Denoising by BV-duality*

Stefan Kindermann†, Stanley Osher‡, Jinjun Xu §

Dedicated to David Gottlieb on his 60th birthday

Abstract

In this paper we apply Y. Meyer's G-norm for image processing problems. We use a definition of the G-norm as norm of linear functionals on BV, which seems to be more feasible for numerical computation. We establish the equivalence between Meyer's original definition and ours and show that computing the norm can be expressed as an interface problem. This allows us to define an algorithm based on the level set method for its solution. Alternatively we propose a fixed point method based on mean curvature type equations. A computation of the G-norm according to our definition additionally gives functions which can be used for denoising of simple structures in images under a high level of noise. We present some numerical computations of this denoising method which support this claim.

Key Words:
denoising, bounded variation, G-norm, image processing, level sets, ROF

1 Introduction

Many image processing methods are based on a minimization of functionals involving the bounded variation norm. It is widely accepted that the BV-space is a good model for images, since it penalizes noise, while keeping important information such as edges. For instance, denoising of images can be efficiently done by minimizing the Rudin-Osher-Fatemi (ROF) functional [27]:

$$J_{ROF}(u) = \lambda \|u - f\|^2_{L^2(\Omega)} + |u|_{BV}. \tag{1}$$

*Research supported by NIH U54RR021813, NSF DMS-0312222, NSF ACI-0321917 and NSF DMI-0327077

†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

§Department of Mathematics, UCLA, 520 Portola Plaza, Los Angeles, Ca, 90095, email: jjxu@math.ucla.edu

In the following $|u|_{BV}$ denotes the BV-seminorm (see [2, 14])

$$|u|_{BV} := \sup_{\phi \in [C_0^\infty(\Omega)]^2, \|\phi\|_{L^\infty} \leq 1} \int_\Omega u(x) \nabla \cdot \phi \, dx, \tag{2}$$

and $\nabla \cdot$ denotes the divergence operator. Many variants of this functional are known (see e.g. the recent work on Bregman iteration [23]), and the application of the BV-space has been found useful not only for denoising, but for image segmentation and deblurring and regularization of ill-posed problems. In image segmentation a generalization of the ROF-functional leads to the Mumford-Shah minimization [22].

The space of functions of bounded variation roughly contains sketches of an image. The ROF-method decomposes a given image f into $f = u + v$, where u is the minimizer of (1) and v is the residual. u is considered the cartoon-part, while the remaining noise and texture is in the v-part. In image processing, the cartoon u is considered to incorporate the most important information, however, recently Y. Meyer [21] put emphasis on the v-part, since it contains information about the texture part of the image. In [21] the G-norm was introduced, as a norm that measures textural parts. The G-norm is basically the dual norm of the BV-seminorm. It turns out that it is also extremely useful in describing the structure of minimizers of (1) and related functionals. In particular, the optimality conditions can be expressed in a concise form involving this norm. Besides this, several variational image decomposition methods have been proposed using the G-norm as the norm for the error $u - f$ instead of the usual L^2-norm – this is also known as Meyer-decomposition (see for instance [3, 4, 19, 21, 32]).

However, one of the problems in dealing with the original definition of the G-norm as in (3) is that it is difficult to compute. Instead in this paper we use its definition as a norm of functionals on BV (6), from which the G-norm of a function can be calculated by solving an optimization problem, which can be handled by standard routines. Moreover, the optimization problem can be rephrased as a problem only involving interfaces, see Subsection 2.1. This allows us to apply the well known machinery of the level set method to this problem.

One of the main points in this paper is that this optimization not only gives the G-norm, but also the function where the extremum in (6) is attained. This extremal function contains much information about the image itself, and can be used to extract features hidden in high noise, which can be used for denoising of extremely degraded images. In Section 5 we present some of the astonishing results.

2 G-norm

In the following we assume that either $\Omega = \mathbb{R}^n$, $n \leq 2$ or $\Omega \subset \mathbb{R}^n$, $n \leq 2$, is a simply connected Sobolev extension domain. We denote by BV the Banach space

$$BV := \{u \in L^1(\Omega) \,|\, \|u\|_{BV} := \|u\|_{L^1} + |u|_{BV} < \infty\},$$

with the BV-seminorm (2).

In the following we will deal with the G-norm, which is roughly the dual norm of BV. It was introduced by Y. Meyer [21] for the case of $\Omega = \mathbb{R}^n$: Let $f \in L^2(\mathbb{R}^n)$, then the G-norm $\|.\|_*$ is defined as

$$\|f\|_* := \inf\{ \|g\|_{L^\infty(\mathbb{R}^n)} \mid \nabla\cdot g = f\}, \tag{3}$$

where the identity $\nabla\cdot g = f$ has to be understood in the sense of distributions.

This definition was generalized to Ω being a bounded domain in [3]. In this case nonzero constant functions have unbounded G-norm. Hence it is necessary to include a normalization condition, namely f has to have zero mean. Let $f \in L^2(\Omega)$ with

$$\int_\Omega f(x)dx = 0, \tag{4}$$

let $\vec{n}$ be the unit normal vector on the boundary, then the generalization of (3) to bounded domain is

$$\|f\|_* := \inf\{\|g\|_{L^\infty} | \nabla\cdot g = f, g.\vec{n} = 0 \text{ on } \partial\Omega\}. \tag{5}$$

For our purposes it is more convenient to use a definition of $\|.\|_*$ as a dual norm.

In the following we have to distinguish between the two cases $\Omega = \mathbb{R}^n$ and Ω being bounded. At first we show that definitions (3) and (5) can be rephrased in the following form:

Lemma 2.1. *Let $f \in L^2(\Omega)$, $n \leq 2$. If Ω is a bounded extension domain, let f additionally satisfy* (4)*. Then both* (3) *and* (5) *are equal to*

$$\|f\|_* := \sup_{u\in W^{1,1}(\Omega), u\neq 0} \frac{\int_\Omega f(x)u(x)dx}{\int_\Omega |\nabla u|dx} = \sup_{u\in BV, u\neq 0} \frac{\int_\Omega f(x)u(x)dx}{|u|_{BV}} \tag{6}$$

Proof. A proof of the first equality can be found in [1]. For the second equality note that any BV-function u can be approximated by smooth $W^{1,1}$-function u_n such that $\|u - u_n\|_{L^1} \to 0$ and $|u|_{BV} = \lim_{n\to\infty} \int_\Omega |\nabla u_n(x)|dx$ [2, Thm. 3.9]. Moreover by the continuous embedding of $BV \to L^2$ for such an approximation the sequence $\|\frac{u_n}{|u_n|_{BV}}\|_{L^2}$ is bounded and thus has a weakly convergent subsequence, whose limit must be $\frac{u}{|u|_{BV}}$. Hence, the suprema are equal. □

We are mainly using definition (6) for our purposes. The next lemma shows that the supremum in (6) is attained at a BV-function:

Lemma 2.2. *Under the same assumptions on Ω and f as in Lemma 2.1, the supremum in* (6) *is attained at a BV-function:*

$$\|f\|_* = \max_{u\in BV(\Omega), u\neq 0} \frac{\int_\Omega f(x)u(x)dx}{|u|_{BV}}. \tag{7}$$

Proof. Let $u_n \in BV$ be a maximizing sequence, then $v_n := \frac{u_n}{|u_n|_{BV}}$ is in BV, with $|v_n|_{BV} = 1$. If Ω is bounded, property (4) allows us to add constants to v_n without changing the value of the ratio in (7). Hence we can assume that each v_n has zero mean. By the continuous embedding into $BV \to L^2$, a sequence of function with zero mean and bounded BV-seminorm has a weakly convergent subsequence in L^2. Moreover since the embedding from BV to L^1, is compact, the limit v of this subsequence is in BV, and by definition a maximizer. If $\Omega = \mathbb{R}^n$, then by the continuous embedding v_n is bounded in $L^2(\mathbb{R}^n)$ and hence has a weakly convergent subsequence. Its limit v is in $L^2(\mathbb{R}^n)$, and by weak convergence satisfies $|v|_{BV} \leq 1$. By continuous embedding [2, Thm 3.47] $v \in L^1$, hence $v \in BV$ and by weak convergence it is a maximizer. □

Note that for bounded domains condition (4) is essential, because if it does not hold, the supremum in (6) is infinity, since we can add arbitrary constants to u.

By definition the following fundamental inequality holds (cf. [21]): For any $f(x), u(x) \in L^2(\Omega)$

$$\int_\Omega f(x)u(x)dx \leq \|f\|_* |u|_{BV} \tag{8}$$

If $\|f\|_* |u|_{BV} < \infty$ then equality holds if and only if u is a maximizer of (6). Moreover it was shown in [15, 21] by the isoperimetric inequality that any L^2-function (which has zero mean in the case of a bounded domain) has bounded G-norm:

$$\|f\|_* \leq C\|f\|_{L^2}.$$

For $\Omega = \mathbb{R}^2$ the constant C can be taken as $C = \frac{1}{2\sqrt{\pi}}$ (see [15]), for bounded domains it depends on the embedding constants of $BV(\Omega) \to L^2(\Omega)$.

2.1 Level set formula

Using the coarea formula [2] G. Strang observed in [30] that the supremum in (6) can be taken over characteristic functions in BV. This has been used in [5] and [12] to construct minimizers for functionals defined for characteristic functions by extending the problem to BV-functions and using appropriate thresholding. For our purposes we go the opposite way: The optimization problem (6) can be reduced to a problem only involving characteristic functions. In this way we are reduced to an interface problem, which can be handled by the level set method.

