AN ITERATIVE REGULARIZATION METHOD FOR TOTAL VARIATION-BASED IMAGE RESTORATION[*]

STANLEY OSHER[†], MARTIN BURGER[‡], DONALD GOLDFARB[§], JINJUN XU[†], AND WOTAO YIN[§]

Abstract. We introduce a new iterative regularization procedure for inverse problems based on the use of Bregman distances, with particular focus on problems arising in image processing. We are motivated by the problem of restoring noisy and blurry images via variational methods by using total variation regularization. We obtain rigorous convergence results and effective stopping criteria for the general procedure. The numerical results for denoising appear to give significant improvement over standard models, and preliminary results for deblurring/denoising are very encouraging.

Key words. iterative regularization, total variation, Bregman distances, denoising, deblurring

AMS subject classifications. 47A52, 49M30, 65J22, 68U18

DOI. 10.1137/040605412

1. Introduction. The aim of this paper is to introduce a new iterative regularization procedure for inverse problems based on the use of Bregman distances, with particular focus on problems arising in imaging. We are motivated by the following classical problem in image restoration:

Given a noisy image $f : \Omega \to \mathbb{R}$, where Ω is a bounded open subset of $\mathbb{R}^2$, we want to obtain a decomposition,

$$f = u + v,$$

where u is the true signal and v is the noise.

This problem has a very long history (cf. [30]). One of the most successful and popular techniques for approximating the solution of this problem is due to Rudin, Osher, and Fatemi [38] and is defined as follows:

$$(1.1) \qquad u = \operatorname*{arg\,min}_{u \in BV(\Omega)} \left\{ |u|_{BV} + \lambda \|f - u\|_{L^2}^2 \right\}$$

for some scale parameter $\lambda > 0$, where $BV(\Omega)$ denotes the space of functions with bounded variation on Ω and $|\cdot|$ denotes the BV seminorm, formally given by

$$|u|_{BV} = \int_{\Omega} |\nabla u|,$$

[†]Department of Mathematics, UCLA, 520 Portola Plaza, Los Angeles, CA 90095 (sjo@math.ucla.edu, jjxu@math.ucla.edu).

[‡]Institut für Industriemathematik, Johannes Kepler Universität, Altenbergerstr. 69, A 4040 Linz, Austria (martin.burger@jku.at).

[§]IEOR Department, Columbia University, New York, NY 10027 (goldfarb@columbia.edu, wotao.yin@columbia.edu).

which is also referred to as the *total variation* of u. We call this variational problem the Rudin–Osher–Fatemi (ROF) model. It has been used and analyzed by several authors in several different contexts (cf. [1, 8, 12, 17, 31, 36, 47]). Also, in [12] and subsequently in [31, 32, 33, 36] the "staircasing" effect of this model was analyzed. No completely satisfying remedy has yet been found; e.g., see our results in Figures 1(c)–1(f) in this work for $0 \leq x \leq 120$. In spite of this phenomenon, the ROF model is still quite popular.

The use of the BV seminorm is essential since it allows us to recover images with edges. It is well known that this would be impossible if the first term in (1.1) were replaced by $J_p(u) := \int_\Omega |\nabla u|^p$ for any $p > 1$, which might seem more attractive at first glance due to differentiability and strict convexity. The main reason for this effect is that for $p > 1$ the derivative of J_p corresponds to a nondegenerate elliptic differential operator of second order and thus has a smoothing effect in the optimality condition, whereas for total variation the operator is degenerate and affects only the level lines of the image.

The main results in this work concern an iterative regularization procedure designed to improve ROF restoration and its generalizations. Instead of stopping after recovering the minimizer u in (1.1), we call this solution u_1 and use it to compute u_2, u_3, etc. This is done using the Bregman distance [6], which we shall define in our context in section 3.1. If we call $D(u, v)$ the Bregman distance between u and v associated with the functional J, our algorithm designed to improve (1.1) is

$$(1.2) \qquad u_k = \operatorname*{arg\,min}_{u \in BV(\Omega)} \left\{ D(u, u_{k-1}) + \lambda \| f - u \|_{L^2}^2 \right\}.$$

We obtain a sequence $\{u_k\}$ which we show monotonically converges to f, the noisy image. However, as k increases, for λ sufficiently small the values u_k also monotonically get closer to $\tilde{u}$, the true noise-free image, in a strong sense defined below, until

$$\| u_{\bar{k}} - f \|_{L^2} < \tau \| \tilde{u} - f \|_{L^2},$$

for any $\tau > 1$. The ideal situation is to take λ small and $\bar{k}$ large so that $\bar{k}\lambda$ converges to a critical time $\bar{t}$ at which the estimate above is satisfied. These results are generalized and made precise in section 3.

Iterative procedures involving Bregman distance have been used before in signal processing algorithms, e.g., in [9, 10]. There, and in all the other applications that we are aware of, the goal was to accelerate the computation of the solution to a fixed problem, e.g., to solve the ROF minimization equation (1.1). The approach probably closest to our iterative method is the one in [9], where each iteration step consists in computing

$$u_k = \operatorname*{arg\,min}_u D(u, u_{k-1}) \qquad \text{subject to } \| Ku - f \| \leq \epsilon$$

for some $\epsilon > 0$. The difference, however, is that for increasing iteration, the residual $\| Ku - f \|$ will, in general, not decrease further during the iteration, and hence the iteration procedure yields a smoothing of the solution rather than a closer approximation of the data f. Here our apparently novel idea is to replace the variational problem (1.1) by a sequence (1.2) so as to obtain an improved restoration, or indeed improved solution, to a wide class of inverse problems. Another new aspect of our approach is that we use an iteration with a Bregman distance (in the generalized sense) corresponding to a nondifferentiable functional, the total variation.

