Deblurring and Denoising of Images by Nonlocal Functionals *

Stefan Kindermann[†], Stanley Osher[‡], Peter W. Jones[§]

Abstract

This paper investigates the use of regularization functionals with nonlocal correlation terms for the problem of image denoising and image deblurring. These functionals are expressed as integrals over the Cartesian product of the pixel space. We show that the class of neighborhood filters can be described in this framework. Using these functionals we can consider the functional analytic properties of some of these neighborhood filters, and show how they can be seen as regularization terms using a smoothed version of the Prokhorov metric. Moreover, we define a nonlocal variant of the well-known bounded variation regularization, which does not suffer from the staircase effect. We show existence of a minimizer of the corresponding regularization functional for the denoising and deblurring problem and we present some numerical examples comparing the nonlocal version to the bounded variation regularization and the nonlocal mean filter.

1 Introduction

Denoising and denoising of images are among the most fundamental problems in image processing. The problem of image denoising is to find a clear image u from a noisy image f. In the deblurring problem a given image f is regarded as a blurry version of an unknown exact image u, which is to be found.

A vast variety of methods for doing this are available touching very different fields of mathematics. Many successful methods for denoising and deblurring can be derived from the energy method. In this framework image processing problems are considered as variational problems where a (cleaner of deblurred) image is computed by a minimization of some energy functional. Typically, such functionals consist of a fidelity term such as the norm of the difference between

*Research supported by NIH U54RR021813, NSF DMS-0312222, NSF ACI-0321917 and NSF DMI-0327077, P. J. is supported by DMS-0140623

[†]Department of Mathematics, UCLA, 520 Portola Plaza, Los Angeles, Ca 90095, on leave from University Linz, Austria, email: kindermann@indmath.uni-linz.ac.at

[‡]Department of Mathematics, UCLA, 520 Portola Plaza, Los Angeles, Ca, 90095, email: sjo@math.ucla.edu

[§]Dep. of Mathematics, Yale University, New Haven, CT 06520, email:jones@math.yale.edu

the clean image and the original noisy image and a regularization term which penalizes high frequency noise.

Besides the variational approach many alternative concepts were developed as well, for instance, filtering by smoothing operators (one of the simplest cases is Gaussian convolution), or approximation by appropriate basis functions - the wide field of Wavelet approximation falls into this field. Moreover, considering the noise as random variable leads to stochastic models. Of course, all these fields are not separated from each other and many algorithms can be viewed under the aspect of filtering, approximation, energy minimization or within the stochastic framework.

In this article we focus on the energy method. The results of such an approach are mainly determined by the choice of the regularization functional. A successful filter in this class is the Rudin-Osher-Fatemi-Method (ROF) [22]. Here the clear image is defined by a variational problem using the bounded variation seminorm as regularization term:

$$u = \mathrm{argmin}_{v \in \mathrm{BV}} \lambda \|v - f\|^2 + |v|_{BV}. \tag{1}$$

$|v|_{\mathrm{BV}}$ denotes the bounded variation seminorm

$$|v|_{\mathrm{BV}} = \sup_{\phi \in C_0^\infty, \|\phi\|_\infty \leq 1} \int_\Omega v \ \nabla\phi(x) \ dx.$$

The success of the BV-norm stems from the fact that it allows discontinuous solutions and hence preserves edges while filtering high frequency oscillations due to noise. Several other methods such as texture models ([17, 19, 25]) or methods for different noise models [21] are derived from the original ROF model.

The ROF-method and similar regularization methods which penalize derivatives are each essentially a local method. The regularization involves only the values and derivatives of v at the same point. Hence, the corresponding Euler-Lagrange equation can be written as a (nonlinear) differential equation.

In contrast to most of the well-known regularization methods we investigate functionals which involve nonlocal terms. In this case the Euler-Lagrange equations are not partial differential equations but include integrals. Our main motivation for such nonlocal models comes from the recent work of [6] using neighborhood filters. Such filters have originally been proposed by Yaroslavsky [26, 27], and further generalized in [23, 24]. In contrast to spatial filtering such as convolution, these filters clean a noisy image by taking an average over similar pixel values (assuming uniform noise). Since similar pixel values can be located far from each other this leads to an essentially nonlocal filtering. These methods are usually considered within filtering theory. However, it is the aim of this paper to show they can also be interpreted within the variational framework. Due to the nonlocality of the neighborhood filters it is not surprising that the corresponding functionals are nonlocal ones.

The main aim of this paper it to relate the neighborhood filter to an energy minimization. The relation is in the sense that applying a neighborhood filter can be interpreted as one step of solving the optimality condition for a certain functional, see Section 3. Moreover, by the interpretation as a minimization problem we can consider functional analytic properties of the penalization functional. Since in general such functionals are not convex, they cannot be related to norms in Banach space. However, we show that the filtering can be interpreted as minimization of a metric similar to the metric of convergence in measure (or Prokhorov metric), see Section 4. In Section 5 we define a generalization of the ideas in Section 3: we propose to minimize an energy using a nonlocal bounded variation (NLBV) functional derived from the space of bounded variation. In Section 5 we study the analytic properties of this. We establish the equivalence of the NLBV-functional with a quotient norm on BV, and prove existence and uniqueness of the corresponding denoising and deblurring functionals. Finally in Section 6 we show some numerical results for denoising and deblurring problems using these nonlocal functionals.

2 Neighborhood Filter

Neighborhood filters have been originally proposed by Yaroslavsky [27, 26]. The main idea is to compute a filtered image u by taking an average of the noisy image f. The average is taken over pixels which have similar greyvalues, and not over pixels which are close in the image, as it is usual for a convolution. The general form of these filters is the following: The value of the cleared image u at position x is defined by

$$u(x) = \frac{1}{C(x)} \int_\Omega K(x,y,f) f(y) dy \qquad C(x) = \int_\Omega K(x,y,f) dy. \tag{2}$$

The choice of the kernel $K(x,y,f)$ (which depends on f) determines the actual filter. $C(x)$ acts as a normalization to ensure that a constant function f is mapped to itself.

In particular, the original Yaroslavsky neighborhood filter – in the following denoted by $\mathrm{YNF}_{1,h}$ – is defined by

$$K(x,y,f) = K_{\mathrm{YNF}_{1,h}}(x,y,f) = \begin{cases} 1 & \text{if } |f(x) - f(y)| \le h \\ 0 & \text{else} \end{cases} .$$

From a computational point of view is more practical to limit the average to a neighborhood of x, which defines the filter $\mathrm{YNF}_{1,h,\rho}$:

$$K(x,y,f) = K_{\mathrm{YNF}_{1,h,\rho}}(x,y,f) = \begin{cases} 1 & \text{if } |f(x) - f(y)| \le h \text{ and } |x-y| \le \rho \\ 0 & \text{else} \end{cases}$$

The previous filters do not take into account differences in the grey values of $f(x)$ larger than h. This can lead to undesirable blocky structures. A version of the original Yaroslavsky filter weights the different grey level by a Gaussian (known as SUSAN-method in [23]). This defines a second kind of Yaroslavsky neighborhood filter denoted by $\mathrm{YNF}_{2,h}$:

$$K(x,y,f) = K_{\mathrm{YNF}_{2,h}}(x,y,f) = e^{-\frac{|f(x)-f(y)|^2}{h^2}}$$

Of course, a similar localized version $\mathrm{YNF}_{2,h,\rho}$ can be defined if the integration is restricted to a neighborhood $|x-y| \leq \rho$, or by an appropriate weighting function (see the bilateral filtering in [24]).

Finally, these filters experienced a significant improvement by Buades, Coll and Morel [6]. They proposed the following ("nonlocal-mean" filter) NLM_h:

$$K(x,y,f) = K_{\mathrm{NLM}_h}(x,y,f) = e^{-\frac{G_\sigma * |f(x-.)-f(y-.)|^2(0)}{h^2}}, \tag{3}$$

here f is assumed to be periodically extended outside of the rectangular domain Ω, as it is common in image processing. This filter mixes grey levels with a spatial filtering. Here $G_\sigma *$ denotes a convolution with a mollifier, i.e. G_σ is a positive function converging to the δ-distribution as $\sigma \to 0$. Hence for small σ, the nonlocal-mean filter approximates the Yaroslavsky filter $\mathrm{YNF}_{2,h}$.

The idea behind the neighborhood filters is to define a measure of similarity between different regions in the image, and that the cleared image at some point x is an average of all pixels which are similar to the grey value $f(x)$ at this point. The nonlocal mean filter differs from the YNF-filters, because not only the pixel values are used to define similarity, but neighborhoods of pixels ("windows") are compared. This is the interpretation of the convolution G_σ. Note that the mollifier is conveniently taken as a function with compact support. The size of the support defines the size of the windows which are compared. A reasonable choice for the size of the support and hence the size of the neighborhoods is to compare ranges from 5×5 to 9×9 pixels for a typical image of size 256×256 pixels. For colored images this can be even smaller [6].

Remark 2.1. Let us remark that a slight modification of the kernel in the previous filter at $x = y$ yields a significant improvement. Obviously, a pixel is always similar to itself, indicated by the fact that all kernel functions $K(x,y,f)$ take their maximum value 1 at $x = y$. By the normalization via $C(x)$, only the relative difference in similarity is taken into account, hence a maximum value at $x = y$, lowers the similarity for all other values. In a numerical procedure it is better to redefine the value of $K(x,y,f)$ at $x = y$, for instance [6] use $K(x,x,f) = \max_{x \neq y} K(x,y,f)$. This increases the ability of the filter to find similarities, away from $x = y$.

3 Nonlocal Functionals

In this section we show how the neighborhood filters can be derived from a variational principle. The main observation is their characterization as one step of a fixed-point iteration to solve the optimality conditions of nonlocal functionals.