For the optimization problem (6) the result of Strang can be formulated as follows.

Lemma 2.3. *Let $u \in BV(\Omega)$, $u \neq 0$ be such that*

$$\frac{\int_\Omega f(x)u(x)dx}{|u|_{BV}} = \lambda,$$

Then for any ϵ there exists a μ such that the upper level set

$$\chi_\mu(x) := 1_{\{x|u(x)\geq\mu\}}$$

satisfies

$$\frac{\int_\Omega f(x)\chi_\mu(x)dx}{\int_\Omega |\nabla\chi_\mu(x)|dx} \geq \lambda - \epsilon,$$

Proof. This has been shown in [30] as follows: u satisfies

$$K_{f,\lambda}(u) := \int_\Omega f(x)u(x)dx - \lambda \int_\Omega |\nabla u(x)|dx = 0. \tag{9}$$

By the coarea formula we can rewrite this as

$$\int_{-\infty}^{\infty} K_{f,\lambda}(\chi_\mu)d\mu = 0.$$

Now suppose that the conclusion in the theorem does not hold, i.e., an ϵ exists such that

$$\frac{\int_\Omega f(x)\chi_\mu(x)dx}{\int_\Omega |\nabla\chi_\mu(x)|dx} < \lambda - \epsilon,$$

for all μ such that $\chi_\mu \neq 0$. Then we have

$$K_{f,\lambda}(\chi_\mu) < -\epsilon \int_\Omega |\nabla\chi_\mu(x)|dx.$$

Integration with respect to μ gives

$$K_{f,\lambda}(u) < -\epsilon \int_\Omega |\nabla u(x)|dx,$$

contradicting (9). □

The following is a simple corollary of Lemma 2.3:

Corollary 2.4. *The supremum in* (6) *can be restricted to characteristic functions with finite perimeter. Moreover if $u \in BV$ is a maximizer, i.e., it satisfies*

$$\|f\|_* = \frac{\int_\Omega f(x)u(x)dx}{|u|_{BV}},$$

then for almost all μ with nonempty level sets $\chi_\mu(x) := 1_{\{x|u(x)\geq\mu\}}$, these level sets satisfy

$$\|f\|_* = \frac{\int_\Omega f(x)\chi_\mu(x)dx}{|\chi_\mu|_{BV}}, \tag{10}$$

In particular, the maximum in (7) *is attained at a characteristic function.*

Proof. Let u be a maximizer of (7) and $\chi_\mu = 1_{\{x|u(x)\geq\mu\}}$ be the characteristic function of its μ-level set. From (8) and (9) with $\lambda = \|f\|_*$ we get $K_{f,\lambda}(\chi_\mu) \leq 0$ and $\int K_{f,\lambda}(\chi_\mu)d\mu = 0$, thus $K_{f,\lambda}(\chi_\mu) = 0$ for almost every μ. If $\chi_\mu(x)$ is nonempty it must satisfy (10). □

Let us give some examples for the G-norm and extremal characteristic functions. First we consider the one-dimensional case: Let $\Omega = I = [a, b]$ be an interval and $f \in L^2(I)$ with $\int_I f(x)dx = 0$. From (5) we find that

$$\|f\|_* = \inf\{ \|F\|_{L^\infty} | F(x) = \int_a^x f(s)ds\}$$

Now $F(x)$ is a continuous function, which attains its maximum and minimum on I. If $c \in (a, b)$ is such that $|F(c)| = \max_x |F(x)|$, then

$$\|f\|_* = |F(c)| = \int_I f(x)\chi(x)dx = \frac{\int_I f(x)\chi(x)dx}{|\chi|_{BV}},$$

where $\chi(x)$ is either the characteristic function of the interval $[a, c]$ or of $[c, b]$ depending on the sign of $F(c)$. Hence in the one-dimensional case χ_μ in (10) can be taken as a characteristic function of an interval with one endpoint at a or b.

Another interesting example is the case $\Omega = \mathbb{R}^2$ and f is a characteristic function itself. Let $f(x) = \chi_C(x)$ be the characteristic function of the set C, then the optimization problem (6) is

$$\|f\|_* = \sup_{D,\, \mathrm{per}(D)<\infty} \frac{|C \cap D|}{|\partial D|}.$$

This can be seen as a generalized isoperimetric problem. An extremal D has maximal area relative to C with minimal perimeter. Recalling the fact that a circle has maximal area with given perimeter, it is not surprising that if $C = B_R$ is a circle with radius R, D can be taken as the same circle, yielding

$$\|\chi_{B_R}\|_* = \frac{R}{2}.$$

This has already been shown in [21]. Note that if we consider the G-norm of this circle on a bounded domain the corresponding value of the norm is smaller, because f has to be normalized with (4). In fact if B_R is contained in Ω, where $\Omega = B_{R_2}$ is another circle and $f(x) = \chi_{B_R}(x) + c$, with $c = \frac{-R^2\pi}{|\Omega|} = -\frac{R^2}{R_2^2}$, then the choice $D = B_R$ gives

$$\frac{\int_\Omega f(x)\chi_\mu(x)dx}{|\chi_\mu|_{BV}} = \frac{R}{2}(1 - \frac{R^2\pi}{|\Omega|}) < \frac{R}{2}. \tag{11}$$

Taking $D = B_R$ gives only a lower bound on $\|f\|_*$. That this is indeed the optimal choice can be proven by definition (5), where $g_i(x) = x_i w(|x|)$ with $w(r) = \frac{1}{2}(1 - \frac{R^2}{R_2^2})$ for $0 \le r \le R$ and $w(r) = \frac{1}{2}(\frac{R^2}{r^2} - \frac{R^2}{R_2^2})$ for $R \le r \le R_2$ (compare [21]). Now $\|g\|_{L^\infty}$ gives an upper bound on $\|f\|_*$, which coincides with the lower bound, hence (11) holds.

Note however that the maximizing set D is not always identical with C. If C is a square, then according to [30] the optimal set D is a rounded square.

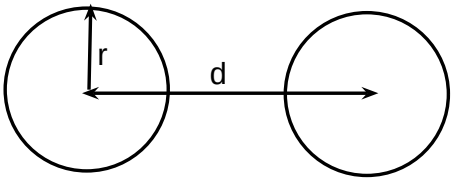

Figure 1: double-circle: two identical circles with radius = R and distance between centers = d.

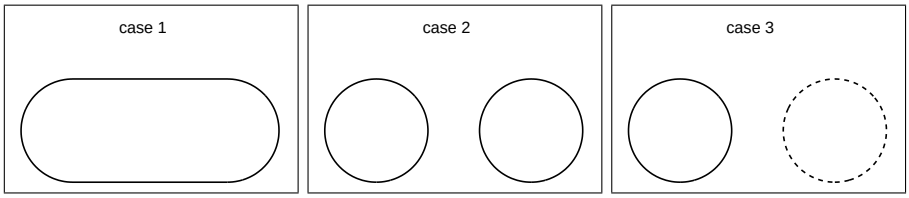

Figure 2: three possible solutions to the double-circle example. case 1 (left), case 2 (middle), case 3 (right)

Another example is the case of two circles. Let C be the union of two circles with radius r, whose centers are separated by a distance $d > 2r$, as shown in Figure 1. If d is small enough then D is not identical with C. In fact taking D as the convex hull of C, as in Figure 2 case 1 (left) gives a different value than taking D as one of the circles (Figure 2 case 3) or taking $D = C$ (case 2):

- case 1: $\frac{|C \cap D|}{|\partial D|} = \dfrac{2\pi r^2}{2\pi r + 2d}$;
- case 2, 3: $\frac{|C \cap D|}{|\partial D|} = \dfrac{r}{2}$.

We assume that the optimal D is amongst these three cases. If this holds, then the G-norm depends on the distance: If the centers of the circle are less than πr apart, then case 1 is a better choice, whereas for $d \geq \pi r$ case 2 or case 3 are. In this sense, the G-norm treats the two circles when close enough as one object, while they are considered separated objects for $d \geq \pi r$.

The last example also shows that the solution to the optimization problem (6) is not unique, since both case 2 and case 3 give the same value.

2.2 Subgradient

In this section we relate the solution to the optimization problem (6) to the set of subgradients of the functional mapping $f \to \|f\|_*$ for $f \in L^2$. Before we prove this we have to consider the set of plain images, which was defined in [15]. We use a slightly generalized definition:

Definition 2.5. *Let $g \in BV(\Omega)$. We say that g is a plain image if there exists an $f \in L^2$ such that*

$$\int_\Omega f(x)g(x)dx = |g|_{BV}, \qquad \text{with } \|f\|_* = 1. \tag{12}$$

We denote by PI the set of plain images: $PI := \{\, g \in BV \,|\, g \text{ is a plain image}\}$.

This is just another way of saying that g is a maximizer in (6) for some function f. In [15] plain images are defined as those for which an $f \in BV$ exists such that (12) holds. There it is proven that these plain images equal the set of cartoon-parts u in an ROF-decomposition [27] $f = u+v$ with $f \in BV$. Following this proof it is straightforward to see that g is a plain image in the sense of Definition 2.5 if and only if g is the cartoon part u of a ROF-decomposition of some function $f \in L^2$.