We also note that previously in [43] the authors constructed a sequence of approximations $\{u_k\}$ using ROF with a quite different approach, used more to decompose images than to restore them. We comment on this in section 3.5.

The ideal result of the minimization procedure (1.1) would be to decompose f into the true signal u and the additive noise v. In practice, this is not fully attainable. We must expect to find some signal in v and some smoothing of textures in u. The concept "texture" is imprecise so far, and the decomposition depends on the scale parameter λ. Large λ corresponds to very little noise removal, and hence u is close to f. Small λ yields a blurry, oversmoothed u. These statements can be quantified, as discussed below.

In his book [30], Meyer did some very interesting analysis on the ROF model. He began by characterizing textures which he defines as "highly oscillatory patterns in image processing" as elements of the dual space of $BV(\Omega)$. This can be motivated by using the rigorous definition of the BV seminorm

$$(1.3) \qquad \int |\nabla u| = |u|_{BV} = \sup_{|g|_\infty \leq 1,\, g \in C_c^1(\Omega)^2} \int u(\nabla \cdot g).$$

Here $|g| = \sqrt{g_1^2 + g_2^2}$. Defining the space G as the distributional closure of the set

$$\left\{ w = \partial_x g_1 + \partial_y g_2 = \nabla \cdot g \mid g \in C_c^1(\Omega)^2 \right\},$$

equipped with the norm $\|w\|_* = \inf_g \sup_{x,y} |g|$, Meyer showed that elements of this dual space G can be regarded as textures. He also showed that the space G arises implicitly in the ROF model as follows: For $f = u + v$, with u defined by (1.1), we have

$$(1.4) \qquad \|f\|_* < \frac{1}{2\lambda} \implies u = 0,\, v = f;$$

$$(1.5) \qquad \|f\|_* \geq \frac{1}{2\lambda} \implies \|v\|_* = \frac{1}{2\lambda},\, \int uv = \frac{1}{2\lambda}|u|_{BV}.$$

The Euler–Lagrange equation arising in the ROF minimization is

$$(1.6) \qquad -\frac{1}{2\lambda}\nabla \cdot \left(\frac{\nabla u}{|\nabla u|}\right) = f - u = v.$$

Of course the expression on the left in (1.6) needs to be defined when $|\nabla u| = 0$. This is easily done; see [30].

We see that the term v, which was usually thrown away and which represents noise, is an element of G with $*$-norm $\leq \frac{1}{2\lambda}$. This expression is (formally) $-\frac{1}{2\lambda}$ times the curvature of the level contour of $u(x, y)$ at each point. Moreover, by (1.5), it does have $*$-norm $\frac{1}{2\lambda}$ if $\|f\|_* \geq \frac{1}{2\lambda}$, as was shown in [30].

The following interesting example of the effect of ROF minimization was analyzed in [30, p. 36]: Let $f(x, y) = \alpha\chi_R(x, y)$, where $\chi_R(x, y) \equiv 1$ if $\sqrt{x^2 + y^2} \leq R$, $\chi_R(x, y) \equiv 0$ otherwise. Meyer showed that (a) $\|f\|_* = \frac{(\alpha R)}{2}$ and (b) the $f = u + v$ decomposition is as follows: If $\alpha\lambda R \geq 1$, then

$$u = \left(\alpha - \frac{1}{\lambda R}\right)\chi_R, \quad v = \frac{1}{\lambda R}\chi_R.$$

Notice that v is independent of α, which is quite surprising. If $\alpha\lambda R \leq 1$, then $u = 0$, $v = f$. Clearly, the ROF $u + v$ decomposition is defective in this case. The function v is certainly not noise.

Meyer then suggested a modified variational problem:

$$(1.7) \qquad u = \underset{u \in BV(\Omega)}{\arg\min} \left\{ |u|_{BV} + \lambda \|f - u\|_* \right\}.$$

Here we can think of a decomposition, $f = u + v$, where u is a cartoon, or primal sketch, and v is texture plus noise. This model is difficult to minimize using the usual Euler–Lagrange equation approach due to the nonsmoothness of both terms involved in the functional. However, it can be solved effectively as the minimization of a smooth function subject to constraints and, in particular, as a second-order cone program [23].

Vese and Osher [45] approximated Meyer's model by

$$(1.8) \qquad (u, g) = \underset{(u,g)}{\arg\min} \left\{ \int |\nabla u| + \lambda \int |f - u - \nabla \cdot g|^2 + \mu \left(\int |g|^p \right)^{\frac{1}{p}} \right\},$$

with $p \geq 1$ and $\lambda, \mu > 0$. As $\lambda, p \to \infty$, (1.8) approaches Meyer's model. The results displayed in [45] were quite good, especially in separating texture from cartoon. Analytical results were also obtained in [45] (following Meyer's approach in [30]):

$$f = u + v + w, \text{ with } v = \nabla \cdot g.$$

$$(1.9) \qquad \text{If } u = 0, \text{ then } \|f - \nabla \cdot g\|_* \leq \frac{1}{2\lambda};$$

$$(1.10) \qquad \text{If } g = 0, \text{ then } \|\nabla(u - f)\|_q \leq \frac{\mu}{2\lambda}, \text{ where } q = \frac{p}{p-1};$$

$$(1.11) \qquad \text{Both } u = 0, g = 0 \iff \|f\|_* \leq \frac{1}{2\lambda}, \|\nabla f\|_q \leq \frac{\mu}{2\lambda}.$$

Yet another approximation to (1.8) was later constructed by Osher, Solé, and Vese [35]:

$$(1.12) \qquad u = \underset{u \in BV(\Omega)}{\arg\min} \left\{ |u|_{BV} + \lambda \|\nabla\Delta^{-1}(f - u)\|^2 \right\};$$

see [35] for details. The $(L^2)^2$-fitting term used in the ROF model is replaced by an $(H^{-1})^2$-fitting term. This is also an $f = u + v$ model. The resulting Euler–Lagrange equation is equivalent to

$$\frac{1}{2\lambda}\Delta\left(\nabla \cdot \frac{\nabla u}{|\nabla u|}\right) = f - u = v,$$

which is easy to solve, e.g., by using gradient descent [35]. This time v is $\frac{1}{2\lambda}$ times the Laplacian of the curvature of the level contours of u, the cartoon version of f.

