkels, M. Burger, M. Droske, O. Nemitz, and M. Rumpf. Cartoon extraction based on anisotropic image classification. In *Vision, Modeling, and Visualization Proceedings*, pages 293–300, 2006.
- [5] G. Bouchitté, I. Fonseca, and L. Mascarenhas. A global method for relaxation. *Arch. Ration. Mech. Anal.*, 145(1):51–98, 1998.
- [6] H. Brézis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*. North-Holland Publishing Co., Amsterdam, 1973. North-Holland Mathematics Studies, No. 5. Notas de Matemática (50).
- [7] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. Numer. Anal.*, 29(1):182–193, 1992.
- [8] R. Chan, S. Setzer, and G. Steidl. Inpainting by flexible Haar wavelet shrinkage. Preprint, University of Mannheim, 2008.
- [9] B. Dacorogna. *Direct methods in the calculus of variations*, volume 78 of *Applied Mathematical Sciences*. Springer, New York, second edition, 2008.

- [10] G. Dal Maso. Integral representation on $BV(\Omega)$ of Γ -limits of variational integrals. *Manuscripta Math.*, 30(4):387–416, 1979/80.
- [11] F. Demengel and R. Temam. Convex functions of a measure and applications. *Indiana Univ. Math. J.*, 33(5):673–709, 1984.
- [12] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.
- [13] I. Fonseca and S. Müller. Relaxation of quasiconvex functionals in $BV(\Omega, R^p)$ for integrands $f(\xi, u, \nabla u)$. *Arch. Ration. Mech. Anal.*, 123(1):1–49, 1993.
- [14] E. Giusti. *Direct Methods in the Calculus of Variations*. World Scientific Publishing, River Edge, NJ, 2003.
- [15] M. Grasmair. *Relaxation of Nonlocal Integrals with Rational Integrands*. PhD thesis, University of Innsbruck, Austria, 2006.
- [16] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7):629–639, 1990.
- [17] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1–4):259–268, 1992.
- [18] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, New York, 2009.
- [19] O. Scherzer and J. Weickert. Relations between regularization and diffusion filtering. *J. Math. Imaging Vision*, 12(1):43–63, 2000.
- [20] J. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner, Stuttgart, 1998. European Consortium for Mathematics in Industry.
- [21] K. Yosida. *Functional Analysis*, volume 123 of *Die Grundlehren der Mathematischen Wissenschaften*. Academic Press Inc., New York, 1965.