Now we can relate the optimization problem (7) to the subgradient:

Theorem 2.6. *Let $F : L^2(\Omega) \to \mathbb{R}$ be the convex functional*

$$F : f \to \|f\|_*$$

Let $f \neq 0$ and denote by $\partial F(f)$ the subgradient of F at f. Then the set of solutions to the optimization problem

$$u = \text{argmax}_{|v|_{BV}=1} \int_\Omega f(x)v(x)dx. \tag{13}$$

equals the set of plain images in the subgradient:

$$(13) \Leftrightarrow u \in \partial F(f) \cap PI.$$

Proof. According to the definition, a function p is in the subgradient set [8, 33] $\partial F(f)$ if and only if

$$\|g\|_* \geq \|f\|_* + < p, g - f > \quad \forall g \in L^2$$

Since $g, f \in L^2$ we can identify a linear functional $< p, . >$ with some function $p \in L^2$. Choosing $g = \lambda f$ for $\lambda > 0$ gives

$$(\lambda - 1)\left(\|f\|_* - \int_\Omega p(x)f(x)dx\right) \geq 0 \quad \forall \lambda \geq 0$$

hence a necessary condition for p to be in the subgradient set is

$$\int_\Omega p(x)f(x)dx = \|f\|_*. \tag{14}$$

Moreover, $p \in \partial F(f)$ if and only if (14) hold and

$$\int_\Omega p(x)g(x)dx \leq \|g\|_* \quad \forall g \in L^2 \tag{15}$$

If u satisfies (13), then (14) holds for $p = u$ and (15) holds by (8). If we take $\frac{f}{\|f\|_*}$ in (14) if follows that u is a plain image. Conversely, let $p \in \partial F(f)$ be a plain image. We can identify p with an L^2-function, which satisfies (14) and (15). From (15) and (12) it follows that $|p|_{BV} \leq 1$. From (14) and (8) we get $|p|_{BV} \geq 1$, thus $|p|_{BV} = 1$. Since $\|f\|_*$ is the maximum value in (13) and since (14) holds, p must be a maximizer in (13). □

Remark 2.7. The optimization problem (6) can also be interpreted as a segmentation problem. Let us assume that f has zero mean (4). By Strang's formula the optimization problem is equivalent to an optimization over characteristic functions χ_D, where $D = \{x|\chi_D(x) > 0\}$

$$\|f\|_* = \max_{\chi_D \in BV} \frac{\int_\Omega f(x)\chi_D dx}{\int |\nabla \chi_D| dx} = \max_D \frac{\int_D f(x)dx}{|\partial D|} \tag{16}$$

$$= \frac{1}{2} \frac{|\int_D f(x)dx - \int_{\Omega \setminus D} f(x)dx|}{|\partial D|}.$$

We can think of the extremal set D in (6) as that one which separates the domain Ω into two sets $D, \Omega \setminus D$, such that the difference between the integrals of f over these two sets is large, while the perimeter of the interface ∂D is as small as possible. From the basic inequality (8) $\int f(x)\chi_D(x)dx \geq -\|f\|_*|\partial D|$ we can also deduce that the optimal D in (16) is a solution to the segmentation problem, which has a more familiar form:

$$\min_D \frac{1}{2} \left(\int_D f(x)dx - \int_{\Omega \setminus D} f(x)dx \right) + \lambda |\partial D|,$$

where λ is chosen as $\lambda = \|f\|_*$.

Remark 2.8. Let us remark that the G-norm is closely related to prescribed mean curvature equations of the form:

$$\nabla \cdot \frac{\nabla u}{\sqrt{|\nabla u|^2 + 1}} = f(x) \tag{17}$$

Such an equation comes up as Euler-Lagrange equation for an area functional of the graph of u:

$$J(u) = \int_\Omega \sqrt{1 + |\nabla u(x)|^2} dx + \int_\Omega f(x)u(x). \tag{18}$$

Equation (17) does not always have a solution, in fact a necessary condition for solvability of (17) is $\|f\|_* < 1$. This can be seen from the inf - definition of the G-norm (5). Considering the Dirichlet problem for equation (17), sufficient conditions for solvability can be derived. Besides smoothness, these involve on the one hand a relation of the curvature of the boundary of Ω and the boundary values of f. Interestingly, a second ingredient is the condition (in our

notation) $\|f\|_* < 1$. In this case the prescribed mean curvature equation allows a differentiable solution. For the precise theorem see [13, Thm. 16.10].

It is unknown to the authors if a similar result holds for the corresponding Neumann-problem with $\frac{\partial}{\partial n} f(x) = 0$ on the boundary. It would be interesting if in this case the condition on the boundary of $\partial\Omega$ is needed, or if $\|f\|_* < 1$ together with sufficiently smoothness on f suffices for a differentiable solution.

3 Duality Denoising

The G-norm is a central ingredient in characterizing minimizers of functionals with BV-regularization. Let us give some examples. At first we state the well-known optimality conditions of the ROF-functional (1). This functional has a unique minimizer in BV. The ROF-image decomposition splits a given image f into $f = u + v$, where u a minimizer of (1) is interpreted as the cartoon-part of the image, whereas the residual v contains noise and texture. The optimality conditions for $J_{ROF}(u)$ are derived in [21]:

Theorem 3.1. *If $\|f\|_* \leq \frac{1}{2\lambda}$, then the minimizer of $J_{ROF}(u)$ is $u = 0$.*

If $\|f\|_* > \frac{1}{2\lambda}$, then the minimizer of $J_{ROF}(u)$ u is nonzero and the ROF-decomposition $f = u + v$ is characterized by the conditions

$$\|v\|_* = \frac{1}{2\lambda} \tag{19}$$

$$\int_\Omega u(x)v(x)dx = \frac{1}{2\lambda}|u|_{BV} \tag{20}$$

Proof. See [21] □

Note that since $u \neq 0$, by definition of $\|v\|_*$ (20) is equivalent to

$$u = \text{argmax}_{\tilde{u}\in BV, \tilde{u}\neq 0} \frac{\int_\Omega v(x)\tilde{u}(x)dx}{|\tilde{u}|_{BV}}. \tag{21}$$

This means, that u can be recovered from the noise part v, which is supposed to contain mostly noise. This observation is a main point for our denoising algorithm.

The optimality condition has been extended to other functionals as well. An alternative to the ROF-method is the $BV - L^1$ decomposition, where $f = u + v$, and u is a minimizer of

$$J_{L^1}(u) = \lambda\|u - f\|_{L^1} + |u|_{BV}. \tag{22}$$

Note that in this case due to lack of strict convexity uniqueness of a minimizer does not hold. The optimality conditions for this functional only involve the signum-function sign(v): (see [25]).

Theorem 3.2. *Let $u \neq 0$ be a minimizer of (22), and suppose $\{x|u(x) = f(x)\}$ has measure zero and $\|\text{sign}(u - f)\|_* > \lambda$, then u satisfies*

$$u = \text{argmax}_{\tilde{u} \in BV} \frac{\int_\Omega \text{sign}(v(x))\tilde{u}(x)dx}{|\tilde{u}|_{BV}}. \tag{23}$$

If $\{ x \mid u(x) = f(x)\}$ has positive measure the optimality condition has to be relaxed.

Instead of changing the norm in the fidelity term $u - f$, a different approach was taken in [11], using an anisotropic BV-seminorm. For this model the BV-norm $|u|_{BV} = \int_\Omega |\nabla u| dx$ is replaced by:

$$|u|_{BV_\phi} := \int_\Omega \phi(\nabla u)dx,$$

with a positively 1-convex function ϕ. Similar optimality conditions for the ROF-model have been deduced, where the G-norm $\|.\|_*$ is replaced by

$$\|f\|_{\phi,*} := \sup_{u \in BV_\phi} \frac{\int_\Omega f(x)u(x)dx}{|u|_{BV_\phi}}$$

The optimality conditions for this model remain the same as for the ROF-model, if in Theorem 3.1 the BV-norm $|.|_{BV}$ and the G-norm $\|.\|_*$ are replaced by their anisotropic variants, $|.|_{BV_\phi}$ and $\|f\|_{\phi,*}$, respectively. In particular a nonzero solution satisfies

$$u = \text{argmax}_{\tilde{u} \in BV_\phi} \frac{\int_\Omega v(x)\tilde{u}(x)dx}{|\tilde{u}|_{BV_\phi}}. \tag{24}$$

3.1 Denoising

The optimality conditions in Theorem 3.1, (23), (24) have one common property: The u-component of a decomposition can be recovered from the v-component without knowledge of f or λ. This is somehow surprising, since the v-component is supposed to contain only noise and texture and almost no signal. In Section 5 we give an example for the recovery of a signal from a v-component, which visually seems to contain almost nothing but noise. Of course, a maximizer in (21) is not unique since the level sets of a maximal u are maximizers as well (cf. Corollary 2.4). But even though we might not be able to find all maximizers we can at least expect to recover some level set of the original image u. The numerical computations in Figure 12 also show that the computed level set corresponds to main edges in the picture.

Note, however, that in theory this works only if we a-priori know that v comes from an ROF- or equivalent decomposition. But since the optimization problem (6) can be performed for any f, we can use this for denoising of any image, simply by computing the maximizer u in (7).

The usefulness of the algorithm is, however, restricted by the condition, that the result u should contain objects which are related to the image. Moreover a

maximizer of (6) is not unique, since u can be multiplied by any nonzero scalar, but again Corollary 2.4 indicates that we might be able to recover the main edges of the image.

In this sense our method is not competing with traditional methods for image denoising, where only little noise and much structure is present. If (6) is applied to a standard image with very little noise, the result of the optimization problem is a characteristic function which contains only basic information on the image. This function cannot always be related to edges of features of the image.