Following [35], we can easily show for this model that

$$(1.13) \qquad \|\Delta^{-1}f\|_* \leq \frac{1}{2\lambda} \iff u = 0, v = f;$$

$$(1.14) \qquad \|\Delta^{-1}f\|_* > \frac{1}{2\lambda} \iff \|\Delta^{-1}v\|_* = \frac{1}{2\lambda} \text{ and } \int(-\Delta^{-1}v)u = \frac{1}{2\lambda}|u|_{BV}.$$

We have found experimentally by looking at the error term v found for optimal choice of parameters that this model does a somewhat better job at denoising images than the ROF model (although there is more computational effort involved), but it does not do as well in separating cartoon from texture as the Vese–Osher model [45]. See also [3] for an explanation of this phenomenon.

Additional work on a cartoon/texture decomposition was done in [2] using duality and in [41] using a combination of sparse representations and total variation regularization. One of the many reasons to separate cartoon from texture is to improve image inpainting algorithms. See [5] for a successful approach to this and [4] for a pioneering paper on this subject.

We also mention here that using duality, Chambolle [11] constructed an algorithm solving for v directly in a way that simplifies the calculations needed to solve (1.1), (1.8), and (1.12). Duality was also used in [14] to solve (1.1).

We note that for each choice of λ there is a δ such that problem (1.1) is equivalent to the constrained minimization problem

$$(1.15) \qquad u = \operatorname*{arg\,min}_{u \in BV(\Omega)} \left\{ |u|_{BV} \quad \text{subject to} \quad \|f - u\|_{L^2}^2 = \delta^2 \right\}.$$

Often, one has a reasonable estimate of δ, whereas it is difficult to know how to choose λ in (1.1), which corresponds to the Lagrange multiplier for the noise constraint in (1.15). The original ROF paper [38] used a projected gradient method to solve (1.15). However, the results using the new procedure described in the next section are invariably much better than the constrained denoising of ROF [38]. The error is much smaller and the edges are sharper with our new model.

We will also show how our new procedure can be used for other image restoration tasks, e.g., restoring blurry and noisy images, thus improving the results of [39]. The decomposition in this case becomes

$$f = Au + v,$$

where A is a given compact operator, often a convolution using, e.g., a Gaussian kernel. If A is not known, this becomes a blind deconvolution problem. See [16] for an interesting approach to blind deconvolution, also minimizing functionals involving the BV seminorm.

2. Using geometry and iterative regularization. Our present work has several immediate antecedents. In [44], Tasdizen et al. processed deformable surfaces via the level set method [34]. The idea was used to

(a) first process the unit normals to a given initial surface and

(b) deform the surface so as to simultaneously process it and fit it to the previously computed surface.

The results were visually very pleasing, but no detailed theoretical analysis has yet been obtained.

In [29], Lysaker, Osher, and Tai borrowed the basic idea discussed above and applied it to images as follows. (This is purely formal analysis; see [29] for implementation details.)

- *Step* 1: Given f, compute an approximate normal to the ideal denoised image u_1, computed by carrying out the minimization in (1.1):

$$(2.1) \qquad \vec{n}_1 = \frac{\nabla u_1}{|\nabla u_1|}.$$

- *Step* 2: Replace the ROF minimization in (1.1) by

$$u_2 = \arg\min_{u \in BV(\Omega)} \left\{ \int (|\nabla u| - \vec{n}_1 \cdot \nabla u) + \lambda \int (f - u)^2 \right\}.$$

This minimization procedure attempts to match normals as well as grey-level values.

In [29] the denoised normal $\vec{n}_1$ was computed by using a one-harmonic map as in [46]:

$$\vec{n}_1 = \arg\min_{|\vec{n}|=1} \left\{ \int |\nabla \vec{n}| + \lambda \int \left(\vec{n} - \frac{\nabla f}{|\nabla f|} \right)^2 \right\}.$$

Unlike all the other methods discussed in this paper, this is not a convex minimization problem, and it does not produce an image u_1 satisfying (2.1). Because of this nonconvexity, we decided here to compute $\vec{n}_1$ by using ROF itself:

- *Step* 1: First, solve the ROF model to obtain

$$u_1 = \arg\min_{u \in BV(\Omega)} \left\{ \int |\nabla u| + \lambda \int (f - u)^2 \right\}.$$

  Then define $\vec{n}_1 = \frac{\nabla u_1}{|\nabla u_1|}$.
- *Step* 2: Perform a correction step to obtain

$$u_2 = \arg\min_{u \in BV(\Omega)} \left\{ \int (|\nabla u| - \vec{n}_1 \cdot \nabla u) + \lambda \int (f - u)^2 \right\}.$$

Then we make the following obvious, but crucial, observation:

$$-\int \vec{n}_1 \cdot \nabla u = \int u \nabla \cdot \vec{n}_1 = \int u \left(\nabla \cdot \frac{\nabla u_1}{|\nabla u_1|} \right).$$

But, from the Euler–Lagrange equations for ROF, we have

$$\nabla \cdot \frac{\nabla u_1}{|\nabla u_1|} = -2\lambda(f - u_1) = -2\lambda v_1$$

(recall $f = u_1 + v_1$), and hence $-\int \vec{n}_1 \cdot \nabla u = -\int 2\lambda u v_1$.