Such functionals take the general form:

$$J(u) := \int_{\Omega \times \Omega} g\left(\frac{|u(x)-u(y)|^2}{h^2}\right) w(|x-y|)dxdy \tag{4}$$

with an appropriate positive weight function w, a differentiable filter function $g : \mathbb{R}^+ \to R$, and a parameter h. We formally derive the directional derivative of $J(u)$: Let $v \in L^2(\Omega)$, $t \in \mathbb{R}$, $w \in L^\infty(\mathbb{R}^+)$, and $g \in C^1(\mathbb{R}^+)$, then the following derivative exists:

$$\begin{aligned}
&\frac{d}{dt} J(u+tv) \mid_{t=0} \\
&= \frac{2}{h^2} \int_{\Omega\times\Omega} g'\left(\frac{|u(x)-u(y)|^2}{h^2}\right)(u(x)-u(y))\,(v(x)-v(y))\,w(x-y)dxdy.
\end{aligned}$$

The integral can be split into a term involving $v(x)$ and $v(y)$. In the latter a change of variables $(x, y) \to (y, x)$ gives the same integral as that with $v(x)$, hence we end up with

$$\frac{d}{dt} J(u+tv) \mid_{t=0} = \frac{4}{h^2} \int_{\Omega\times\Omega} g'\left(\frac{|u(x)-u(y)|^2}{h^2}\right)(u(x)-u(y))\,w(x-y)dxv(y)dy.$$

This shows that under the mentioned assumptions on w and g the Frechet-derivative of J as a functional from $L^2(\Omega)$ to $\mathbb{R}$ is given by

$$J'(u) = \frac{4}{h^2} \int_{\Omega} g'\left(\frac{|u(x)-u(y)|^2}{h^2}\right)(u(x)-u(y))\; w(|x-y|)dy. \tag{5}$$

If u is a stationary point of $J(u)$, then by (5), it is a solution to the fixed-point equation:

$$\begin{aligned}
u(x) &= F_{g,w}(u) := \frac{1}{C(x)} \int_{\Omega} g'\left(\frac{|u(x)-u(y)|^2}{h^2}\right) u(y)w(|x-y|)dy \quad (6)\\
C(x) &= \int_{\Omega} g'\left(\frac{|u(x)-u(y)|^2}{h^2}\right) w(|x-y|)dy. \quad (7)
\end{aligned}$$

Note that $F_{g,w}$ has the form of a neighborhood filter (2). If we set $u_0 = f$, and perform a fixed-point iteration

$$u_{n+1} = F_{g,w}(u_n)$$

then the first step is identical to a nonlocal filtering as above.

The Yaroslavsky neighborhood filter $\text{YNF}_{1,h}$ is obtained by the choice $w = 1$ and

$$g_1(s) := \begin{cases} s & 0 \le s \le 1 \\ 1 & \text{else} \end{cases} \quad \Rightarrow \quad g_1'(s) = \begin{cases} 1 & 0 \le s \le 1 \\ 0 & \text{else} \end{cases} .$$

For the corresponding localized version $\text{YNF}_{1,h,\rho}$ we can choose w as a cut-off function

$$w(x) = \begin{cases} 1 & |x| \le \rho \\ 0 & \text{else} \end{cases}$$

In a similar manner we obtain $\text{YNF}_{2,h}$ by

$$g_2(x) := 1 - e^{-x} \;\Rightarrow\; g_2'(x) = e^{-x},$$

which gives the functional

$$J_{1,h}(u) := \int_{\Omega\times\Omega} \left(1 - e^{-\frac{|u(x)-u(y)|^2}{h^2}}\right) w(|x-y|)dxdy. \tag{8}$$

The Buades-Coll-Morel nonlocal-mean filter is not directly related to a functional of the form (4). In analogy to the nonlocal-mean filter a suitable generalization of (4) would be a functional including a convolution (with fixed $\sigma > 0$):

$$J_{2,h}(u) := \int_{\Omega\times\Omega} \left(1 - e^{-\frac{G_\sigma * |u(x-.)-u(y-.)|^2(0)}{h^2}}\right) w(|x-y|)dxdy \tag{9}$$

The formal derivative of the $J_{2,h}$ is given by

$$\begin{aligned} J'_{2,h}(u) &= \frac{4}{h^2}\int_\Omega K_\sigma(x,y)\,(u(x)-u(y))\,w(|x-y|)dxdy \\ K_\sigma(x,y) &= G_\sigma * K_1(x-.,y-.) \\ K_1(x,y) &= \int_\Omega e^{-\frac{G_\sigma * |u(x-.)-u(y-.)|^2(0)}{h^2}} dxdy \end{aligned}$$

Using the assumption that u is periodically extended outside of Ω, the derivative can be derived similar as in (5) observing that the convolution satisfies the identity $\int_\Omega (G_\sigma * q)(x)p(x) = \int_\Omega (G_\sigma * q)(x)p(x)dx$ for any periodic p, q. There is a slight difference between the filter defined by the optimality conditions of (9) and the nonlocal mean filter. The latter uses $K_1(x,y)$ as kernel in (2), while the optimality conditions leads to the kernel $K_\sigma(x,y)$. So except from the convolution of K_1 the corresponding fixed point iteration is identical to the nonlocal mean filter NLM_h.

3.1 Regularization functional

The previous fixed-point iteration can be considered as finding a stationary point of the functional $J(u)$ in (4). Note that this functional does not include any fidelity term $\|u-f\|$. The interpretation of the neighborhood filters as a minimization step allows us to generalize the iteration (6) to a regularization functional similar to (1). We may add a least squares term to (4) and look for a stationary point of the corresponding functional. The nonlocal functional term is then considered as a regularization term. Following this idea, we may define the following functional:

$$L(u) = J_{1,h}(u) + \lambda\|u-f\|^2. \tag{10}$$

The condition for a stationary point can be written in fixed-point form as:

$$u(x) =: \frac{1}{C(x)}\left(\lambda f + \int_\Omega g'\left(\frac{|u(x)-u(y)|^2}{h^2}\right) u(y)w(|x-y|)dy\right) \tag{11}$$

$$C(x) = \lambda + \int_\Omega g'\left(\frac{|u(x)-u(y)|^2}{h^2}\right) w(|x-y|)dy. \tag{12}$$

A similar formula holds for the NLM-functional. Note that the first step of the fixed-point iteration with starting point $u_0 = f$ takes the general form (2) with

$$K(x, y, f) = \lambda\delta(x - y) + g'\left(\frac{|f(x) - f(y)|^2}{h^2}\right) w(|x - y|). \tag{13}$$

This filter is simply a modification of the neighborhood filter at $x = y$. This should be compared with the modification of the neighborhood filters mentioned in Remark 2.1. Indeed, the only difference between (13) and the corresponding kernel in (6) happens at $x = y$. If the value of λ is close to the mean value of $\int_\Omega g'\left(\frac{|f(x)-f(y)|^2}{h^2}\right) w(|x - y|)dy$, and if additionally $w(0) = 0$, this modification is very similar to that in Remark 2.1, but with $K(x, x, f) = \int K(x, y, f)dy$.

The functional (10) is, in general, not convex. However, for the choice $g(x) := 1 - e^{-x}$ and for sufficiently large λ the quadratic term dominates and for this case (10) convexity holds. Let us formulate this as a theorem:

Theorem 3.1. *For bounded Ω and weighting function $w \in L^\infty$ there exists a $\lambda_0(h) > 0$, such that, for all $\lambda > \lambda_0(h)$, $L(u)$ in (10) with $g(x) = 1 - e^{-x}$ is convex.*

Proof. For brevity we introduce the operator $\Delta_{x,y}u := u(x) - u(y)$. For fixed $v \in L^2(\Omega)$ the second derivative in direction v is given by

$$\frac{d^2}{dt^2}L(u+tv) = \int_{\Omega\times\Omega} \tilde{g}\left(\Delta_{x,y}u + t\Delta_{x,y}v\right) \left(\Delta_{x,y}v\right)^2 w(|x-y|)dxdy + 2\lambda \int_\Omega v(x)^2 dx,$$

where

$$\tilde{g}(\tau) = \frac{d^2}{d\tau^2}(1 - e^{-\frac{\tau^2}{h^2}}) = \frac{2}{h^2}e^{-\frac{\tau^2}{h^2}}(1 - 2\frac{\tau^2}{h^2}).$$

A minimum of $\frac{d^2}{dx^2}e^{-x^2} = 2e^{-x^2}(1 - 2x^2)$ is attained at $x = -\sqrt{\frac{3}{2}}$ with minimal value $-4e^{-\frac{3}{2}}$. Hence we can estimate

$$\begin{aligned}\frac{d^2}{dt^2}L(u + tv) &\geq -\frac{4}{h^2}e^{-\frac{3}{2}} \int_{\Omega\times\Omega} (v(x) - v(y))^2 w(|x - y|)dxdy + 2\lambda \int_\Omega v(x)^2 dx \\ &\geq \left(-\frac{4}{h^2}e^{-\frac{3}{2}} 4|\Omega|\|w\|_{L^\infty} + 2\lambda\right) \|v\|^2_{L^2}.\end{aligned}$$

For $\lambda > \frac{8}{h^2}e^{-\frac{3}{2}}|\Omega|\|w\|_{L^\infty}$ this will be positive, hence $L(u)$ is convex. □

For practical applications the previous theorem is of limited use as the lower bound on λ for which (10) is convex is too large to be used in numerical computations. For our computations we observed that a smaller value of $\lambda < \lambda_0(h)$ usually gives better results.

Let us mention, that a nonlocal functional of the form (4) with $g(t) = \arctan(t)$ and $w(s) = e^{-s^2}$ has been introduced by De Giorgi as an approximation to the Mumford-Shah functional. This relation has been studied further in [13] and in [8] for the discrete case.

Moreover, (8) bears some resemblance to the Perona-Malik equation [20]. Indeed if h is small and the weighting function w approaches the δ-distribution, the corresponding functional is approximately

$$\int_\Omega (1 - e^{-\frac{|\nabla u(x)|^2}{\kappa}})dx$$

with some parameter κ. The corresponding steepest descent flow is a variant of the Perona-Malik equation. However, this is not really similar to our approach, since it is local. We do not consider h asymptotically small. But it is worth noticing that up to a time scaling this flow can be considered a descent flow for the metric (16) with $f(x) = \nabla u(x)$ and $g(x) = 0$.

4 Functional analytic properties

The general form of the nonlocal functional (4) allows a range of possibilities for constructing a norm or a metric. Besides the fact that it is nonlocal, the resulting regularization procedure depends essentially on the filter properties determined by the choice of g and w. We give some examples of standard (semi) norms which are a special case of (4).