Instead, the results for the duality denoising are best for images which contain almost no structure but only noise, and where the information behind the noise is a very simple image (e.g a characteristic function). Roughly speaking we think our method should be applied to a noise model of the form

$$f = \text{simple image} + \text{large noise},$$

In Section 5 we give some numerical example, where the algorithm works well. The simple images in these cases are characteristic functions, which are hidden in noise, which can be larger then the information itself.

4 Numerical Computation

In this section we present some numerical algorithms to compute approximate maximizers of the optimization problem (6), by solving the Euler-Lagrange equations. First of all consider the functional for which an extremum is to be found (we use $\int_\Omega |\nabla u| dx$ as a notation for $|u|_{BV}$):

$$E(u) := \frac{\int_\Omega f(x)u(x)dx}{\int_\Omega |\nabla u(x)|dx} \tag{25}$$

A formal differentiation with respect to u gives the Gateaux-derivative

$$E'(u) = \frac{1}{|u|_{BV}} \left(f(x) + \frac{\int_\Omega f(x)u(x)}{\int_\Omega |\nabla u(x)|dx} \nabla . \frac{\nabla u(x)}{|\nabla u(x)|} \right).$$

Here $\nabla . \frac{\nabla u(x)}{|\nabla u(x)|}$ is the usual useful but sloppy notation for the subgradient of the BV-seminorm. We want to find the maximum of $E(u)$, hence we may derive an evolution equation, following the steepest ascent (u_t denoting the time-derivative):

$$u_t = \frac{1}{|u|_{BV}} \left(f(x) + E(u) \nabla . \frac{\nabla u(x)}{|\nabla u(x)|} \right)$$

The factor $\frac{1}{|u|_{BV}}$ is a constant in the space variable x and might be set to 1 by an appropriate scaling.

4.1 Level Sets

The preceding algorithm is one straightforward possibility. It does not take into account any knowledge of the structure of the solution. However, by Strang's formula (Corollary 2.4) we know that a solution to the optimization problem (6) is a characteristic function. Using this we can apply the well-known formalism of the level set method [26] to calculate the interface of this characteristic function. As usual in the level set method, we define the interface of the characteristic function as the zero-level set of an additional function ϕ. For a given subdomain $D \subset \Omega$, we define the level set function $\phi(x)$ on Ω as

$$\phi(x) = \begin{cases} < 0 & x \in D, \\ = 0 & x \in \partial D, \\ > 0 & x \in D^c. \end{cases} \tag{26}$$

Using the one-dimensional Heaviside function H and delta function δ, which are defined as

$$H(\phi) = \begin{cases} 0 & \text{if } \phi \leq 0 \\ 1 & \text{if } \phi > 0 \end{cases}, \qquad \delta(\phi) = H'(\phi), \tag{27}$$

the volume integral of function f over D can be written as the integral over the whole domain Ω

$$\int_D f(x)\,dx = \int_\Omega f(x)(1 - H(\phi(x)))\,dx, \tag{28}$$

and the perimeter can be obtained as

$$\text{per}(\partial D) = |\partial D| = \int_\Omega \delta(\phi(x))|\nabla\phi(x)|\,dx. \tag{29}$$

Thus, the optimization problem with energy (25) can be written as an optimization problem over level set functions:

$$\|f\|_* = \sup_\phi \tilde{E}(\phi)$$

$$\tilde{E}(\phi) := \frac{\int_\Omega f(x)(1 - H(\phi))\,dxdy}{\int_\Omega \delta(\phi)|\nabla\phi|\,dxdy}. \tag{30}$$

The gradient of $\tilde{E}(\phi)$ can be found from E by the identity $\tilde{E}(\phi) = E(H(-\phi))$, thus:

$$\tilde{E}'(\phi) = -E'(H(-\phi))|\nabla\phi|$$

By rescaling we get the following evolution equation for the level set function ϕ

$$\phi_t = |\nabla\phi|\left[\tilde{E}(\phi(t))\left(\nabla \cdot \frac{\nabla\phi}{|\nabla\phi|}\right) - f\right]. \tag{31}$$

The steady state solution of (31) will be at least a local extremum of (30).

For the level set method we do not assume that the zero level set of ϕ is inside Ω. In fact, from (10) it follows that we have to look for the supremum over all characteristic functions and not only those which are supported in Ω. It is important to note that by definition of the BV-norm the perimeter of a characteristic set is the length of $\partial D \cap \Omega$, so the boundary of Ω does not contribute to the perimeter. Therefore one should not use Dirichlet boundary condition on ϕ, but homogeneous Neumann boundary conditions.

Another point in the level set algorithm is that the normalization condition (4) is not really needed. The question, whether to normalize f or not depends on whether we want to compute the G-norm on a bounded (5) or on the whole domain (3). If f is supported in Ω, then both cases work fine. However, if this is not the case and f is not normalized, then the level sets of ϕ can disappear over the boundary of $\partial\Omega$, because nothing prevents ϕ from having its zero contour outside Ω. This does not happen if f is normalized, and for images this is the correct way to proceed.

For the level set algorithm it is crucial to have a good initial guess, because the algorithm may get stuck in a local extremum. For example if f is the characteristic function of two separated circles, a local extremum is the case where u is taken as the characteristic function of one of these circle. If the initial guess is not too far away from this extremum it will converge towards it, although the global minimum might be different.

4.2 Initial guess

We suggest a heuristic method to find a good initial guess. Again this is based on Strang's formula. Suppose we have a reasonably good candidate for a maximum of $E(u)$. Then Lemma 2.3 tells us, that by looking at the upper level sets of u, $\chi_\mu(x) := H(u-\mu)$, we can find a function, which has a value $E(\chi_\mu)$ as good as $E(u)$ or even better. Hence, it seems reasonable to choose as an initial guess for the level set algorithm the interface of χ_μ, where χ_μ is that upper level set, which has maximum value $E(\chi_\mu)$. This yields the following level set scan algorithm:

Level set scan:

1. Given u, take a finite number of values μ_i, $i = 1 \dots k$, $\mu_i \in [\min(u), \max(u)]$
2. **For** $i = 1$ to k
   Set $\chi_{\mu_i} = H(u - \mu_i)$ (upper level set at $u = \mu_i$)
   Compute $E(\chi_{\mu_i})$
   end
3. Set $\chi_\mu := \text{argmax}_{i=1\dots k}(|E(\chi_{\mu_i})|)$

Note that we compute the perimeter numerically, so in order to obtain reliable results, we reject characteristic functions, which are too small or too large. We only maximize over those functions χ_{μ_i} whose area is within $95\% - 5\%$ of the total area of the computational domain Ω and whose perimeter is larger than a

small threshold. In this way we avoid level-sets which contain only few pixels. The reason for this is that a numerical error in the perimeter leads to a large relative numerical error in E (and thus unreliable result) if the perimeter is small.

For our numerical computations we observe that the algorithm is rather robust in the selection of the number of levels μ_i . We used 100 values to test for.

What remains to discuss is how to find – in a simple way – an input function u for the level set scan. We use the solution to the Poisson equation

$$\Delta u = f, \tag{32}$$

with zero Neumann boundary conditions. (Note that we assume that f has zero mean, so a solution exists). The reason for this choice comes from the analogy to the corresponding optimization problem for the H^1- norm. In fact, the solution to (32) is proportional to an extremum of the functional $u \to \frac{\int_\Omega f(x)u(x)dx}{(\int_\Omega |\nabla u|^2 dx)^{1/2}}$. Summing up, the procedure is of the following kind:

1. Solve the Poisson equation (32) for u
2. Perform a level set scan on the solution u
3. The output of the level set scan is an initial guess to the evolution equation (31).

4.3 Fixed point algorithm

The level set approach gives a quite precise description of the interface we are looking for. It performs well, if f is a simple function itself, such as a characteristic function. However, the problem of local minima is inherent to this algorithm. We therefore suggest a second method, which does not rely on the level set evolution, and which seems to work better for complicated functions f, such as images. The main idea is to solve the equations for the steady state of the optimization problem: Replacing the BV-norm by its well-known ϵ-regularization the steady state solution satisfies

$$\begin{aligned} \nabla . \frac{\nabla u(x)}{\sqrt{\epsilon^2 + |\nabla u(x)|^2}} &= \frac{1}{\lambda} f(x) \\ \lambda &= \frac{\int_\Omega f(x)u(x)dx}{\int |\nabla u(x)| dx} \end{aligned}$$

Solving this equation iteratively gives the following sequence u_n, λ_n, defined by

$$\lambda_{n-1} = \frac{\int_\Omega f(x)u_{n-1}(x)dx}{\int |\nabla u_{n-1}(x)| dx} \tag{33}$$

$$\nabla . \frac{\nabla u_n(x)}{\sqrt{\epsilon^2 + |\nabla u_{n-1}(x)|^2}} = \frac{1}{\lambda_{n-1}} f(x) \tag{34}$$

As initial guess we use u_0 the solution to the Poisson equation (32). After a number of iterations, we can perform a level set scan on u_n to find a final solution χ_μ from the upper-level sets of u_n.

Of course, this algorithm has many variants. For instance λ_n can be kept fixed for a couple of u-iterations, which amounts to solving the nonlinear elliptic equation in (33) by a iterative fixed-point method. Note that in our case the choice of ϵ in (34) does not play a role, since by an appropriate scaling $u \to \epsilon\tilde{u}$, ϵ can be set to 1. The values of $u(x)$ are not important, what is important are its level sets. So the scaling does not affect the outcome of the algorithm.