We can thus rewrite Step 2 as

$$u_2 = \arg\min_{u \in BV(\Omega)} \left\{ \int |\nabla u| + \lambda \int [(f - u)^2 - 2u v_1] \right\}$$

$$= \arg\min_{u \in BV(\Omega)} \left\{ \int |\nabla u| + \lambda \int (f + v_1 - u)^2 - \lambda \int (v_1^2 + 2v_1 f) \right\}.$$

Since the last integral above is independent of u, we have

$$u_2 = \arg\min_{u \in BV(\Omega)} \left\{ \int |\nabla u| + \lambda \int (f + v_1 - u)^2 \right\}.$$

Remarkably, we are led to the concept that v_1, the "noise" computed by the ROF procedure, should be *added* back to f, the original noisy image, and the sum then processed by the ROF minimization procedure.

2.1. Iterative regularization: Total variation denoising. Clearly, the above regularization process can be repeated. Moreover, the first step can be put into this iterative framework by choosing initial values $u_0 = 0$ and $v_0 = 0$. We shall give precise reasons why this is a good procedure, using the concept of Bregman distance [6, 18] from convex programming, in the next section. Specifically, we are proposing the following *iterative regularization* procedure:

- *Initialize*: $u_0 = 0$ and $v_0 = 0$.
- *For $k = 0, 1, 2, \ldots$*: compute u_{k+1} as a minimizer of the modified ROF model, i.e.,

$$(2.2) \qquad u_{k+1} = \arg\min_{u \in BV(\Omega)} \left\{ |u|_{BV} + \lambda \| f + v_k - u \|_{L^2}^2 \right\},$$

and update

$$(2.3) \qquad\qquad\qquad v_{k+1} = v_k + f - u_{k+1}.$$

We certainly need a stopping criterion, which gives some information for which k we would obtain an approximation as close as possible to the true noise-free image g. In the next section we shall show that the *discrepancy principle* is a reasonable stopping rule; it consists in stopping the iterative procedure the first time the residual $\|u_k - f\|_{L^2}$ is of the same size as the noise level δ. We will prove that some distance measure between the iterate and the true image decreases monotonically until the stopping index is reached and that the regularization procedure enjoys the usual *semiconvergence* properties of iterative regularization methods; i.e., the reconstructed image obtained at the stopping index converges to the true noise-free image as the noise level tends to zero (in a stronger topology than the one of $L^2(\Omega)$). Note that if we do not stop the iteration properly, the iterates would just converge to the noisy image f in $L^2(\Omega)$, and the total variation of the iterates could become unbounded, which is clearly undesirable.

These facts indicate that, for denoising f, a good strategy is to proceed iteratively until the result gets noisier, say, until u_{k+1} is more noisy than u_k. Of course, if we happen to have a good estimate of δ, we can use the discrepancy principle.

It is interesting to further understand how the iterative procedure (2.2), (2.3) works. If we consider why u_2 might look less noisy than u_1, we have

$$u_2 = f + v_1 - v_2 = u_1 + 2v_1 - v_2.$$

This implies that for u_2 to be less noisy than u_1, we need $2v_1 - v_2$ to have more signal than noise. This is indeed the case if the stopping index is greater than one.

It is also clear that our results depend on λ. If λ is very large, we may approximate the noisy image too well, and the stopping index may be satisfied already after the first step. In such a case we may expect a bad reconstruction. If λ is small, we oversmooth initially and can make sure that the stopping index is not satisfied after one step. Our numerical results confirm that the images u_k, $k = 1, 2, \ldots$, become less blurry and noisy until we reach the stopping index. Later they eventually become noisy, converging to the original noisy image f. In numerical experiments we also found out that if λ is sufficiently small, a further decrease does not have a large impact on the final reconstruction. Roughly speaking, by dividing λ by two, the number of iterations needed until the stopping index is reached doubles, and the final reconstruction is almost the same. This fact induces the conjecture that there exists

a limiting flow of images on which our procedure can be interpreted as an implicit time discretization with time step λ. If this is the case, then the dependence of the results on λ is somehow one-sided; i.e., only too large values of λ will create bad reconstructions. The existence and approximation of a continuous flow of images is beyond the scope of this paper, but we plan to investigate it in future research.

Example. It is instructive to see what this procedure does to the specific clean image mentioned in the previous section:

$$(2.4) \qquad f = \alpha \chi_R = \begin{cases} \alpha & \text{if} \quad \sqrt{x^2 + y^2} \leq R; \\ 0 & \text{if} \quad \sqrt{x^2 + y^2} > R. \end{cases}$$

If $\alpha \lambda R \geq 1$, Meyer's result gives us

$$f = u_1 + v_1 = \left(\alpha - \frac{1}{\lambda R} \right) \chi_R + \frac{1}{\lambda R} \chi_R.$$

Then

$$f + v_1 = \left(\alpha + \frac{1}{\lambda R} \right) \chi_R = \alpha \chi_R + \frac{1}{\lambda R} \chi_R = u_2 + v_2.$$

This follows because we merely replace α by $\alpha + \frac{1}{\lambda R}$ in the equation above. So $u_2 = f$, as do all the u_k, $k \geq 2$. The objection that ROF degrades clean images by shrinking extrema is no longer valid.