For $g(t) = t^p$, $1 \le p < \infty$, $w(|x-y|) = |x-y|^{-(n+\nu p)}$ with $\nu \in (0,1)$, (4) defines a seminorm, which for sufficiently smooth boundary can be made into an equivalent norm of the fractional Sobolev space $W^{\nu,p}(\Omega)$ ([1]):

$$\|u\|^p_{W^{\nu,p}} = \|u\|^p_{L^p} + J(u)$$

On the subspace of functions satisfying $\int_\Omega u(x)dx = 0$, $J(u)^{1/p}$ is in fact a norm. Choosing an increasing weighting function $w(|x-y|) = |x-y|^{(nq+\nu q-1)}$ and $g(t) = t^q$, $1 < q < \infty$, $\frac{1}{q} + \frac{1}{p} = 1$ we obtain a norm on this subspace, which is similar to a dual norm on $W^{\nu,p}$: If $\int_\Omega u(x)dx = 0$ then

$$|u|_{(W^{\nu,p})^*} = \sup_{|\phi|_{W^{\nu,p}} \le 1} \int_\Omega u(x)\phi(x)dx \le J(u)^{1/p}$$

Thus, the weight function w controls the place of $J(u)$ in a Sobolev scale. The interesting case is the choice of g and w in the case of (8) and (9). Different from the previous examples, this does not give a seminorm. Nevertheless, we can associate a metric to $J(u)$. In the following we show how the regularization term $J(u)$ can be related to a variant of the Prokhorov metric.

Let us give the definition of the Prokhorov metric in the case of random variables:

Definition 4.1. *Let X, and Y be random variables. Then the Prokhorov metric is defined as*

$$\rho(X,Y) = \inf\{h > 0 \mid h \in \mathbb{R} : \mu_X(B) \le \mu_Y(B^h) + h, \forall \text{ Borel-sets } B\}$$

where μ_X, μ_Y denotes the probability distribution of X and Y, respectively:

$$\mu_X(A) = \mathcal{P}\{X \in A\} \qquad \mu_Y(A) = \mathcal{P}\{Y \in A\},$$

and $B^h := \{x \mid \|x - y\| \leq h, y \in B\}$.

This definition can slightly be generalized to real-valued measurable functions defined on a subset of $\mathbb{R}^n$, with the Lebesgue-measure used instead of a probability measure:

Definition 4.2. *Let $f, g : \Omega \to \mathbb{R}^n$ be measurable functions, then we define the Prokhorov distance to the 0-function by*

$$\rho_0(f, 0) = \inf\{h > 0 \mid \text{meas}\{x \mid |f(x)| > h\} \leq h\} \tag{14}$$

The distance between two functions is defined as

$$\rho_0(f, g) := \rho_0(f - g, 0) \tag{15}$$

If Ω is equipped with a probability measure, then (14) coincides with Definition 4.1. However, for Definition 4.2 a probability measure is not necessary. According to [12] (15) defines a metric and it forms a complete metric space in the set of measurable functions. The Prokhorov metric has mainly been studied in the stochastic setting [4, 15]. Recently, a convergence theory for inverse problems using the Prokhorov metric has been developed [10].

The relation to the nonlocal functionals is indicated by the observation that the Gaussian is just a smooth approximation to a cut-off function:

$$\int_\Omega \left(1 - e^{-\frac{|f(x)|^2}{(2h)^2}}\right) dx \sim \int_{\{x \mid |f(x)| > h\}} dx = \text{meas}\{x \mid |f(x)| > h\}$$

Hence we may define two variants of a smoothed version of the Prokhorov metric. One is related to the $\text{YNF}_{2,h}$ filter and the other to the nonlocal mean filter NLM_h. Additionally we include a dependence on a regularization parameter α:

Theorem 4.3. *Let f, g be measurable functions, $\alpha > 0$ then*

$$\rho_{1,\alpha}(f,g) := \inf\{h > 0 \mid \int_\Omega \left(1 - e^{-\frac{|f(x)-g(x)|^2}{h^2}}\right) dx \leq \alpha h\} \tag{16}$$

$$\rho_{2,\alpha}(f,g) := \inf\{h > 0 \mid \int_\Omega \left(1 - e^{-\frac{G_\sigma * |f(x-.)-g(x-.)|^2(0)}{h^2}}\right) dx \leq \alpha h\} \tag{17}$$

define a metric. The infimum is attained at the unique number h, where equality holds in (16) and (17).

Proof. Consider the functionals

$$\rho_{1,\alpha,h}(f) := \int_\Omega \left(1 - e^{-\frac{|f(x)|^2}{h^2}}\right) dx, \ \rho_{2,\alpha,h}(f) := \int_\Omega \left(1 - e^{-\frac{G_\sigma * |f(x-.)|^2(0)}{h^2}}\right) dx$$

$\rho_{1,\alpha}(f)$ is positive and vanishes iff $f(x) = 0$ a.e. Hence the infimum in (16) is 0 iff $f(x)$ is identical 0 a.e. The same holds for for $G * |f(x-.)|^2$. If G is a positive mollifier, the term vanishes iff $|f|$ vanishes, thus we conclude that $\rho_{2,\alpha}(f,g) = 0$ iff $f = g$ a.e.

Next we have to show the triangle inequality. For $a, b \in \mathbb{R}, \nu \in [0,1]$ we start with the inequality:

$$e^{-(\nu a+(1-\nu)b)^2} \geq \min\{e^{-a^2}, e^{-b^2}\}. \tag{18}$$

This is a consequence of the fact that $\exp(-x^2)$ attains no interior minimum on a compact interval. For arbitrary h_1, h_2, set $a = \frac{f_1(x)}{h_1}, b = \frac{f_2(x)}{h_2}$, $\nu = \frac{h_1}{h_1+h_2}$. With (18) and $\exp(-x^2) \leq 1$ we obtain

$$\begin{aligned} 1 - e^{-\frac{|f_1(x)+f_2(x)|^2}{(h_1+h_2)^2}} &\leq 1 - \min\left\{e^{-\left(\frac{|f_1(x)|}{h_1}\right)^2}, e^{-\left(\frac{|f_2(x)|}{h_2}\right)^2}\right\} \\ &\leq 1 - e^{-\left(\frac{|f_1(x)|}{h_1}\right)^2} + 1 - e^{-\left(\frac{|f_2(x)|}{h_1}\right)^2}. \end{aligned}$$

If f_1,f_2 satisfy

$$\rho_{1,\alpha}(f_1, h_1) \leq \alpha h_1, \quad \rho_{1,\alpha,h_2}(f_2) \leq \alpha h_2,$$

then the previous estimate shows that

$$\rho_{1,\alpha,h_1+h_2}(f_1 + f_2) \leq \alpha(h_1 + h_2),$$

which proves the triangle inequality for the metric $\rho_{1,\alpha}$.

For $\rho_{2,\alpha}$ the triangle inequality follows in a similar manner from

$$e^{-\left[G_\sigma * |\nu f_1(x)+(1-\nu)f_2(x)|^2\right](0)} \geq e^{-\nu\left[G_\sigma * |f_1(x)|^2\right](0)+(1-\nu)\left[G * |f_2(x)|^2\right](0)},$$

and (18).

By basic calculus $\rho_{1,\alpha,h}(f), \rho_{2,\alpha,h}(f)$, is strictly monotonically increasing in h unless $f = 0$. Since αh is increasing, by the mean value theorem there exists an h where equality holds in the definition of (16) and (17), moreover this is precisely the number where the infimum is attained. □

Note that for any choice of $\alpha > 0$ there exists a corresponding h such that equality holds in (16) and (17). The opposite is also true: For a given h we can find a $\alpha(h)$ leading to equality in the definition. The later is more convenient, as we can consider h as a regularization parameter and $\alpha(h)$ be implicitly defined by (16) and (17). Using h as parameter we can consider the metric $\rho_{i,\alpha}$ on the space of measurable functions defined on $\Omega \times \Omega$: this allows us to rewrite the functionals (8), (9) as

$$J_{i,h}(u) = \alpha(h)\rho_{i,\alpha(h)}\left(u(x) - u(y), 0\right).$$

This shows the similarity of the neighborhood filter with a regularization functional. In fact (10) takes the form of a regularization functional

$$L(u) = \lambda\|u - f\|^2 + \alpha(h)\ \rho_{i,\alpha(h)}\left(u(x) - u(y), 0\right),$$

with an error term and a regularization term defined via the Prokhorov-like metric of a nonlocal term $u(x)-u(y)$. The equivalence to standard regularization method is, however, not complete. Given h the corresponding regularization parameter $\alpha(h)$ will also depend on u.

5 Nonlocal BV

The functionals in the previous section do not take into account the gradient of the image f. This might suffice for denoising purposes, where the problem is not ill-posed in the space L^2. For the deblurring case this is a drawback, since adding a Prokhorov term will in general not regularize the problem in a reasonable space. In fact, the set of functions, which are bounded in the Prokhorov metric is not compact in L^2. Hence, if the problem is ill-posed a functional involving gradients will do better. For image deblurring it is desirable to avoid blurring of edges, so a good choice is to use nonlocal functionals derived from the space of bounded variations. Following the ideas of Section 3 we therefore propose the following nonlocal regularization term: for $u \in W^{1,1}(\Omega)$ we define

$$\mathrm{NLBV}(u) := \int_{\Omega\times\Omega} |\nabla u(x) - \nabla u(y)| dxdy \tag{19}$$

or

$$\mathrm{NLBV}_\tau(u) := \int_{\{(x,y)\in\Omega\times\Omega \mid |x-y|\leq\tau\}} |\nabla u(x) - \nabla u(y)| dxdy \tag{20}$$

If u is merely in $L^1(\Omega)$, we can extend the previous definition by duality similar to the definition of the BV-seminorm we have:

$$\int_{\Omega\times\Omega} |\nabla u(x) - \nabla u(y)| dxdy = \sup_{\substack{\phi\in C_0^\infty(\Omega\times\Omega)\\ \|\phi\|_\infty\leq 1}} \int_{\Omega\times\Omega} (\nabla u(x) - \nabla u(y))\, \phi(x,y) dxdy.$$

An integration by parts leads to the generalization or definitions (19) and (20) to functions $u \in L^1(\Omega)$:

$$\mathrm{NLBV}(u) := \sup_{\phi\in C_0^\infty(\Omega\times\Omega), \|\phi\|_\infty\leq 1} \int_{\Omega\times\Omega} u(x)\left(\nabla_x\phi(x,y) - \nabla_x\phi(y,x)\right) dxdy \tag{21}$$

and

$$\mathrm{NLBV}_\tau(u) = \sup_{\phi\in X, \|\phi\|_\infty\leq 1} \int_{\Omega\times\Omega} u(x)(\nabla_x\phi(x,y) - \nabla_x\phi(y,x)) dxdy \tag{22}$$

where $X = \{\phi \in C_0^\infty(\{(x,y) \in \Omega\times\Omega \mid |x-y| \leq \tau\})\}$. The corresponding space $\mathrm{NLBV}(\Omega)$ is defined as (using (21))

$$\mathrm{NLBV}(\Omega) := \{u \in L^1(\Omega) \mid \mathrm{NLBV}(u) < \infty\}.$$

This space shares many properties with the space of functions of bounded variation. It includes discontinuous functions just as the space $\mathrm{BV}(\Omega)$ (see Lemma 5.2). By adding the L^1-norm to the $\mathrm{NLBV}(\Omega)$-functional we end up with a norm, making $\mathrm{NLBV}(\Omega)$ a Banach space (compare Lemma 5.3). (A similar conclusion applies to the NLBV_τ-functional). However, due to the nonlocal form, the functional also measures similarities in an image, which are spatially separated.