There is a crude variant of this algorithm, where λ_n is not computed by the ratio (34), but is chosen as geometrically increasing sequence. This has the effect that if λ_n is larger than $\|f\|_*$ a solution u_n to Equation (33) does not exist, but will diverge to infinity. However, although the values of u_n diverge, its level sets seem to be stable over a long period of iteration. So performing such an iteration together with a stopping criterium and a level set scan still gives reasonable results.

5 Numerical Implementation and Results

In this section we discuss the numerical implementation of the algorithms and present some result. Let us focus on the level set approach: the evolution Equation (31) can be rewritten as

$$\phi_t + f|\nabla\phi| = \tilde{E}(\phi(t))\kappa|\nabla\phi|, \tag{35}$$

where κ denotes the curvature of the zero level set of ϕ. In the level set framework (cf. [24]), the left hand side of (35) is a motion in normal direction with time-independent speed $f(x, y)$, which is hyperbolic, and the right hand side is a curvature-dependent term, which is parabolic.

The hyperbolic term can be viewed as a Hamilton-Jacobi part. Let $H(\phi_x, \phi_y) = f\sqrt{\phi_x^2 + \phi_y^2}$. We can approximate this by the Lax-Friedrichs scheme [7]

$$\hat{H} = H\left(\frac{\phi_x^+ + \phi_x^-}{2}, \frac{\phi_y^+ + \phi_y^-}{2}\right) - \alpha^x\left(\frac{\phi_x^+ - \phi_x^-}{2}\right) - \alpha^y\left(\frac{\phi_y^+ - \phi_y^-}{2}\right), \tag{36}$$

where α^x and α^y can be simply chosen as $\alpha^x = \alpha^y = \max_{x,y}|f(x, y)|$, and $\phi_x^\pm, \phi_y^\pm$ can be approximated by first-order forward/backward difference

$$\begin{aligned}(\phi_x^\pm)_{i,j} &\approx (D_x^\pm\phi)_{i,j} = \pm(\phi_{i\pm1,j} - \phi_{i,j}),\\ (\phi_y^\pm)_{i,j} &\approx (D_y^\pm\phi)_{i,j} = \pm(\phi_{i,j\pm1} - \phi_{i,j}),\end{aligned}$$

or by higher order ENO (cf. [28, 29]) or WENO schemes (cf. [16, 17, 20]).

The parabolic term $\tilde{E}(\phi(t))\kappa|\nabla\phi|$ can be discretized by central differencing

schemes

$$\begin{aligned}
\kappa &= \nabla \cdot \frac{\nabla\phi}{|\nabla\phi|} \approx D^- \left(\frac{D^+\phi}{|D^+\phi|} \right) \\
&\approx D_x^- \left(\frac{D_x^+\phi}{\sqrt{(D_x^+\phi)^2 + (D_y^+\phi)^2 + \delta^2}} \right) + D_y^- \left(\frac{D_y^+\phi}{\sqrt{(D_x^+\phi)^2 + (D_y^+\phi)^2 + \delta^2}} \right), \\
|\nabla\phi| &\approx \sqrt{(D_x^0\phi)^2 + (D_y^0\phi)^2} \approx \sqrt{\left(\frac{\phi_x^+ + \phi_x^-}{2} \right)^2 + \left(\frac{\phi_y^+ + \phi_y^-}{2} \right)^2}.
\end{aligned}$$

A tiny value δ^2 (e.g. $\approx 10^{-6}$) is added in computing κ to avoid the singularity occurring at $|\nabla\phi| = 0$. Note that in the last approximation a higher order scheme for $\phi_x^{\pm}$ and $\phi_y^{\pm}$ (from the computing of numerical Hamiltonian term), gives a more accurate $|\nabla\phi|$.

Once $|\nabla\phi|$ is obtained, we can get $\tilde{E}(\phi)$ by numerical integration (e.g., the trapezoidal scheme)

$$\tilde{E}(\phi) \approx \frac{\int_\Omega f(x)(1 - H_\epsilon(\phi))\, dx}{\int_\Omega \delta_\epsilon(\phi)|\nabla\phi|\, dx}.$$

After discretizing the spatial derivative terms, either first order forward Euler or higher order Runge-Kutta (cf. [28]) time discretization can be used to advance the time evolution. Since (35) contains both hyperbolic Hamilton-Jacobi term and parabolic term, the combined CFL condition is given by (cf. [24])

$$\Delta t_n \left(\frac{2\max_{x,y}|f(x,y)|}{h} + \frac{4\max_{x,y}\tilde{E}(\phi(t^n))}{h^2} \right) < 1, \tag{37}$$

where $h = \min(\Delta x, \Delta y)$ and $t^{n+1} = t^n + \Delta t_n$.

In our numerical experiments we use first order forward Euler scheme for time and first order $D^{\pm}$ for spatial difference. The results show that this is good enough for our problems. One important thing to be added to the numerical part is the reinitialization of level set function ϕ. To get a good approximation of the perimeter, it is recommended to choose ϕ as the signed distance function to its 0-isocontour. Since (35) does not guarantee that ϕ maintains this property we add a reinitialization step after each ϕ^n. As in [31], we use the reinitialization equation

$$\begin{cases} \phi_\tau + S(\phi^n)(|\nabla\phi| - 1) = 0, \\ \phi(\tau = 0) = \phi^n, \end{cases} \tag{38}$$

where $S(\phi)$ is a sign function and is numerically smeared out as

$$S(\phi^n) = \frac{\phi^n}{\sqrt{(\phi^n)^2 + (\Delta x)^2}}. \tag{39}$$

A high order scheme should be used for the reinitialization part. In practice, we use WENO + TVD Runge-Kutta scheme and only update (38) one step for every ϕ^n.

For level-set methods special care has to be taken for the evaluation of the Heaviside function $H(\phi)$ and the delta function $\delta(\phi)$ in (30). A general way to overcome this problem in numerical implementation is smearing $H(\phi)$ (cf. [24]). For example, we define the smeared-out Heaviside function as

$$H_\epsilon(\phi) = \begin{cases} 0 & \phi < -\epsilon, \\ \frac{1}{2} + \frac{\phi}{2\epsilon} + \frac{1}{2\pi}\sin\left(\frac{\pi\phi}{\epsilon}\right) & -\epsilon \le \phi \le \epsilon, \\ 1 & \phi > \epsilon, \end{cases} \tag{40}$$

where ϵ is a tunable parameter that determines the size of the bandwidth of numerical smearing (a typical good choice is $\epsilon = 1.5\Delta x$). Then the corresponding derivative is the first-order accurate smeared-out approximation of delta function

$$\delta_\epsilon(\phi) = \begin{cases} 0 & \phi < -\epsilon, \\ \frac{1}{2\epsilon} + \frac{1}{2\epsilon}\cos\left(\frac{\pi\phi}{\epsilon}\right) & -\epsilon \le \phi \le \epsilon, \\ 0 & \phi > \epsilon, \end{cases} \tag{41}$$

Another C^∞ regularization of H and δ was introduced by Chan and Vese in [6]:

$$H_{2,\epsilon}(\phi) = \frac{1}{2}\left(1 + \frac{2}{\pi}\tan^{-1}\left(\frac{\phi}{\epsilon}\right)\right), \quad \delta_{2,\epsilon}(\phi) = \frac{1}{\pi}\frac{\epsilon}{\phi^2 + \epsilon^2}. \tag{42}$$

The difference of $\delta_{2,\epsilon}$ to δ_ϵ is that its support is global, not only around $\phi = 0$.

Because of its simplicity we use (40) with $\epsilon = 1.5\Delta x$ for our numerical computations. For a more sophisticated approach to the approximation of Dirac-function we refer to the recent work [10].

The numerics of the fixed point algorithm in Subsection 4.3 is fairly simple. The corresponding equation (34) can be solved by standard finite differences or finite elements on a rectangular grid. The computation of $|\nabla u_n|$ and the BV-norm can be done as for the ROF-method [24]. The fixed point iteration is stopped as soon as the value of the functional to be maximized $\lambda_n = E(u_n)$ becomes stationary, i.e. the iteration is stopped when the criteria

$$\left|\frac{|\lambda_n - \lambda_{n-1}|}{\lambda_0}\right| \le h \tag{43}$$

with a small h holds. After that a level set scan gives the final result.

5.1 Numerical results

In this part we present some results we obtained with the algorithm described above.

Example 1: we start from a simple example. Here f is the normalized characteristic function $f = \chi_{B_R} - c$, of a circle as mentioned above in (11). The

domain Ω is a 128×128 pixel square (in the examples which follow, we use the same domain if not specified otherwise), (x_0, y_0) is the center of the domain and we choose the radius $R = 20$. This leads to a normalization constant $c \sim -0.08$. The theoretical $G-$norm from (11) is $\|f\|_* = 9.2$. (Although (11) was obtained for the case Ω being a circle we believe that it also holds in this case). We choose ϕ^0 by the level set scanning method as stated in Subsection 4.2 and use the level set method (31) to compute the final result shown in Figure 3. We can see that the initial guess ϕ^0 is almost same as the optimal solution ϕ as we need. The computed G-norm is 9.12, with a small numerical error which we believe is due to the grid size. To show this, in Figure 4 we use two finer grids to compute the G-norm of this circle and get $\|f\|_* = 9.17$ for $h = 0.5$ and $\|f\|_* = 9.20$ for $h = 0.25$. The latter one matches the theoretical result very well.