If $\alpha \lambda R < 1$, we have

$$f = u_1 + v_1 = 0 + \alpha \chi_R;$$
$$f + v_1 = 2\alpha \chi_R.$$

Let n be the smallest integer for which $n\alpha \lambda R \geq 1$. We have $u_{n-1} = 0$, $v_{n-1} = (n-1)\alpha \chi_R$. But $u_n = (n\alpha - \frac{1}{\lambda R})\chi_R$, $v_n = \frac{1}{\lambda R}\chi_R$. Finally, $u_{n+1} = f$, as do all u_k for $k \geq n+1$. This illustrates the strongly nonlinear nature of this iterative procedure. We go from a sequence of "restored" images, all of which are totally black, to the true result in two steps.

The above results also apply to the radially symmetric piecewise constant image f (2.4) if radially symmetric noise that is not too large is added to it. This follows from an analysis of the ROF model by Strong and Chan [42]. Strong and Chan present numerical results that show that their analytical results predict quite well the actual performance of ROF, even on digital images with no radial symmetry.

Chambolle [11] has shown that the problem dual to the total variation regularization (restoration) problem (1.1) is

$$(2.5) \qquad v = \arg \min_{p \in \mathcal{K}} \left\{ \|p - f\|_{L^2}^2 \right\},$$

where

$$(2.6) \qquad \mathcal{K} \equiv \mathrm{cl} \left\{ \frac{1}{2\lambda} \nabla \cdot g \mid g \in \mathcal{C}_c^1(\Omega, \mathbb{R}^2), \; |g(x)| \leq 1 \; \forall x \in \Omega \right\},$$

with closure taken in the space G; i.e., v is a projection of f onto the convex set $\mathcal{K}$. A simple alternative proof of this in the finite-dimensional case can be found in [23]. This minimization problem determines the "noise" v in f, whereas the minimization problem (1.1) determines the "signal" $u = f - v$ in f. The dual version of the iterative regularization procedure (2.2)–(2.3) becomes the following:

- *Initialize*: $v_0 = 0$.
- *For* $k = 0, 1, 2, \ldots$: compute v_{k+1} as the minimizer of the modified dual problem, i.e.,

$$(2.7) \qquad v_{k+1} = \arg \min_{p \in \mathcal{K}} \left\{ \|p - (f + v_k)\|_{L^2}^2 \right\}.$$

Note that

$$(2.8) \qquad u_{k+1} = (f + v_k) - v_{k+1}.$$

Although we do not use this procedure in this paper, we include it for its simplicity and elegance.

Note that had we not introduced the dual iterates v_k and the update (2.3) for them, the expression (2.2) for u_{k+1}, in terms of only the primal iterates u_k, would have had the much more complicated form

$$(2.9) \quad u_{k+1} = \arg \min_{u \in BV(\Omega)} \left\{ |u|_{BV} + \lambda \left\| (k+1)f - \sum_{j=0}^{k} u_j - u \right\|_{L^2}^2 \right\} \text{ for } k = 0, 1, \ldots,$$

where $u_0 = 0$.

2.2. Iterative regularization: General case. The above regularization procedure generalizes to other inverse problems and other regularization models, as we will outline in the following and detail in succeeding papers. Specifically, the algorithm can be generalized to regularization models of the form

$$\min_u \left\{ J(u) + H(u, f) \right\},$$

where J is a convex nonnegative regularization functional (for total variation regularization we have $J(u) = \int |\nabla u|$) and the fitting functional H is convex nonnegative with respect to u for fixed f. As usual for convex functionals (cf. [20]) we shall denote the subdifferential of J at a point u by

$$\partial J(u) := \{ p \in BV(\Omega)^* \mid J(v) \geq J(u) + \langle p, v - u \rangle \ \forall v \in BV(\Omega) \}.$$

After initializing $u_0 = 0$ and $p_0 = 0 \in \partial J(u_0)$, the iterative procedure is given by the sequence of variational problems

$$u_k = \arg \min_u \left\{ J(u) + H(u, f) - \langle u, p_{k-1} \rangle \right\}$$

for $k = 1, 2, \ldots$, where $\langle \cdot, \cdot \rangle$ denotes the standard duality product and p_{k-1} is a subgradient of J at u_{k-1}.

As particular examples we may consider the following:

- The Vese–Osher model (1.8), where we recall that we also minimize over g at each step. The kth step yields the decomposition $f + w_k = u_{k+1} + \nabla \cdot g_{k+1} + w_{k+1}$, with $w_0 = 0$, via the minimization problem (1.8), with f replaced by $f + w_k$ for $k \geq 0$.
- The Osher–Solé–Vese model, where we merely decompose $f = u_1 + v_1$ and iterate via $f + v_k = u_{k+1} + v_{k+1}$ for $k \geq 1$.

In principle, the iteration procedure can be written down for arbitrary functionals H and J, but the well-definedness of the algorithm is not obvious since one needs the existence of u_k as the minimizer of a variational problem and the existence of an associated subgradient p_k for the next step. This will introduce some conditions on J and H that we shall discuss in further detail below.

3. Analyis of the iterative regularization procedure. In the following we shall provide a detailed analysis for the most important case of functionals we are interested in, namely

$$(3.1) \qquad\qquad J(u) := |u|_{BV}$$

and

$$(3.2) \qquad\qquad H(u, f) := \frac{\lambda}{2}\|f - Ku\|^2,$$

with $K : L^2(\Omega) \to \mathcal{H}$ being a bounded linear operator whose kernel does not include the space of continuous functions, and $\mathcal{H}$ being some Hilbert space. In this case it is easy to see that $N(u) := J(u) + \sqrt{H(u,0)}$ is an equivalent norm on $BV(\Omega)$. We shall discuss the case of more general J and H in section 3.4. For quadratic H we can use Fréchet derivatives instead of subgradients; they are given by

$$\partial_u H(\cdot, f) = \lambda K^*(Ku - f),$$

where K^* denotes the adjoint of K.