We can as well consider more general functionals of the same type:

$$F(u) := \int_{\Omega\times\Omega} g(|\nabla u(x) - \nabla u(y)|^2) w(x-y) dx dy, \tag{23}$$

where g is a nonnegative differentiable function, and w a positive weight function. For the sake of generality we state the Euler-Lagrange equation for the denoising functional with (23) as regularization term:

$$J(u) = \lambda \|u - f\|_{L^2}^2 + F(u).$$

A minimizer of $J(u)$ satisfies

$$\lambda(u(x) - f(x)) = 2\nabla \left(\int_\Omega g'(|\nabla u(x) - \nabla u(y)|^2)(\nabla u(x) - \nabla u(y)) w(x-y) dy \right). \tag{24}$$

For the case of the nonlocal BV-functional $g'(x) = \frac{1}{\sqrt{x}}$. Another important choice for g in (23) comes up at the numerical computations where the nondifferentiable function $\frac{1}{\sqrt{x}}$ is replaced by the well-known ϵ-regularization $g = \frac{1}{\sqrt{x+\epsilon}}$.

One of the advantages of the NLBV-functional is that it does not suffer from the staircase effect, since linear functions are preferred over piecewise constant ones. If is obvious, that $\mathrm{NLBV}(g) = 0$ and $\mathrm{NLBV}_\tau(g) = 0$, if g is an affine function. However, the functional does not vanish for piecewise constant functions, hence we expect, that a regularization procedure with the NLBV-seminorm will prefer affine functions over piecewise constant one, which should avoid staircasing. The numerical examples in Section 6 will show this.

For further analysis we need lower semicontinuity and approximation properties of the NLBV-functionals:

Lemma 5.1. *Let $u \in L^1(\Omega)$ and $u_n \in W^{1,1}(\Omega)$ be a sequence converging to u either in the L^1-norm topology or weakly in L^2, then*

$$\mathrm{NLBV}(u) \le \liminf_{n\to\infty} \mathrm{NLBV}(u_n) \qquad \mathrm{NLBV}_\tau(u) \le \liminf_{n\to\infty} \mathrm{NLBV}_\tau(u_n).$$

Moreover for any $u \in L^1(\Omega)$ with $\mathrm{NLBV}(u) < \infty$ (or $\mathrm{NLBV}_\tau(u) < \infty$) a sequence of C^∞-functions u_n exists converging in L^1 to u and which satisfy

$$\lim_{n\to\infty} \mathrm{NLBV}(u_n) = \mathrm{NLBV}(u) \quad (\textit{or } \lim_{n\to\infty} \mathrm{NLBV}_\tau(u_n) = \mathrm{NLBV}_\tau(u).)$$

Proof. The proof is analogue to the BV-case (cf. [2, Thm. 3.9]) □

The functional NLBV can be interpreted as an equivalent norm on a suitable quotient space of BV. In fact, we have the following lemma:

Lemma 5.2. *Let Ω be a bounded extension domain, $u \in \mathrm{NLBV}(\Omega)$ then*

$$|\Omega| \inf_{g(x) \text{ affine}} |u-g|_{\mathrm{BV}} \leq \mathrm{NLBV}(u) \leq 2|\Omega| \inf_{g(x) \text{ affine}} |u-g|_{\mathrm{BV}}. \tag{25}$$

Proof. Define the set $\tilde{X} := \{\psi \in C_0^\infty(\Omega\times\Omega) \mid \|\phi\|_\infty \leq 1\}$. Let g be an arbitrary affine function $g(x) = hx + c$, $h, c \in \mathbb{R}$, then

$$\begin{aligned}
\mathrm{NLBV}(u) \quad &= \mathrm{NLBV}(u-g) \\
&\leq \sup_{\phi,\psi\in\tilde{X}} \int_{\Omega\times\Omega} (u(x)-g(x))\left(\nabla_x\phi(x,y) - \nabla_x\psi(y,x)\right) dxdy \\
&\leq 2 \sup_{\phi\in\tilde{X}} \int_\Omega (u(x)-g(x))\, \nabla_x \int_\Omega \phi(x,y)dydx \\
&\leq 2 \sup_{\eta(x)=\int\phi(x,y)dy,\phi\in\tilde{X},} \int_\Omega (u(x)-g(x))\,\nabla_x \frac{\eta(x)}{\|\eta\|_\infty} dx\, \|\eta\|_\infty \\
&\leq 2|u-g|_{BV}|\Omega|,
\end{aligned}$$

where the last inequality follows from $|\int_\Omega \phi(x,y)dy| \leq |\Omega|$ for $\phi \in \tilde{X}$ and the definition of the BV-seminorm. Since g was arbitrary the right hand side of (25) follows.

For the left hand side let $u \in W^{1,1}(\Omega)$, take $h := \frac{1}{|\Omega|}\int_\Omega \nabla u(y)dy$, and $g(x)$ as the affine function $g(x) = hx + c$.

$$\begin{aligned}
|u(x)-g(x)|_{\mathrm{BV}} \quad &= \int_\Omega \left|\frac{1}{|\Omega|}\int_\Omega \nabla u(x) - \nabla u(y)dy\right|dx \\
&\leq \frac{1}{|\Omega|}\int_\Omega\int_\Omega |\nabla u(x)-\nabla u(y)|dydx = \frac{1}{|\Omega|}\mathrm{NLBV}(u).
\end{aligned} \tag{26}$$

For general $u \in \mathrm{NLBV}(\Omega)$ by Lemma 5.1 we can find a sequence of smooth functions $u_n \in C^\infty(\Omega)$ with $\lim_{n\to\infty}\|u-u_n\|_{L^1} = 0$, $\lim_{n\to\infty}\mathrm{NLBV}(u_n) = \mathrm{NLBV}(u)$. With $g_n(x) := \frac{1}{|\Omega|}\int_\Omega \nabla u_n(y)dy\, x$ and (26) we find that $|u_n(x) - g_n(x)| \leq \frac{1}{|\Omega|}\mathrm{NLBV}(u_n) \leq C_1$ for some constant C_1. According to the Poincaré-inequality for BV ([2]) for each n we can find constants m_n such that

$$\|u_n(x) - g_n(x) - m_n\|_{L^1} \leq C_2|u_n(x)-g_n(x)|_{\mathrm{BV}} \leq C_3.$$

This means that the sequence $u_n(x) - g_n(x) - m_n$ is uniformly bounded in BV, and by compact embedding [2] has a convergent subsequence in L^1. But since u_n converges in L^1 the sequence of affine functions $\tilde{g}_n(x) := g_n(x) - m_n$ must have a L^1 convergent subsequence $\lim_{n'}\tilde{g}_{n'} =: g(x)$, which limit has to be an affine function. By weak lower semicontinuity we get

$$|u(x)-g(x)|_{BV} \leq \liminf_{n\to\infty} |u_n - \tilde{g}_n|_{BV} = \frac{1}{|\Omega|}\liminf_{n\to\infty}\mathrm{NLBV}(u_n) = \frac{1}{|\Omega|}\mathrm{NLBV}(u).$$

This proves (25). □

Although Lemma 5.2 establishes an equivalence of norms, this does not necessarily mean that the results for denoising with these norms as regularization are equivalent. Since in general convergence of a denoising algorithm for functionals like (1) (e.g as $\lambda \to \infty$) happens only in the weak-$*$-topology and not in the norm topology. This phenomenom can also be observed when comparing denoising with an anisotropic BV norm (see [11]) to the usual one. Although in this case the norms are equivalent the computational result are significantly different. Moreover, a denoising algorithm using the quotient norm in (25) seems to be more involved than the corresponding NLBV-minimization.

The result of Lemma 5.2 shall be compared to other seminorms, which also annihilate affine functions. One possible candidate would be a generalization of the BV-functional using the second derivative, for instance, $\|\nabla^2 u\|_{L^1}$. This functional is, however, much stronger, as it uses second derivatives, whereas NLBV does not. This difference is also obvious in the embedding theorems - at least in one dimension. Note that a higher order Sobolev space using $\|\nabla^2 u\|_{L^1}$ (such as $W^{2,1}(\mathbb{R})$) only contains continuous functions, while NLBV allows discontinuous ones as well.

For the NLBV$_\tau$-functional the equivalence to a quotient norm is not uniform in τ as the constants in such an inequality depend strongly on τ, and a similar inequality does not hold in the limiting case $\tau \to 0$. An illuminating example is the case of the functional (23), with $g(x) = x$, $\Omega = \mathbb{R}^n$, and w a suitable weight function with $\int w(x)dx = 1$. This can be seen as a nonlocal form of the $H^1(\mathbb{R})$-norm, and it can be completely solved by Fourier analysis. Since the Euler-Lagrange equation only involves derivatives and convolution with w in Fourier space (24) takes the form

$$\lambda \left(\hat{u}(\xi) - \hat{f}(\xi) \right) = -2|\xi|^2 \left(1 - \hat{w}(\xi)\right) \hat{u}(\xi),$$

hence

$$\hat{u}(\xi) = \frac{\hat{f}(\xi)}{\lambda + 2|\xi|^2(1 - \hat{w}(\xi))}. \tag{27}$$

For example, if w is taken as a Gaussian this filters high frequencies, just the same as a H^1 regularization, since $\hat{w}(\xi) \sim 0$ for ξ large. However, the low frequencies are filtered less, because $1 - \hat{w}(\xi) \sim 0$ for ξ small. This shows that the nonlocal functionals with a weight function act as an additional low-pass filter, while not altering the high frequency behavior.