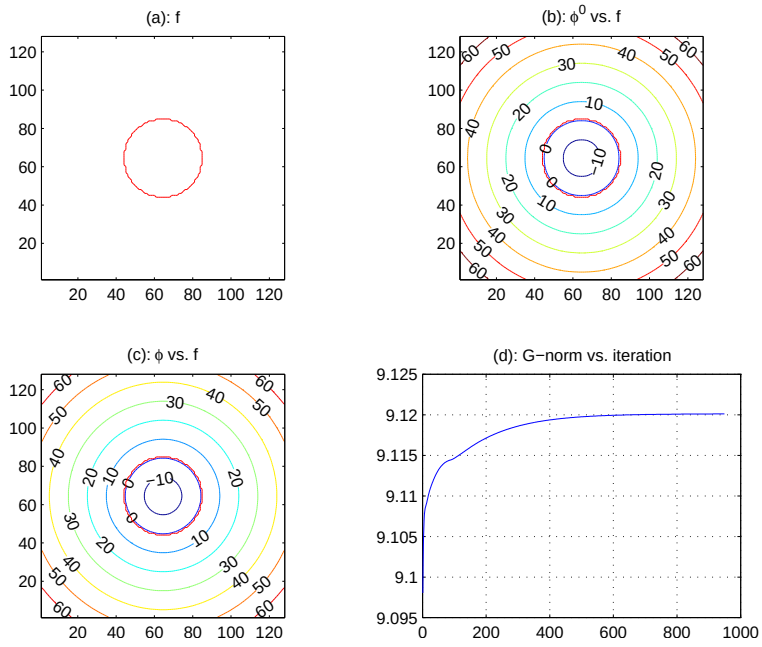

Figure 3: Example 1. (a): 0-isocontour of f, radius $R = 20$, inner=0.92, outer=-0.08; (b): contour of ϕ^0 vs. f; (c) contour of ϕ vs. f; (d) $\tilde{E}(\phi^n)$ vs. n.

Example 2: we take f as a normalized characteristic function of a square with width $L = 40$ and use the level set method. The result is shown in Figure 5. Again we can see that the initial guess ϕ^0 is very close to the final answer ϕ. We state here that although we used about 2600 iterations to get the steady state (where $|\tilde{E}(\phi^n) - \tilde{E}(\phi^{n-1})| < 10^{-6}$), we can also stop at few hundred steps since the change of $\tilde{E}(\phi^n)$ is only about 0.1, which is very small compared to the grid size $\Delta x = \Delta y = 1$. This conclusion also applies to Example 1.

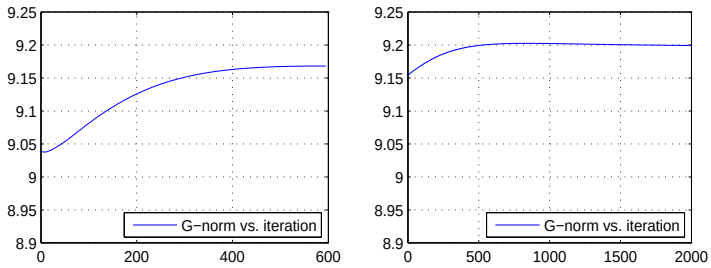

Figure 4: Example 1. circle from Figure 3, finer grids. G-norm $\tilde{E}(\phi^n)$ vs. n. Left: $h = 0.5$ (256×256 grid), $\|f\|_* = 9.17$; right: $h = 0.25$ (512×512 grid), $\|f\|_* = 9.20$.

Example 3: in this example we consider a double-circle experiment, where f is a normalized characteristic function of two circles whose radii are both r and the distance between two centers is d, as shown in Figure 1.

We take the radius of the circles $r = 18$, after normalization we find that the solution to (6) is the convex hull of the circles (case 1 in Figure 2) if d is less than about 50, and case 2 and case 3 in Figure 2 are the maximizers if the distance d is larger than 50. We take two values of d. Figure 6 shows the result for $d_1 = 44 < 50$, the optimal ϕ is then the one stated in case 1. Figure 7 shows the result for $d = 64 > 50$, the optimal ϕ is the one stated in case 2.

Example 4: in this example we try to compute the $G-$norm on "real" images. Given f as gray level image data, we normalize it to mean 0 and perform two tests. The results are shown in Figure 8 and 9 for a shape image (Example 4.1) and a fingerprint image (Example 4.2). In Figure 10 we show the corresponding level sets for the fixed-point algorithm. It can be seen, that the results for the finger image coincide in both cases. However, for the shape image the fixed-point iteration gives a slightly different result. However, the corresponding G-norm is close to the one from the level set approach, with about 1% difference. In our opinion the reason for the different result comes from the fact that the level set method got stuck in a local minimum.

Denoising Example 5: We now turn to apply our algorithms to the denoising case. In Figure 11 we show the data for this case. This image was computed as the v-component of an ROF-decomposition for the well-known Lenna image with Gaussian random noise. All the high-level information was extracted and the remaining image contains mostly noise and texture. However, note that by (21) in principle u can be extracted from v without knowledge of the original image or the parameter λ. We applied the fixed-point algorithm to the image in Figure 11 and found the result shown in Figure 12. This shows the computed characteristic function (as a black and white image), which is a maximizer of (6) for $f = v$. This is an astonishing result, since in the data Figure 11 hardly any information can be seen, yet in the computed result in Figure 12 the Lenna-image can be clearly perceived. In this case our algorithm clearly beats human

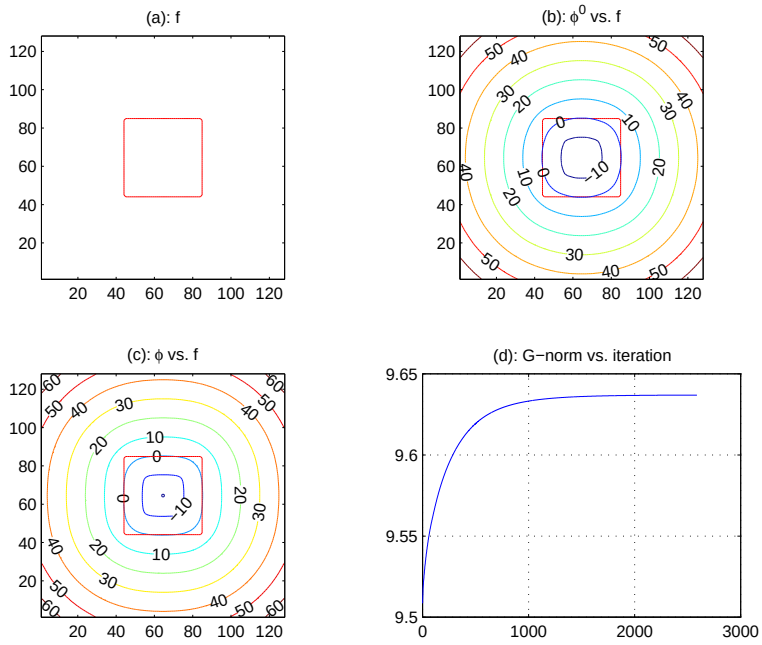

Figure 5: Example 2. (a): 0-isocontour of f, width $L = 40$, inner=0.902, outer=-0.098; (b): contour of ϕ^0 vs. f; (c) contour of ϕ vs. f; (d) $\tilde{E}(\phi^n)$ vs. n

perception.

We also tried to process Figure 11 with the level set approach, but the results were rather poor for two reasons: 1. Using a bad initial guess the level set iteration seems to get stuck in local minima. 2. Computing an initial guess by solving the Poisson equation and performing a level set scan only gives a very ragged result, with lots of small regions which could not be used as an initial guess because of non-smoothness. Probably a more advanced pre-processing can give a suitable initial guess, but we leave this topic to future work.

Let us mention, that a simpler variant of the fixed-point algorithm also gives reasonable results for this case. With the fixed-point algorithm the main edges can already be seen after one iteration. Indeed as the first iteration is a Poisson-equation solver $\Delta^{-1} f$ (with data f) even this simple linear method reveals some structures within the noisy image. The resulting image is shown in Figure 13. However, it should be noted that there is still some noise is present after applying the Poisson solver. Moreover the differences in the gray levels are very small and can only be seen after scaling the image to e.g 256 gray levels. This indicates that the important information on the image is not contained in the gray values of the filtered image $\Delta^{-1} f$ but in its level sets. The conclusion is that the Poisson solver together with a level-set scan is an simple method which might be good enough in some cases to find an initial guess. However,

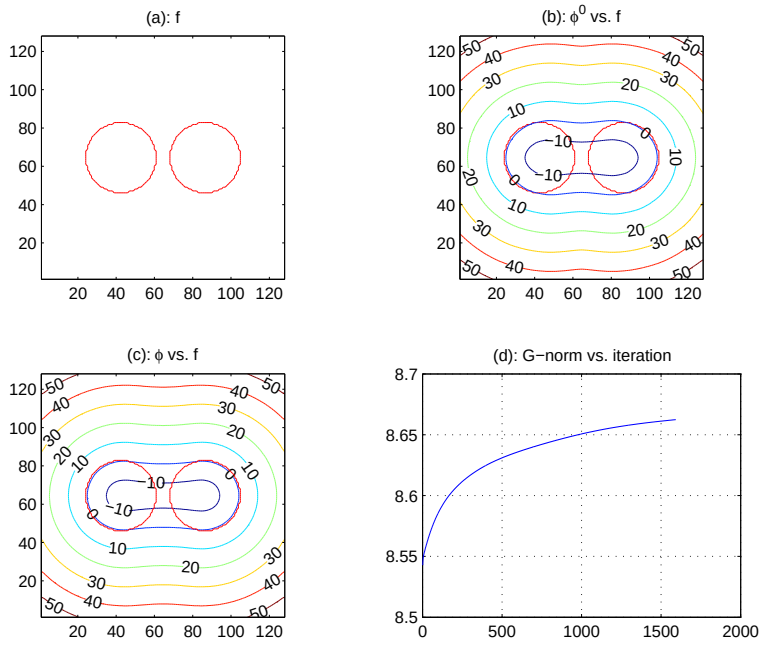

Figure 6: Example 3.1 double circle experiment. $d = 44, r = 18$.

the results are not very regular due to noise (and hence cannot always be used as initial guess for a level-set iteration for instance) and since the gray-levels are very close, a level-set scan might give a wrong result. So for a more reliable result it is important to perform more iterations.