Note that due to the definition of $H(\cdot, f)$ on the larger space $L^2(\Omega)$, its gradients can be considered as elements of this space, too, while the gradients of J are in the larger space $BV(\Omega)^*$, in general. This will have some interesting implications for the regularity of subgradients of $J(u_k)$ we obtain through our iterative minimization procedure. Moreover, note that we can extend J to a convex functional on $L^2(\Omega)$ by setting $J(u) = \infty$ for $u \in L^2(\Omega)\backslash BV(\Omega)$. The identity

$$\partial_u(J + H(\cdot, f)) = \partial J + \partial_u H(\cdot, f)$$

holds (in $BV(\Omega)^*$) for any $f \in L^2(\Omega)$. For a proof of this assertion we refer the reader to [20, Proposition 5.6].

Our general iterative regularization procedure can be formulated as follows.

ALGORITHM 1. Let $u_0 = 0$, $p_0 = 0$, and for $k = 1, 2, \ldots$
- compute u_k as a minimizer of the convex functional

$$(3.3) \qquad Q_k(u) := H(u, f) + J(u) - J(u_{k-1}) - \langle p_{k-1}, u - u_{k-1}\rangle,$$

  where $\langle \cdot, \cdot \rangle$ denotes the usual duality product;
- compute $p_k = p_{k-1} + \lambda K^*(f - Ku_k) \in \partial J(u_k)$.

Note that in principle we could also start with different initial values that satisfy $p_0 \in \partial J(u_0)$. Since for $u_0 \neq 0$ an analytic expression for the subgradient is not known, one would have to solve another complicated optimization problem to determine p_0, which seems to not be desirable from a practical standpoint.

3.1. Iterative regularization and Bregman distances. Before considering the well-definedness of the above algorithm, we establish the connection to Bregman distances. For $p \in \partial J(v)$, we define the (nonnegative) quantity

$$D^p(u, v) \equiv D_J^p(u, v) \equiv J(u) - J(v) - \langle p, u - v\rangle,$$

which is known as a generalized Bregman distance associated with $J(\cdot)$ (cf. [6, 18, 28] for an extension to nonsmooth functions). For simplicity, we will drop the dependence on $J(\cdot)$ from the notation $D_J^p(u, v)$ in the following.

For a continuously differentiable functional, there is a unique element in the subdifferential and consequently a unique Bregman distance. In this case the distance is just the difference at the point u between $J(\cdot)$ and the first-order Taylor series approximation to $J(\cdot)$ at v. Moreover, if $J(u)$ is strictly convex, $D^p(u, v)$ is also strictly convex in u for each v, and as a consequence $D^p(u, v) = 0$ if and only if $u = v$.

Even for a continuously differentiable and strictly convex functional, the quantity $D^p(u, v)$ is not a distance in the usual (metric) sense, since, in general, $D(u, v) \neq D(v, u)$ and the triangle inequality does not hold. However, it is a measure of closeness in the sense that $D^p(u, v) \geq 0$ and $D^p(u, v) = 0$ if $u = v$ (if and only if for strictly convex functionals). For the case of a nonsmooth and not strictly convex functional such as the total variation, it is not clear if one can introduce a Bregman distance for arbitrary u and v, since $\partial J(v)$ might be empty or multivalued. However, one can consider a multivalued version of the Bregman distance in this case, i.e., as the set including all $D^p(u, v)$ for all $p \in \partial J(v)$. As we shall prove below, this issue is not important for our purpose, since the iterative regularization algorithm automatically selects a unique subgradient.

As we shall see below, we shall obtain convergence of the reconstructions in the weak-* topology of $BV(\Omega)$ (and by compact embedding also in $L^2(\Omega)$), which is the same kind of convergence one obtains for the reconstructions of the ROF model (cf. [1]). From this viewpoint one may consider the Bregman distance only as an auxiliary term used in the convergence analysis. However, we also obtain monotone decrease of some Bregman distances between the true image and the computed reconstruction. This may be interpreted as an additional indicator of the quality of the reconstruction, though the meaning of the Bregman distance associated with the total variation is difficult to interpret. However, at least for some cases the convergence of Bregman distances can be used to interpret the convergence speed of discontinuities (cf. [7]).

3.2. Well-definedness of the iterates. In the following we show that the iterative procedure in Algorithm 1 is well defined, i.e., that Q_k has a minimizer u_k and that we may find a suitable subgradient p_k. The latter will be obtained from the optimality condition for the minimization of Q_k, which yields an interesting decomposition of f involving "noise" at levels k and $k - 1$ and signal at each level k.

PROPOSITION 3.1. *Assume that J and H are given by* (3.1) *and* (3.2), *respectively, and let $u_0 = 0$ and $p_0 := 0 \in \partial J(u_0)$. Then, for each $k \in \mathbb{N}$ there exists a minimizer u_k of Q_k, and there exists a subgradient $p_k \in \partial J(u_k)$ and $q_k = \partial_u H(u_k, f) = \lambda K^*(u_k - f)$ such that*

$$(3.4) \qquad\qquad p_k + q_k = p_{k-1}.$$

If, in addition, K has no null space, then the minimizer u_k is unique.

Proof. We prove the above result by induction. For $k = 1$, we have $Q_1(u) = J(u) + H(u, f)$, and the existence of minimizers, as well as the optimality condition $p_1 + q_1 = p_0 = 0$, is well known [1]. Moreover, with $r_1 := \lambda(f - Ku_1) \in \mathcal{H}$ we have $p_1 = K^* r_1$.