In the extremal case $w \sim 1$, which corresponds to the NLBV-functional, $\hat{w}$ converges to the δ-distribution and $\mathrm{supp}(\hat{w}) \to \{0\}$. As can be seen from (27) no low-pass filtering happens at all in this case. Since there are no affine functions in $H^1(\mathbb{R})$, the corresponding quotient space is equivalent to the usual Sobolev space $H^1(\mathbb{R})$. However, including a weight w gives a norm which is not equivalent to $H^1(\mathbb{R})$. This indicates, that the NLBV$_\tau$ might be weaker then the quotient norm.

The next issue is to prove that the denoising functional with the NLBV-seminorm has a solution in BV. For this we need a Poincaré-type inequality. We assume that Ω is a Sobolev extension domain, i.e. for all Sobolev spaces $W^{k,p}(\Omega)$ a linear and bounded operator exists, which extends functions u from $W^{k,p}(\Omega)$ to $W^{k,p}(\mathbb{R}^n)$. Any domain with Lipschitz boundary is a Sobolev extension domain. For geometric conditions on Ω to be a extension domain we refer to [16].

Lemma 5.3. *Let Ω be a bounded Sobolev extension domain and let $u \in W^{1,1}(\Omega)$, then there exists a constant $C(\Omega)$ such that*

$$\int_\Omega |\nabla u(x)| dx \le C(\Omega) \left(\int_{\Omega\times\Omega} |\nabla u(x) - \nabla u(y)| dx dy + \|u\|_{L^1(\Omega)} \right).$$

Proof. By a suitable normalization we may assume that $\int_\Omega |\nabla u(x)| dx = 1$. Take $0 < \epsilon < 1$ arbitrarily but fixed.

First we consider the case $\int_{\Omega\times\Omega} |\nabla u(x) - \nabla u(y)| dx dy \ge \epsilon$, then we obtain by the normalization condition

$$\int_\Omega |\nabla u(x)| dx \le \frac{1}{\epsilon} \int_{\Omega\times\Omega} |\nabla u(x) - \nabla u(y)| dx dy. \tag{28}$$

For the case $\int_{\Omega\times\Omega} |\nabla u(x) - \nabla u(y)| dx dy \le \epsilon$ we define $\vec{a} = \frac{1}{|\Omega|} \int_\Omega \nabla u(x) dx$ and get

$$\begin{aligned}
\int_\Omega |\nabla u(x) - \vec{a}| dx &= \int_\Omega |\nabla u(x) - \frac{1}{|\Omega|} \int_\Omega \nabla u(y) dy| dx \\
&= \frac{1}{|\Omega|} \int_\Omega | \int_\Omega \left(\nabla u(x) - \nabla u(y) \right) dy | dx \\
&\le \frac{1}{|\Omega|} \int_\Omega \int_\Omega |\nabla u(x) - \nabla u(y)| dy dx \le \frac{\epsilon}{|\Omega|}.
\end{aligned}$$

Now define the affine linear function $v(x) := \vec{a}x + d$ with $d \in \mathbb{R}$, then $\nabla v(x) = \vec{a}$ and by the previous inequality we get

$$\begin{aligned}
|\vec{a}| &= | \int_\Omega \nabla v(x) dx | = \int_\Omega |\nabla v(x)| dx = \int_\Omega |\nabla u(x) + \left(\nabla v(x) - \nabla u(x) \right)| dx \\
&\ge \int_\Omega |\nabla u(x)| - \int_\Omega |\nabla v(x) - \nabla u(x)| dx \ge 1 - \frac{\epsilon}{|\Omega|}.
\end{aligned}$$

Now let us consider the case $\Omega = [0,1]^n$, hence $|\Omega| = 1$. In the definition of v the constant d can be chosen such that $\int_\Omega u(x) - v(x) dx = 0$. In this case the Poincaré inequality [1] yields

$$\int_\Omega |u(x) - v(x)| dx \le C_0 \int_\Omega |\nabla u(x) - \nabla v(x)| dx \le C_0 \epsilon$$

On the other hand, since v is a linear function there exists a positive constant C_1, depending only on Ω, with $\|v\|_{L^1(\Omega)} \ge C_1 \|\nabla v\|_{L^1(\Omega)} \ge C_1(1-\epsilon)$. This implies that for ϵ sufficiently small

$$\|u\|_{L^1(\Omega)} \ge \|v\|_{L^1(\Omega)} - \|u - v\|_{L^1(\Omega)} \ge C_1(1-\epsilon) - C_0\epsilon \ge \frac{C_1}{2}.$$

With this ϵ we obtain

$$\int_\Omega |\nabla u(x)|dx \le \frac{2}{C_1}\|u\|_{L^1(\Omega)}. \tag{29}$$

The choice of ϵ only depends on the constants C_0, C_1, which only depend on Ω, hence (28) and (29) imply the result for $\Omega = [0,1]^n$.

Now let Ω be an extension domain. For any $u(x), v(x) \in W^{1,1}(\Omega)$ there exist extensions $U(x), V(x)$ to $\mathbb{R}^n$ such that

$$\begin{aligned}\int_{\mathbb{R}^n} |\nabla U(x)|dx &\le C_\Omega \int_\Omega |\nabla u(x)|dx \\ \int_{\mathbb{R}^n} |\nabla U(x) - \nabla V(x)|dx &\le C_\Omega \int_\Omega |\nabla u(x) - \nabla v(x)|dx.\end{aligned}$$

Proceeding as before we assume the case $\int_{\Omega\times\Omega} |\nabla u(x) - \nabla u(y)|dxdy \le \epsilon$. Consider S a n-dimensional cube which contains Ω. By appropriate scaling we can assume that $S = [0,1]^n$. Define $\vec{a} = \frac{1}{|\Omega|}\int_\Omega \nabla u(x)dx, v(x) = ax + d$ and denote by $V(x)$ its extension to $\mathbb{R}^n$. Since the extension operator is linear, we can find a constant d such that $\int_S U(x) - V(x)dx = 0$. With this constant the Poincaré inequality can be applied on S, hence

$$\begin{aligned}\int_\Omega |u(x) - v(x)|dx &\le \int_S |U(x) - V(x)|dx \le C_0 \int_S |\nabla U(x) - \nabla V(x)|dx \\ &\le C_0 C_\Omega \int_\Omega |\nabla u(x) - \nabla v(x)|dx \le C_0 C_\Omega \epsilon.\end{aligned}$$

Again we can estimate

$$\begin{aligned}\|u\|_{L^1(\Omega)} &\ge \|v\|_{L^1(\Omega)} - \|u - v\|_{L^1(\Omega)} \\ &\ge \tilde{C}_\Omega \|a\| - C_0 C_\Omega \epsilon \ge \tilde{C}_\Omega \left(1 - \frac{\epsilon}{|\Omega|}\right) - C_0 C_\Omega \epsilon > \frac{1}{2}\tilde{C}_\Omega.\end{aligned}$$

for ϵ sufficiently small and some generic constant $\tilde{C}_\Omega$ only depending on the domain. Hence in this case

$$\int_\Omega |\nabla u(x)|dx \le \frac{2}{\tilde{C}_\Omega}\|u\|_{L^1(\Omega)}.$$

Now ϵ depends only on the generic constants so the result for the general case follows using (28). □

Sobolev extension domains can be quite complicated, for instance the von Koch snowflake has this property. For more details on the Poincaré-inequality see [5, 7]. Let us remark that the same proof holds if we only assume that the L^1-Poincaré inequality holds for Ω.

We are now in the position to prove existence of a minimizer for the denoising functional with the NLBV and NLBV_τ-regularization: For $\lambda, \tau > 0$ let us define the functionals for a given image f:

$$\begin{aligned}J_1(u) &:= \lambda\|u - f\|_{L^2}^2 + \text{NLBV}(u) &\quad (30)\\ J_2(u) &:= \lambda\|u - f\|_{L^2}^2 + \text{NLBV}_\tau(u) &\quad (31)\end{aligned}$$

Theorem 5.4. *Let Ω be a bounded Lipschitz domain and $f \in L^2(\Omega)$, $\lambda, \tau > 0$. Then each of the functionals* (30), (31) *attains a unique minimum in* $\mathrm{BV}(\Omega)$.

Proof. Let $u_n \in L^2(\Omega)$ be a minimizing sequence for either $J_1(u)$ or $J_2(u)$. It follows that a constant C_1 exists, such that $\|u_n\|_{L^2} \leq C_1$. Consequently u_n has a weakly convergent subsequence $u_{n'}$ in $L^2(\Omega)$ with limit $u \in L^2(\Omega)$. By the weak lower semicontinuity of the L^2-norm and Lemma 5.1 u has to be a minimum for (30) or (31), respectively. If u is a minimizer of J_1 we can conclude directly from Lemma 5.3 that $u \in BV(\Omega)$, as $\|u\|_{L^1} \leq C(\Omega)\|u\|_{L^2} \leq C_3$ and $\mathrm{NLBV}(u) \leq C$ holds. For the case that u is a minimizer of J_2, we take an arbitrary point $x_0 \in \bar{\Omega}$ and $\rho \leq \frac{\tau}{2}$ fixed. Consider the sets $Z_\rho(x_0) := B_\rho(x_0) \cap \Omega$. If $\partial\Omega$ is Lipschitz, then these sets are Sobolev extension domains and the Cartesian product $Z_\rho(x_0) \times Z_\rho(x_0)$ is contained in $\{(x,y) \in \Omega \times \Omega \| x - y| \leq \tau\}$, hence

$$\int_{Z_\rho(x_0) \times Z_\rho(x_0)} |\nabla u(x) - \nabla u(y)| dx dy \leq \mathrm{NLBV}_\tau(u) \leq C.$$

This estimate also holds for the weak definition (21) on the left hand side with $\Omega = Z_\rho(x_0)$. Moreover since $\|u\|_{L^1(Z_\rho(x_0))} \leq C(\tau)\|u\|_{L^2(Z_\rho(x_0))} \leq C(\tau)\|u\|_{L^2(\Omega)}$ Lemma 5.3 implies that $|u|_{BV(Z_\rho(x_0))} \leq C(\tau)$ for any x_0 and $\rho \leq \frac{\tau}{2}$. As $\bar{\Omega}$ is compact, it can be covered by finitely many sets $\Omega = \bigcup_{i=1}^{N(\tau)} Z_{\rho_i}(x_i)$. By means of a partition of unity [14] we can decompose any function $\phi \in C_0^\infty(\Omega)$ such that $\phi = \sum_{i=1}^{N} \phi_i(x)$ and ϕ_i is supported in $Z_{\rho_i}(x_i)$ and $\|\phi_i\|_\infty \leq \|\phi\|_\infty$. Hence

$$\begin{aligned} \int_\Omega u(x) \nabla.\phi(x) dx &= \sum_{i=1}^{N(\tau)} \int_{Z_{\rho_i}(x_i)} u(x) \nabla.\phi_i(x) dx \\ &\leq N(\tau) \max_i |u|_{BV(Z_{\rho_i}(x_i))} \leq C(\tau), \end{aligned}$$

with some constant $C(\tau)$ only depending on τ and Ω. This finally proves that $u \in BV(\Omega)$. The uniqueness follows from the convexity of NLBV, and NLBV_τ-functionals and the strict convexity of $\|u - f\|_{L^2}^2$. □

For simplicity we proved Theorem 5.4 for Lipschitz domains, but it is also true for extension domains for Sobolev spaces.