Furthermore we should report, that after many iterations of the fixed-point method the images deteriorates again as it becomes more noisy. In our opinion numerical errors are to blame for this, and it might be related to the ill-posedness of the problem. The effect of this so-called semiconvergence [9] is well-known for ill-posed problems, and can be treated simply by stopping the iteration after a fixed time, according to some stopping criteria. Although this is an important issue and there exists quite an extensive theory on stopping criteria and parameter choice rules (see e.g. [9]), for the sake of brevity we do not address the ill-posedness in this work. As a rule of thumb we stop the iteration after the stopping criteria (43) with $h = 10^{-3}$ is fulfilled or at most at 10 to 20 iterations.

The results of applying our algorithm to Figure 11 are that good, because we know in advance that the input to the algorithm is the v-component of the ROF-decomposition. If the input is any function, the output u is not always related to features in the image. However, we still can apply the method to noisy images, and use it as a denoising method. The denoising method works well, if the exact image is simple - e.g. the characteristic function of a simple

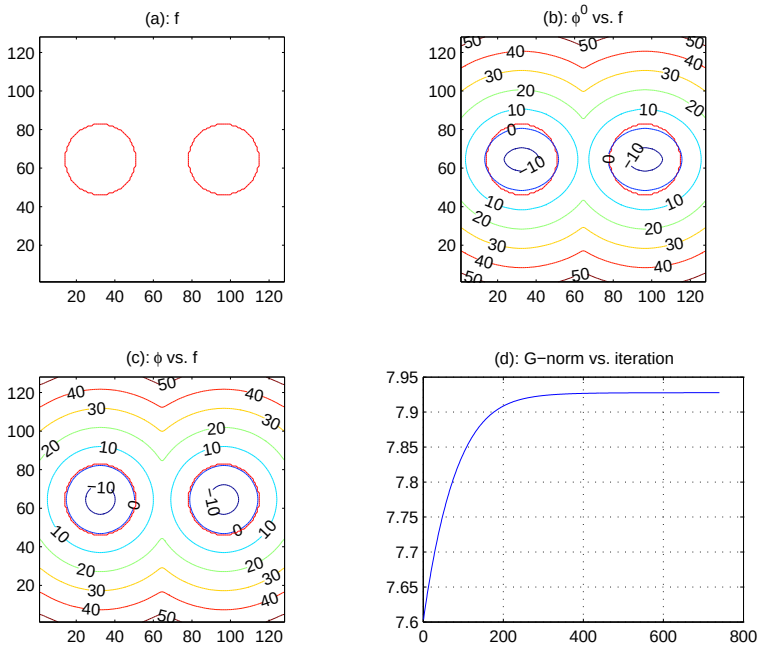

Figure 7: Example 3.2 double circle experiment. $d = 64, r = 18$.

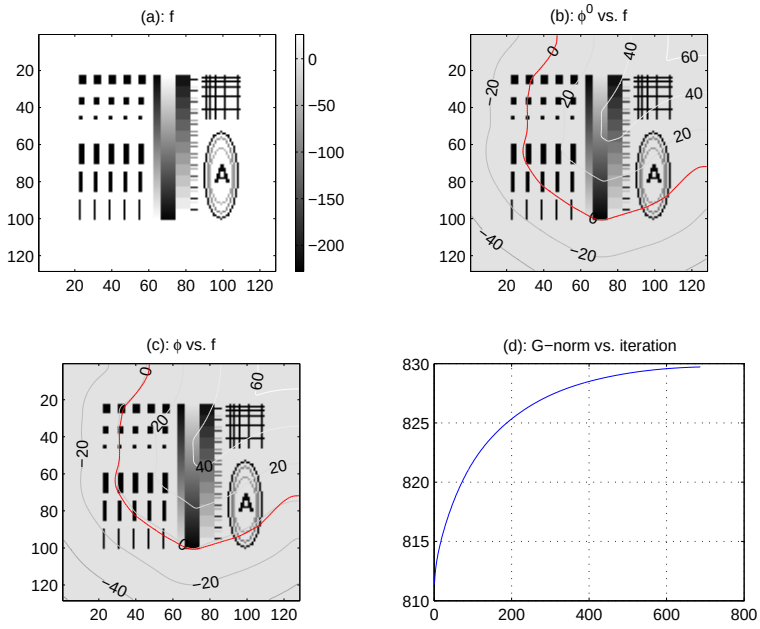

Figure 8: Example 4.1 normalized shape image. mean 0.

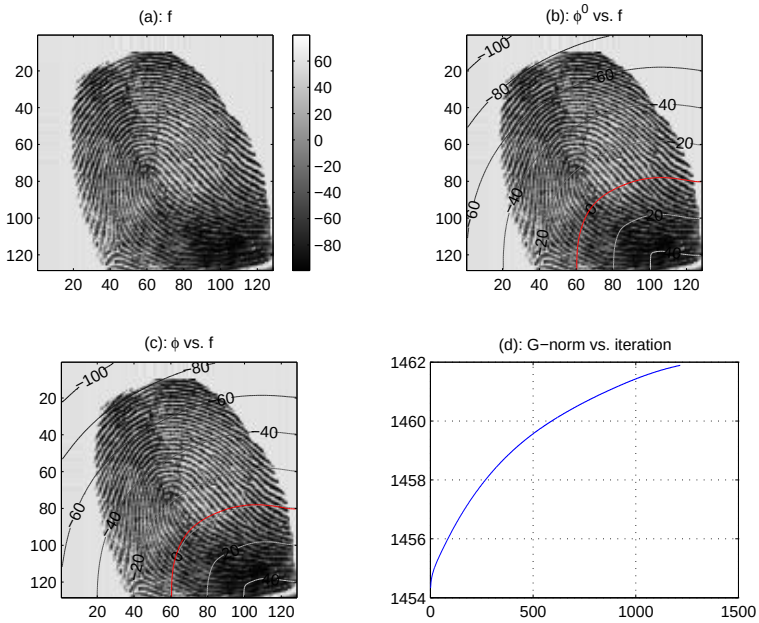

Figure 9: Example 4.2 normalized finger image. mean 0.

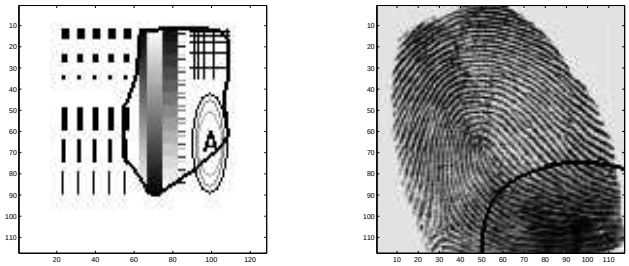

Figure 10: Example 4.1 and 4.2 using fixed-point iteration

geometric structure, and the noisy image contains a high level of Gaussian noise. This is done in the Examples 6–9, Figures 14–17. Here the exact image was a characteristic function of a circle or a square, and the data were highly noisy. In all these cases the left picture contains the noisy data, the center picture shows the result of the fixed-point algorithm, and the right hand side the exact solution.

Denoising Examples 6–7: In Example 6 we used a characteristic function of a circle and added Gaussian noise. The L^2-norm of the noise is about 700% the norm of the exact data, so hardly anything can be seen in the data Figure 14 (left). Nevertheless our method still gives a reasonable approximation or the circle. Example 7 is the same but with the characteristic function of a square as an exact image. Again, the computed result is almost the exact image, except that the corners are rounded. But the result is to be expected, since the rounded square is the extremal set.

Denoising Examples 8–9: Examples 8 and 9 are inspired by Equation (23). Note that for the L^1-BV decomposition u can be found from the sign-function of sign(v), which contains even less information than v. Hence, it should be possible to extract information, for noisy binary (or black and white) images. To create the data for Example 8 we used the same exact solution (a characteristic function of a circle) as in Example 6, then added a high amount of Gaussian noise, and then we applied the sign-function to obtain a noisy binary black and white image, seen in Figure 16 on the left. Note that these data have only two gray levels, but still Figure 16 in the center shows that our method can extract the exact image quite precisely. In Example 9 the same procedure was applied to the square example similar to Example 7.

Let us note that similar results can be found by using the anisotropic version of the BV-norm (24), as explained in Section 3. For instance, using the norm $|u| = \int |u_x| + |u_y| dx$ gives results, where the corners of rectangles lined up with the coordinate axes are not rounded. Hence this method is favorable for rectangular objects with no rounded boundaries.