Now we proceed from $k - 1$ to k and assume that $p_{k-1} = K^* r_{k-1}$ for $r_{k-1} \in \mathcal{H}$. Under the above assumptions, the functional

$$Q_k : u \mapsto J(u) - J(u_{k-1}) + H(u, f) - \langle p_{k-1}, u - u_{k-1} \rangle$$

is weak-* lower semicontinuous (due to convexity and local boundedness; cf. [20]), and it is bounded below by $H(u, f)$ due to the properties of subgradients. Moreover, we can estimate

$$Q_k(u) = J(u) - J(u_{k-1}) - \langle r_{k-1}, f - Ku_{k-1} \rangle$$
$$+ \frac{\lambda}{2} \|Ku - f - \lambda^{-1} r_{k-1}\|^2 - \frac{1}{2\lambda} \|r_{k-1}\|^2$$
$$\geq J(u) - J(u_{k-1}) - \langle r_{k-1}, f - Ku_{k-1} \rangle - \frac{1}{2\lambda} \|r_{k-1}\|^2.$$

Since only the first term on the right-hand side of this inequality is not constant, the boundedness of $Q_k(u)$ implies the boundedness of $J(u)$, and consequently the boundedness of $N(u)$. This shows that the level sets of Q_k are bounded in the norm of $BV(\Omega)$, and therefore they are weak-* compact. Hence, there exists a minimizer of Q_k due to the fundamental theorem of optimization. Moreover, if K has no null space, the strict convexity of $H(\cdot, f)$ and convexity of the other terms imply the strict convexity of Q_k, and therefore the minimizer is unique. Since

$$\partial(-\langle p_{k-1}, \cdot \rangle) = \{-p_{k-1}\},$$

the optimality conditions for this problem imply

$$p_{k-1} \in \partial J(u_k) + \partial_u H(u_k, f),$$

which yields the existence of $p_k \in \partial J(u_k)$ and $q_k = \partial_u H(u_k, f) = \lambda K^*(Ku_k - f)$ satisfying (3.4). With $r_k := r_{k-1} - \lambda(Ku_k - f) \in L^2(\Omega)$ and $p_k := K^* r_k$ we obtain (3.4). □

Note that as a result of (3.4) we obtain that

$$p_k = -\sum_{j=1}^{k} q_j = \lambda \sum_{j=1}^{k} K^*(f - Ku_j);$$

i.e., the subgradient p_k is equal to the adjoint applied to the sum of residuals $f - Ku_j$. Moreover, the iterative algorithm constructs a sequence of minimizers u_k such that there exists $p_k \in L^2(\Omega) \cap \partial J(u_k)$ (for smoothing K we even have p_k in the image of K^*), which can be thought of as a regularity property of u_k, respectively of its level sets. This corresponds to the results of Meyer [30] for the ROF model, showing that the indicator function of a ball may be a solution but not the indicator function of a square. In the same way we could show that the indicator function of a square (or more generally a function whose level sets are squares) cannot arise as an iterate in our regularization procedure. However, the method may still converge to such solutions as $k \to \infty$.

We again consider some special cases:
- *Denoising*: If

    (3.5)
    $$H(u, f) = \lambda \int (f - u)^2,$$

    i.e., K is the identity, we have $\partial_u H(u, f) = 2\lambda(u - f)$, and hence

    $$p_k + 2\lambda(u_k - f) = p_{k-1}, \ k = 1, 2, \ldots, \quad p_0 = 0.$$

    If we set $p_k \equiv 2\lambda v_k$, we obtain our usual decomposition:

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

- *Deblurring*: If

$$(3.6) \qquad H(u, f) = \lambda \int (Au - f)^2$$

for $A : L^2(\Omega) \to L^2(\Omega)$ being a compact linear operator (typically a convolution operator), we have $\partial H_u(u, f) = 2\lambda(A^*(Au - f))$, where A^* is the L^2-adjoint operator, and hence

$$p_k + 2\lambda A^*(Au_k - f) = p_{k-1}.$$

Notice that since $p_0 = 0 = A^*0$ we may conclude inductively that $p_k \in \mathcal{R}(A^*)$, and hence there exist v_k with $p_k = 2\lambda A^* v_k$. Hence, we can alternatively write an update formula for v_k given by

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

Finally, we mention that the Osher–Solé–Vese model (1.12) can also be interpreted as deblurring, with the compact operator $A = \nabla \Delta^{-1}$ and transformed output $\tilde{f} = Af$.

3.3. Convergence analysis. We shall now study some convergence properties of the iterative regularization process. Our analysis below is motivated by that of Hanke [25], who analyzed Levenberg–Marquardt methods for ill-posed problems (also related to nonstationary iterative Tikhonov regularization (cf. [26, 24]) and inverse scale space methods (cf. [40])), which turns out to be a special case of our iterative regularization strategy when using a quadratic regularization functional $J(u) = \|u\|^2$ for some Hilbert space norm.

First, we show two important monotonicity properties of the residual and of the Bregman distance.

PROPOSITION 3.2. *Under the above assumptions, the sequence $H(u_k, f)$ obtained from the iterates of Algorithm 1 is monotonically nonincreasing; we even have*

$$(3.7) \qquad H(u_k, f) \leq H(u_k, f) + D^{p_{k-1}}(u_k, u_{k-1}) \leq H(u_{k-1}, f).$$

Moreover, let u be such that $J(u) < \infty$; then we have

$$(3.8) \qquad D^{p_k}(u, u_k) + D^{p_{k-1}}(u_k, u_{k-1}) + H(u_k, f) \leq H(u, f) + D^{p_{k-1}}(u, u_{k-1}).$$

Proof. From the definition of subgradient and because u_k minimizes $Q_k(u)$ we have

$$H(u_k, f) \leq H(u_k, f) + J(u_k) - J(u_{k-1}) - \langle u_k - u_{k-1}, p_{k-1} \rangle$$
$$= Q_k(u_k) \leq Q_k(u_{k-1}) = H(u_{k-1}, f),$$

which implies (3.7).