We may as well use the nonlocal BV-functional for solving ill-posed problems such as deblurring. Let K be a blurring operator, and f a given blurred image. The problem reduces to solving the operator equation

$$Ku = f.$$

One possibility to regularize this equation is to minimize the functional

$$T(u) = \lambda\|Ku - f\|^2 + \mathrm{NLBV}(u). \tag{32}$$

The following theorem proves existence and uniqueness for (32):

Theorem 5.5. *Let Ω be a Lipschitz domain. Let $K : L^1(\Omega) \to L^2(\Omega)$ be a linear and continuous operator. Moreover, suppose that the following conditions hold:*

$$K(ax+b) = 0 \Leftrightarrow (a,b) = (0,0) \quad \forall a,b \in \mathbb{R} \tag{33}$$

Then (32) *has a unique minimum in* $\mathrm{BV}(\Omega)$*.*

Proof. A minimizing sequence $u_n \in L^1(\Omega)$ is uniformly bounded in the sense that constants C_1, C_2 exists with $\mathrm{NLBV}(u_n) \leq C_1$ and $\|Ku_n\|_{L^2} \leq C_2$. From (25), the Poincaré-inequality for BV, and the compact embedding of BV into L^1, a sequence of affine functions $g_n'(x)$ and a subsequence $u_{n'}$ of u_n exists such that $u_{n'} - g_{n'}$ converges in L^1. By continuity $K(u_{n'} - g_{n'})$ converges in L^2. From $\|Ku_n\|_{L^2} \leq C_2$ it follows that $\|Kg_{n'}\|_{L^2}$ is uniformly bounded. Now the vector space of affine function can be identified with $\mathbb{R}^2$. In particular all norms on this space are equivalent. From the nondegenerarcy condition (33), $(ax+b) \to \|K(ax+b)\|_{L^2}$ and $(ax+b) \to \|ax+b\|_{L^1}$ define equivalent norms on this finite dimensional space. Since $Kg_{n'}$ is bounded and by the finite dimensionality we find a subsequence $g_{n''}$ converging in L^1 to some affine function g. Hence also $u_{n''}$ converges in L^1 to some $u \in L^1$ and by continuity $Ku_{n''} \to Ku$ in L^2. From the semicontinuity of the L^2-norm and Lemma 5.1 u is a minimizer, which is unique by strict convexity. u is in BV by Lemma 5.3 and Lemma 5.1. □

The hypothesis (33) is a generalization of the analogue condition for BV-regularization, where only $Kb = 0 \Leftrightarrow b = 0$ is needed (compare [9]). If we restrict ourselves to dimensions $n = 1$ or $n = 2$, the condition $K : L^1(\Omega) \to L^2(\Omega)$ can be replaced by K being linear and continuous beween L^2 (similar as in [9]). Instead of the compact embedding of BV $\to L^1$, the continuous embedding BV $\to L^2$ and weak compactness can be used to prove existence of a minimizer.

5.1 G-norm

As it was done for the BV-norm we can look at a corresponding dual G-norm. Following [17] we can define a dual norm for the NLBV and NLBV_τ functionals (using their definition in weak form). Since we are interested in the case of a bounded domain, we have to include Neumann-boundary conditions as in [3]: Denote by N the outward normal of Ω then for $f(x) \in L^2$

$$\begin{aligned} \|f\|_{*,nl} \quad := \quad & \inf\{\|\phi\|_{L^\infty(\Omega\times\Omega)^n}\}, \text{ where} \\ & f = \nabla\psi, \\ & \psi.N = 0 \text{ on } \partial\Omega \text{ and } \psi = \int_\Omega \phi(x,y) - \phi(y,x)dy. \end{aligned} \tag{34}$$

And similarly for the NLBV_τ, where the infimum is taken over $L^\infty(X)^n$ functions as in (22).

Just as in [17] a minimizer of (30) with $f \in L^2(\Omega)$, can be characterized in the following way: If $\|f\|_{*,nl} \leq \frac{1}{2\lambda}$, then a minimizer u vanishes: $u = 0$. If $\|f\|_{*,nl} > \frac{1}{2\lambda}$, then a minimizer is characterized by $f = u + v$, where

$$\|v\|_{*,nl} = \frac{1}{2\lambda} \qquad \int_\Omega u(x)v(y) = \frac{1}{2\lambda}\text{NLBV}(u).$$

In view of the equivalence of the nonlocal functional to a quotient norm (Lemma 5.2), this G-norm is equivalent to a corresponding one in a dual quotient space. This leads to the definition

$$\|f\|_* := \inf\{\|\xi\|_{L^\infty(\Omega,R^2)} \,|\, f = \text{div}\xi, \quad \xi.N = 0 \text{ on } \partial\Omega, \quad \int_\Omega \xi(x)dx = 0\} \quad (35)$$

If follows if $\|f\|_* < \infty$, then $\int_\Omega f(x)dx = 0$ and $\int_\Omega xf(x)dx = 0$. The difference to the definition in [3] is the additional condition $\int_\Omega \xi(x)dx = 0$. If follows fairly easily that this yields an equivalent norm to $\|f\|_{*,nl}$: If ξ satisfies the conditions in (35) then $\phi(x,y) := \frac{\xi(x)-\xi(y)}{2}$ can be taken for (34). On the other hand, if ψ is as in the definition of $\|f\|_{*,nl}$, then it is in the set (35) and $\|\psi\|_{L^\infty} \leq 2|\Omega|\|\phi\|_{L^\infty}$.

5.2 Bregman iteration

In [18] a significant improvement of the bounded-variation regularization was introduced. Instead of the BV-seminorm the so-called Bregman distance is used. For a convex functional $J(u)$ the corresponding Bregman distance is defined as

$$D(u,v) = J(u) - J(v) - \langle \partial J(v), u - v\rangle.$$

Here $\partial J(u)$ denotes the subgradient of $J(u)$ at u. The idea behind the Bregman regularization is to use an iterative procedure, where u_k is defined as minimization of

$$\lambda\|u - f\|^2 + D(u, u_{k-1}).$$

Compared to the BV-regularization, this approach shows better results as more signal will remain in the cleared image. It has proven in [18] that the sequence u_k converges to f. For denoising, the iteration is stopped as soon as the residue is of the order of the noise level. The Bregman iteration can be implemented similar to BV-denoising but gives a better result. Moreover, it is a quite general procedure that can be used for any convex functional.

Contrary to the case of neighborhood filters, the nonlocal BV functional is convex and hence, the Bregman iteration can be defined. The idea is to replace the BV-seminorm by the Bregman distance. This yields the following algorithm (see [18]):

1. Set $u_0, v_0 = 0$
2. Define u_k by

$$u_{k+1} = \text{argmin}_u \lambda\|u - f - v_k\|^2 + \text{NLBV}(u)$$

3. $v_{k+1} = v_k + f - u_{k+1}$

The iteration is stopped, as soon as the residue $\|u_k - f\| \leq \delta$. This procedure is expected to give a more detailed image, with less structure in the remaining noise $u_k - f$. In the next section we present the numerical results.

6 Numerical Results

In this section we present some numerical results for the neighborhood filters and nonlocal functionals. At first we consider the denoising problem in one dimension. The exact solution shown in Figure 1 is a piecewise constant and piecewise linear function. To this we added uniformly distributed random noise with a signal to noise ratio SNR ~ 8 (Fig. 2). We computed the result with a discretization of 200 points in the unit interval $[0, 1]$. In Figure 3 and Figure 4 we show the denoised version for the ROF-method and the nonlocal tv-method. The results were computed by an explicit Euler method for the steepest descent flow of (1) with $\lambda = 160$ and (31) with $\lambda = 120$ and $\tau = 60$ pixels. It can be seen clearly, that the nonlocal-tv method does not suffer from staircasing and performs better than BV. However, in the result for the NLBV-functional (Fig. 4) the corner of the graph are cut, e.g. at the point $x \sim 0.6$. To some extent this can be expected as linear pieces in the graph are preferred due to (25). On the other hand the discontinuity at $x \sim 0.4$ is clearly preserved (in contrast to a similar functional using second order derivatives).

In Figures 5-10 we present the results for an exact solution (Fig. 5) which has a cusp, with the same parameters as in the first example. As it can be seen from the picture the solution for the nonlocal BV regularization is smoother compared to the BV-case. Figure 9 shows the result for the Bregman iteration after 3 steps for the NLBV-functional and Figure 10 shows the result after 3 steps of the Bregman iteration for the BV-case. The Bregman iteration for NLBV gives a better approximation to the exact solution, however, in this case we observe difficulties to approximate the cusp. Compared to the analogous result for BV, the nonlocal version appears to have fewer steps. The Bregman iteration for the first problem with exact solution in Figure 1 essentially does not improve from the result in Figure 4, since the residual is already in the range of the noise level.