Concerning the question of complexity of our algorithms it can be said, that each iteration of the fixed-point algorithm, level-set iteration and the level-set scan can be done quite fast all in the order of $O(n)$, where n is the number of pixels. In our test examples we always used images up to size 512×512 pixels, and in these cases about 20 fixed-point iterations were enough.

Finally, let us mention that the idea of reconstructing the cartoon part u from the noise part $v = f - u$ is possible for other variational image decomposition methods as well. For linear filter method such as Tikhonov regularization with an H^1-norm this can be done easily by solving a Poisson equation – as can be seen from the optimality conditions. For the ROF-method we are faced with the nonlinear optimization problem (21) as discussed in this paper. A similar reconstruction can also be done for the Meyer-decomposition. In this case the optimality conditions [18] indicate that a problem like (21) has to be solved twice with different functions f. It is not clear if such a reconstruction can be performed for a method like the Bregman iteration [23]. In this case the cartoon part of the k-th iterate u_k is not directly coupled to the error $f - u_k$ by some

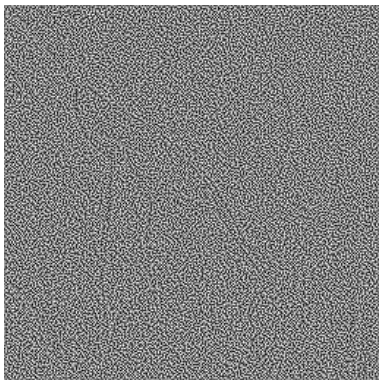

Figure 11: Example 5: Data

Figure 12: Example 5: Result

Figure 13: Example 5: Result after one iteration

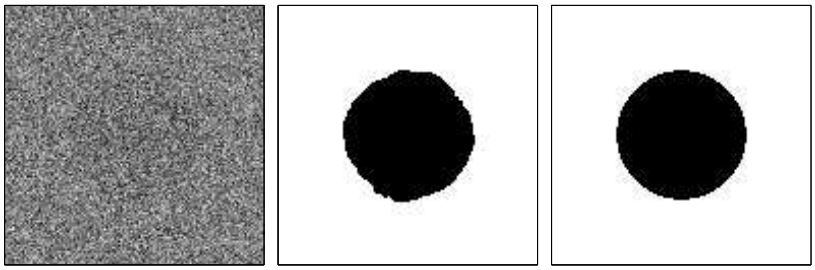

Figure 14: Example 6: Input-data, Result, Exact Solution

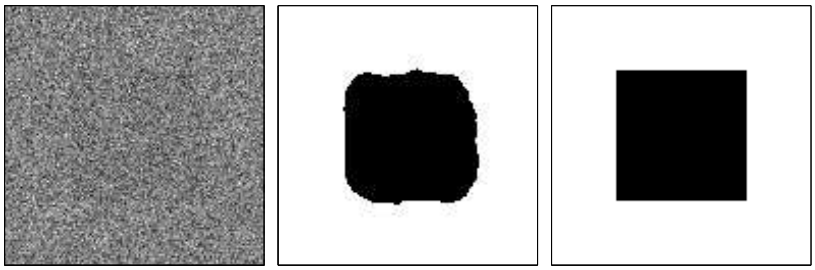

Figure 15: Example 7: Input-data, Result, Exact Solution

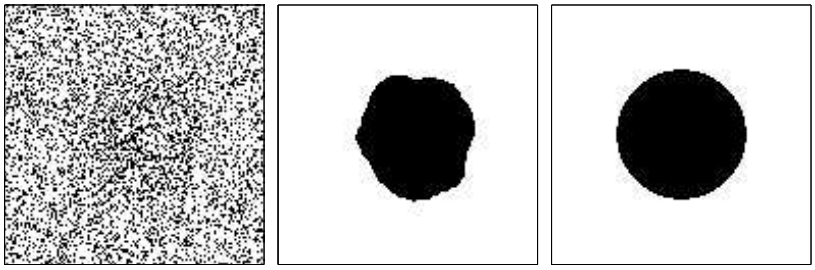

Figure 16: Example 8: Input-data, Result, Exact Solution

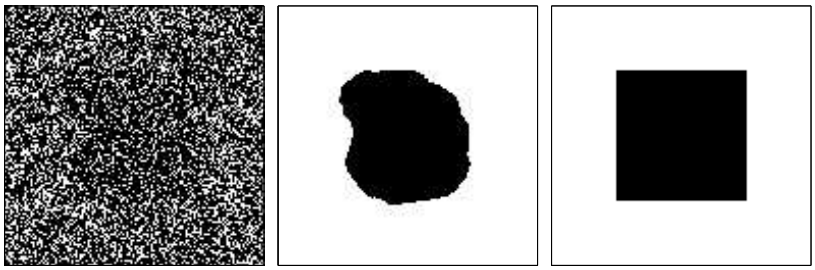

Figure 17: Example 9: Input-data, Result, Exact Solution

optimality condition, but depends also on the previous iterates.

6 Conclusion

We presented two algorithms for computing the G-norm. The methods are based on two observations: The dual representation (6) is the basis for our computation, while Strang's formula Corollary 2.4 is somehow used as stabilization, since we can restrict the optimization problem to the much smaller set of characteristic functions. Our algorithms can be used as a form of denoising, where almost no visual information is contained in the data. The result show this method can extract features which are hardly visible to the human eye.

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), pp. 163–182.

[4] F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle, *Image decomposition into a bounded variation component and an oscillating component*, Journal of Math. Imag. and Vis., 22 (2005), pp. 71–88.

[5] T. Chan, S. Esedoḡlu, and M. Nikolova, *Algorithms for finding global minimizers of image segmentation and denoising models*, preprint, CAM-report, UCLA, 2004.

[6] T. F. Chan and L. A. Vese, *Active contours without edges.*, IEEE Trans. Image Process., 10 (2001), pp. 266–277.

[7] M. G. Crandall and P.-L. Lions, *Two approximations of solutions of Hamilton-Jacobi equations*, Math. Comp., 43 (1984), pp. 1–19.

[8] I. Ekeland and R. Témam, *Convex analysis and variational problems*, vol. 28, SIAM, Philadelphia, PA, 1999.

[9] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.

[10] B. Enquist, A.-K. Tornberg, and R. Tsai, *Discretization of dirac delta functions in level set methods*, J. Comp. Phys., 207 (2005), pp. 28–51.

[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] S. Esedoḡlu and R. Tsai, *Threshold dynamics for the piecewise constant mumford-shah functional*, J. Comput. Phys., 211 (2006), pp. 367–384.

[13] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equation of Second Order*, Springer, Berlin, Heidelberg, 1977.

[14] E. Giusti, *Minimal surfaces and functions of bounded variation*, vol. 80 of Monographs in Mathematics, Birkhäuser Verlag, Basel, 1984.

[15] A. Haddard and Y. Meyer, *Variational methods in image processing*, preprint, CAM-report, UCLA, 2004.

[16] G.-S. Jiang and D. Peng, *Weighted ENO schemes for Hamilton-Jacobi equations*, SIAM J. Sci. Comput., 21 (2000), pp. 2126–2143.

[17] G.-S. Jiang and C.-W. Shu, *Efficient implementation of weighted ENO schemes*, J. Comput. Phys., 126 (1996), pp. 202–228.

[18] S. Kindermann and S. Osher, *Saddle point formulation for a cartoon-texture decomposition*, preprint, CAM-report, UCLA, 2005.

[19] T. M. Le and L. Vese, *Image decomposition using total variation and div(BMO)*, Multisc. Model. and Simul, 4 (2005), pp. 390–423.

[20] X.-D. Liu, S. Osher, and T. Chan, *Weighted essentially non-oscillatory schemes*, J. Comput. Phys., 115 (1994), pp. 200–212.

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

[22] D. Mumford and J. Shah, *Optimal approximations by piecewise smooth functions and associated variational problems*, Comm. Pure Appl. Math., 42 (1989), pp. 577–685.

[23] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterative regularization method for total variation based image restauration*, Multiscale Model. and Simul., 4 (2005), pp. 460–489.

[24] S. Osher and R. Fedkiw, *Level set methods and dynamic implicit surfaces*, vol. 153 of Applied Mathematical Sciences, Springer-Verlag, New York, 2003.

[25] S. Osher and O. Scherzer, *G-norm properties of bounded variation regularization*, Commun. Math. Sci, 2 (2004), pp. 237–254.

[26] S. Osher and J. A. Sethian, *Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations*, J. Comput. Phys., 79 (1988), pp. 12–49.

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

[28] C.-W. Shu and S. Osher, *Efficient implementation of essentially nonoscillatory shock-capturing schemes*, J. Comput. Phys., 77 (1988), pp. 439–471.

[29] ——, *Efficient implementation of essentially nonoscillatory shock-capturing schemes. II*, J. Comput. Phys., 83 (1989), pp. 32–78.

[30] G. Strang, *L^1 and L^∞ approximation of vector fields in the plane*, in Nonlinear partial differential equations in applied science (Tokyo, 1982), vol. 81 of North-Holland Math. Stud., North-Holland, Amsterdam, 1983, pp. 273–288.

[31] M. Sussman, P. Smereka, and S. Osher, *A level set approach for computing solutions to incompressible two-phase flow.*, J. Comput. Phys, 114 (1994), pp. 146–159.

[32] L. A. Vese and S. J. Osher, *Image denoising and decomposition with total variation minimization and oscillatory functions*, J. Math. Imaging Vision, 20 (2004), pp. 7–18.

[33] E. Zeidler, *Nonlinear functional analysis and its applications. III*, Springer-Verlag, New York, 1985.