Let us now treat the two dimensional denoising problem: Figure 11 shows the noisy image of pixel size 512×512 which was created by adding random noise to the pixel value and which is uniformly distributed on the interval [-20,20]. This results in a signal to noise ratio of about 12.

Figure 12 shows the denoised version u and the residual $f - u$ using the Buades-Coll-Morel nonlocal mean filter (3). For this result we used $h^2 \sim 120$, and a mollifier G_σ whose support has the size of 9 pixels. For the numerical computation we restricted the integral in (2) to all pixel y within a 30×30 window of x. It can be seen, that the residual hardly contains any texture.

For Figure 13 we use the nonlocal BV method. We compute the image by

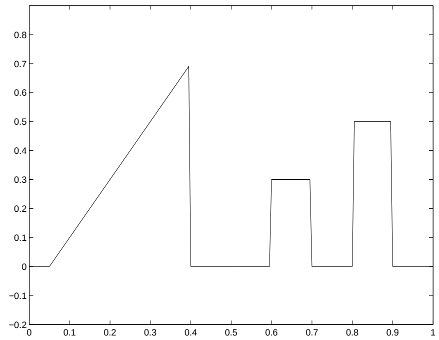

Figure 1: Exact Solution

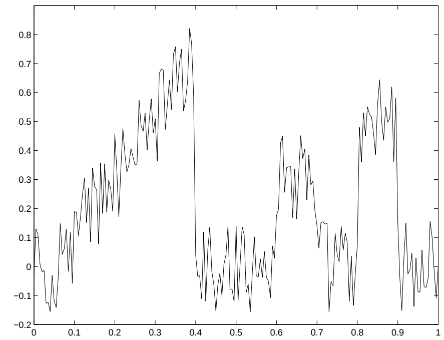

Figure 2: Noisy Data

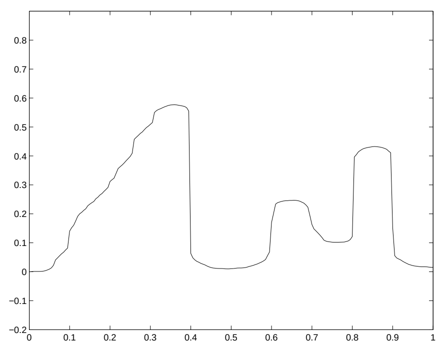

Figure 3: Solution for BV

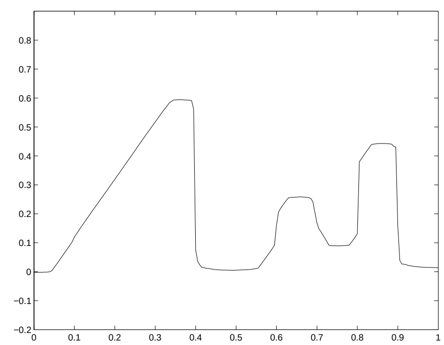

Figure 4: Solution for nonlocal BV

solving the Euler-Lagrange equation for the functional

$$J(u) = \lambda \|u - f\|^2 + \int_{\Omega\times\Omega} \sqrt{(\nabla u(x) - \nabla u(y))^2 + \epsilon^2} dxdy$$

with $\lambda = 0.1$, $\epsilon = \frac{1}{512}$

It can be seen that the nonlocal BV version has no staircasing effect. However, comparing it to the nonlocal mean filter we observe that more texture remains in the residual. This might come from the use of gradients for the nonlocal BV-functionals.

In Figure 14 we show the first three iterations for the Bregman iteration for the NLBV-functional. The image size was 256×256 in this case and we choose $\lambda = 0.01$. In this case the procedure can be stopped after the second iteration. It can be seen, that the Bregman step from the first to the second image improves the quality, by adding more details to u.

In Figures 15–18 we show the result for denoising applied to a textured image. The image size was 128×128 pixel. Figure 15 displays the noisy image, which was created by adding uniformly distributed noise resulting in a signal-to-noise ration SNR~ 9.

Fig. 16 shows the result for the Nonlocal Mean filter (3) with the same parameter setting as for Fig. 12. Fig. 17 and Fig. 18 show the result for the BV-functional (1) and the NLBV_τ functional, where λ was chosen the same for both functionals $\lambda = 0.05$, and $\tau = 9$ pixels. It can be seen that the nonlocal BV-version is better than pure bounded variation regularization. Compared to the nonlocal mean filter both the BV and the NLBV functionals do smooth too much.

The next results in Figures 19–22 are the analogous result for a noisy moon image, with similar settings as before (image size 128×128, SNR~ 8, $\lambda = 0.09$ for BV and NLBV). There is very few difference between the results, but the nonlocal BV gives smoother shadows than the BV-denoising. Still, the nonlocal mean filter performs very well.

In the next figures (Fig. 23-28) we show the results for the deblurring problem. Figure 23 shows the data, which are computed by blurring the original Lenna image with a polynomial smoothing kernel

$$Kf = \int k(x-y)f(y)dy \qquad k(s) = \frac{1}{\alpha}\begin{cases} 1 - \frac{x^2}{\alpha} & x^2 \leq \alpha \\ 0 & \text{else} \end{cases}$$

and $\alpha = 0.001$. For all the deblurring problem the image size was 256×256. Figure 24 compares the results of the nonlocal BV-method on the left with the usual BV-regularization on the right. The computation was done by an explicit Euler method for the steepest descent flow for the functional (32) (but with the NLBV_τ and the BV-functional instead of NLBV.) The regularization parameter λ was $\lambda = 50$ for BV and $\lambda = 5$ for the nonlocal BV, with $\tau = 9$. It is obvious, that the nonlocal-BV method gives smoother result whereas BV tends to blocky structures. Moreover the nonlocal variant does preserve edges as well as BV does. However, there are some artefacts visible around the main edges for the nonlocal version. In Figures 25-28 we present another result for the deblurring problem. In this case $\lambda = 10$ was chosen the same for the BV and NLBV case. The nonlocal result Fig. 28 appears to be slightly better with more details, whereas in Fig. 27 these details are more blocky.

Considering the numerical complexity of the nonlocal method it can be said that the NLBV_τ-regularization does not differ much from the usual BV-regularization if τ is small. One Euler step of the nonlocal method is about τ^2 times a corresponding BV-step. In fact, choosing a large τ does not seem to change the solution very much, for the NLBV-case τ can be chosen smaller than the corresponding parameter ρ for the neighborhood filters.

Summing up, we observed that the nonlocal BV functional shows smoother result than the BV and seems to be slightly better for textured images and small details. However, it generally cannot improve the nonlocal mean filter for the denoising case.

Acknowledgement

The authors would like to thank Jean-Michel Morel and Antoni Buades for providing detailed information on the neighborhood filters and on their implementation and for very useful discussions. Moreover we thank the anonymous referees for important suggestions for improvement of this paper.

References

[1] R. A. Adams, *Sobolev Spaces*, Academic Press, New York San Francisco London, 1975.

[2] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford University Press, New York, 2000.

[3] G. Aubert and J.-F. Aujol, *Modeling very oscillating signals. application to image processing*, Appl. Math. Optim, 51 (2005).

[4] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons, Inc., 1968.

[5] B. Bojarski, *Remarks on Sobolev imbedding inequalities*, in Complex analysis, Joensuu 1987, vol. 1351 of Lecture Notes in Math., Springer, Berlin, 1988, pp. 52–68.

[6] A. Buades, B. Coll, and J. M. Morel, *On image denoising methods*, preprint, CMLA, Cachan, France, 2004.

[7] S. Buckley and P. Koskela, *Sobolev-Poincaré implies John*, Math. Res. Lett., 2 (1995), pp. 577–593.

[8] A. Chambolle, *Finite-differences discretizations of the Mumford-Shah functional*, M2AN Math. Model. Numer. Anal., 33 (1999), pp. 261–288.

[9] A. Chambolle and P.-L. Lions, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188.

[10] H. Engl, A. Hofinger, and S. Kindermann, *Convergence rates in the Prokhorov metric for assessing uncertainty in ill-posed problems*, Inverse Problems, 21 (2005), pp. 399–412.

[11] S. Esedoḡlu and S. J. Osher, *Decomposition of images by the anisotropic Rudin-Osher-Fatemi model*, Comm. Pure Appl. Math., 57 (2004), pp. 1609–1626.

[12] H. Federer, *Geometric measure theory*, Springer Verlag, Berlin-Heidelberg-New York, 1969.

[13] M. Gobbino, *Finite difference approximation of the Mumford-Shah functional*, Comm. Pure Appl. Math., 51 (1998), pp. 197–228.

[14] L. Hörmander, *The Analysis of Linear Partial Differential Operators. I.*, Springer, Berlin, 1990.

[15] P. Huber, *Robust Statistics*, John Wiley & Sons, Inc., New Jersey, 1981.

[16] P. W. Jones, *Quasiconformal mappings and extendability of functions in Sobolev spaces*, Acta Math., 147 (1981), pp. 71–88.

[17] Y. Meyer, *Oscillating patterns in image processing and nonlinear evolution equations*, vol. 22 of University Lecture Series, American Mathematical Society, Providence, RI, 2001.

[18] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterative regularization method for total variation based image restauration*, Multiscale Modelling and Simulation, (to appear 2005).

[19] S. Osher, A. Solé, and L. Vese, *Image decomposition and restoration using total variation minimizaiton and the H^{-1} norm.*, Multiscale Model. Simul., 1 (2003), pp. 349–370.

[20] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 12 (1990).

[21] L. Rudin, P.-L. Lions, and S. Osher, *Multiplicative denoising and deblurring: Theory and algorithms*, in Geometric Level Set Methods in Imaging Vison and Graphics, S. Osher and N. Paragios, eds., Springer, New York, 2003, pp. 103–121.

[22] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.

[23] S. M. Smith and J. M. Brady, *SUSAN-a new approach to low level image processing*, Int. Journal of Computer Vision, 23 (1997), pp. 45–78.

[24] C. Tomasi and R. Manduchi, *Bilateral filtering for gray and color images*, in ICCV, Bombay, 1998, pp. 839–846.

[25] L. Vese and S. Osher, *Modeling textures with total variation minimization and oscillating patterns in image processing.*, J. Sci. Comput., 19 (2003), pp. 553–572.

[26] L. P. Yaroslavsky, *Digital picture processing. An introduction.*, Springer Verlag, 1985.

[27] L. P. Yaroslavsky and M. Eden, *Fundamentals of digital optics*, Birkhäuser, Boston, 1996.

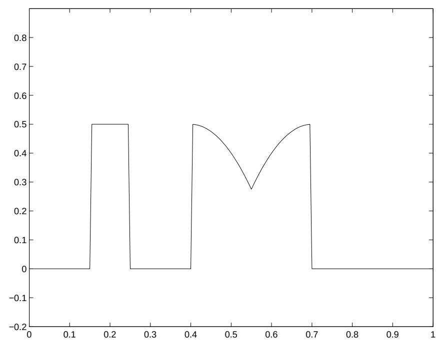

Figure 5: Exact Solution

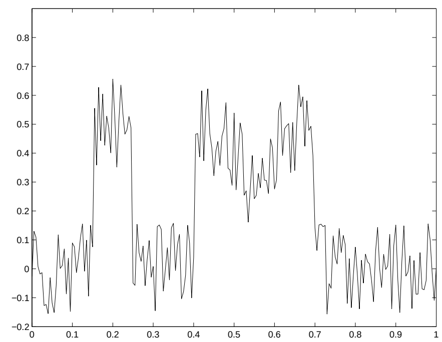

Figure 6: Noisy Data

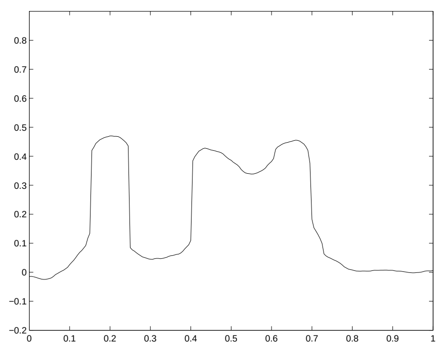

Figure 7: Solution for NLBV

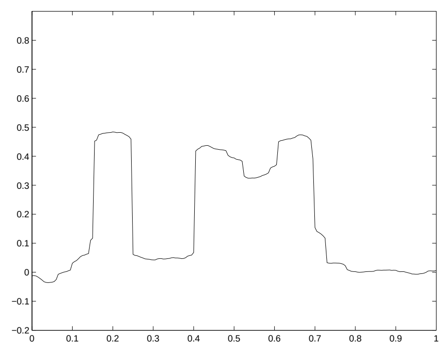

Figure 8: Solution for BV

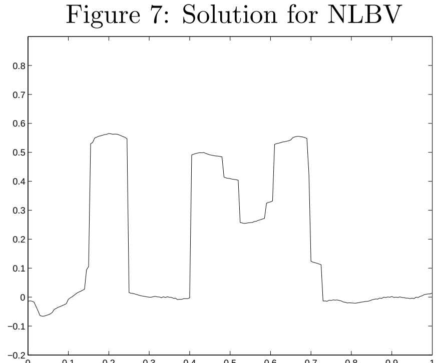

Figure 9: Solution after 3 steps of Bregman iteration for NLBV

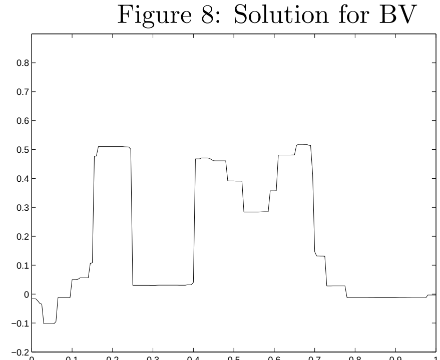

Figure 10: Solution after 3 steps of Bregman iteration for BV

Figure 11: Noisy image

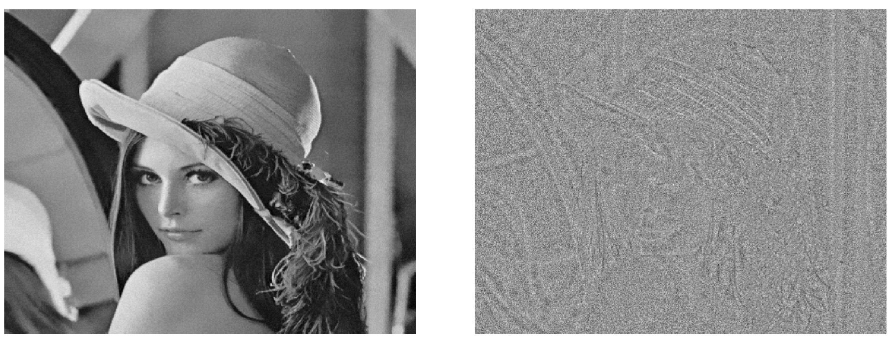

Figure 12: Nonlocal Mean Filter and the residual $f - u$

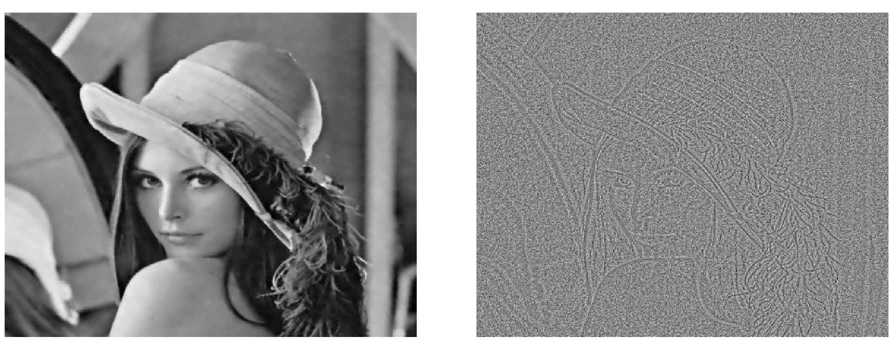

Figure 13: Result for Nonlocal BV and residual $f - u$

Figure 14: Result Bregman iteration u_0,u_1,u_2 and corresponding residual

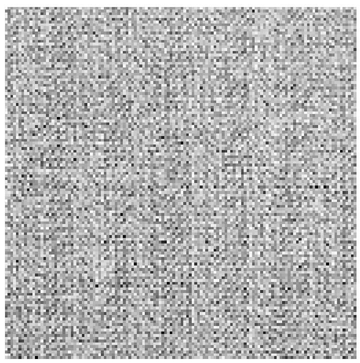

Figure 15: Noisy image

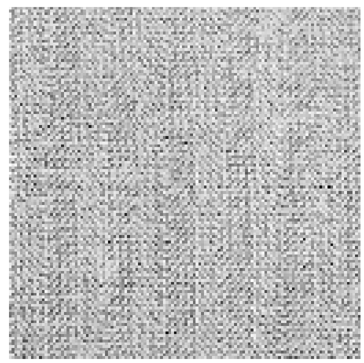

Figure 16: Nonlocal mean Filter

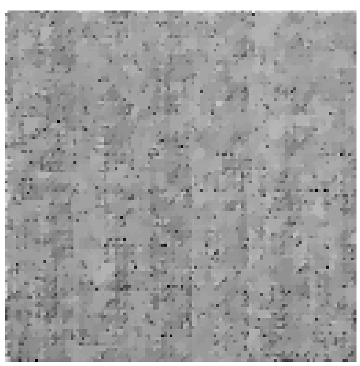

Figure 17: Solution for BV

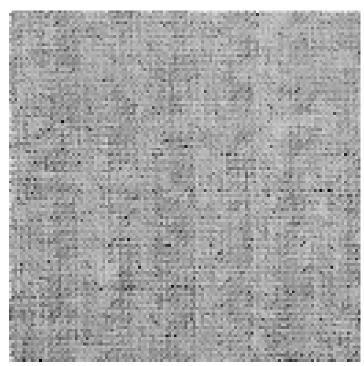

Figure 18: Solution for nonlocal BV

Figure 19: Noisy image

Figure 20: Nonlocal mean Filter

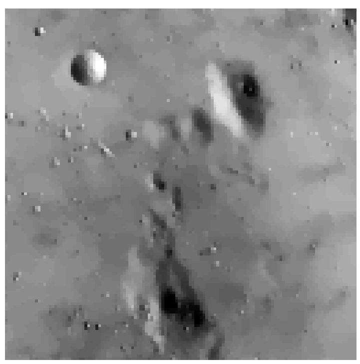

Figure 21: Solution for BV

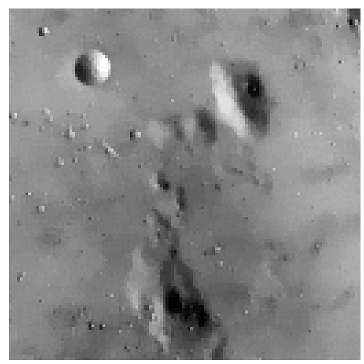

Figure 22: Solution for nonlocal BV

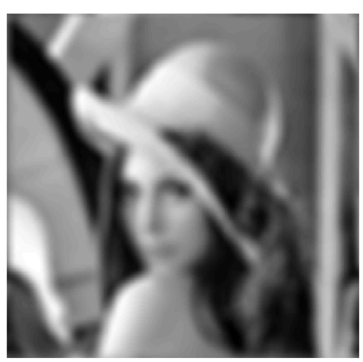

Figure 23: Blurred Image

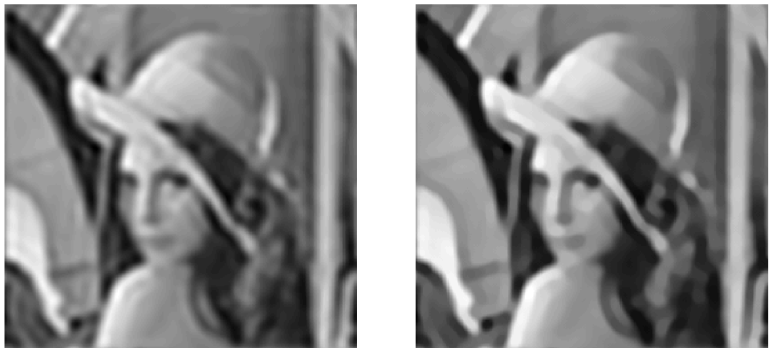

Figure 24: Result deblurring: NLBV and BV

Figure 25: Original image

Figure 26: Blurred image

Figure 27: Deblurring with BV

Figure 28: Deblurring: nonlocal BV