Next we use the following interesting identity for Bregman distances, which seems to have first been pointed out in [18]:

$$D^{p_k}(u, u_k) - D^{p_{k-1}}(u, u_{k-1}) + D^{p_{k-1}}(u_k, u_{k-1})$$
$$= J(u) - J(u_k) + \langle u_k - u, p_k \rangle$$
$$- J(u) + J(u_{k-1}) - \langle u_{k-1} - u, p_{k-1} \rangle$$
$$+ J(u_k) - J(u_{k-1}) + \langle u_{k-1} - u_k, p_{k-1} \rangle$$
$$= \langle u_k - u, p_k - p_{k-1} \rangle.$$

Replacing $p_k - p_{k-1}$ by $-q_k$, using (3.4), and using the fact that q_k is a subgradient of $H(\cdot, f)$ at u_k, we obtain

$$D^{p_k}(u, u_k) - D^{p_{k-1}}(u, u_{k-1}) + D^{p_{k-1}}(u_k, u_{k-1}) = \langle q_k, u - u_k \rangle.$$
$$\leq H(u, f) - H(u_k, f),$$

which is equivalent to (3.8). $\quad\square$

If there exists a minimizer $\tilde{u}$ of $H(\cdot, f)$ with $J(\tilde{u}) < \infty$, then we obtain, in particular from the choice $u = \tilde{u}$ in (3.8),

$$
\begin{aligned}
D^{p_k}(\tilde{u}, u_k) &\leq D^{p_k}(\tilde{u}, u_k) + D^{p_{k-1}}(u_k, u_{k-1}) \\
&\leq D^{p_k}(\tilde{u}, u_k) + D^{p_{k-1}}(u_k, u_{k-1}) - H(u_k, f) + H(\tilde{u}, f) \\
&\leq D^{p_{k-1}}(\tilde{u}, u_{k-1}).
\end{aligned}
$$

(3.9)

This result allows us to conclude a general convergence theorem.

THEOREM 3.3 (exact data). *Assume that there exists a minimizer $\tilde{u} \in BV(\Omega)$ of $H(\cdot, f)$ such that $J(\tilde{u}) < \infty$. Then*

$$(3.10) \qquad\qquad H(u_k, f) \leq H(\tilde{u}, f) + \frac{J(\tilde{u})}{k},$$

and, in particular, u_k is a minimizing sequence.

Moreover, u_k has a weak-* convergent subsequence in $BV(\Omega)$, and the limit of each weak-* convergent subsequence is a solution of $Ku = f$. If $\tilde{u}$ is the unique solution of $Ku = f$, then $u_k \to \tilde{u}$ in the weak-* topology in $BV(\Omega)$.

Proof. We now sum (3.8) arriving at

(3.11)

$$D^{p_k}(\tilde{u}, u_k) + \sum_{\nu=1}^{k} \left[D^{p_{\nu-1}}(u_\nu, u_{\nu-1}) + H(u_\nu, f) - H(\tilde{u}, f) \right] \leq D^0(\tilde{u}, u_0) = J(\tilde{u}).$$

From $D^{p_{\nu-1}}(u_\nu, u_{\nu-1}) \geq 0$ and the monotonicity of $H(u_\nu, f)$ due to (3.7), we further conclude that

$$D^{p_k}(\tilde{u}, u_k) + k \left[H(u_k, f) - H(\tilde{u}, f) \right] \leq J(\tilde{u}),$$

and the nonnegativity of the first term implies (3.10).

For $f = K\tilde{u}$, (3.10) implies (together with the monotonicity of $\|Ku_k - f\|^2$)

$$k\lambda \|Ku_k - f\|^2 \leq \lambda \sum_{\nu=1}^{k} \|Ku_\nu - f\|^2 \leq J(\tilde{u}).$$

From (3.10) and (3.4) we obtain

$$
\begin{aligned}
J(\tilde{u}) &\geq \sum_{\nu=1}^{k} D^{p_{\nu-1}}(u_\nu, u_{\nu-1}) = J(u_k) - \sum_{\nu=1}^{k} \langle p_{\nu-1}, u_\nu - u_{\nu-1} \rangle \\
&= J(u_k) - J(u_0) - \langle p_{k-1}, u_k - \tilde{u} \rangle + \sum_{\nu=1}^{k-1} \langle p_\nu - p_{\nu-1}, u_\nu - \tilde{u} \rangle \\
&= J(u_k) - \sum_{\nu=1}^{k-1} \langle q_\nu, u_k - \tilde{u} \rangle - \sum_{\nu=1}^{k-1} \langle q_\nu, u_\nu - \tilde{u} \rangle.
\end{aligned}
$$

Since $q_\nu = 2\lambda K^*(Ku_\nu - f)$ in this case, we may further estimate

$$J(\tilde{u}) \geq J(u_k) - 2\lambda \sum_{\nu=1}^{k-1} \langle Ku_\nu - f, Ku_k - f \rangle - 2\lambda \sum_{\nu=1}^{k-1} \|Ku_\nu - f\|^2$$

$$\geq J(u_k) - k\lambda \|Ku_k - f\|^2 - 3\lambda \sum_{\nu=1}^{k-1} \|Ku_\nu - f\|^2$$

$$\geq J(u_k) - 4J(\tilde{u}).$$

Thus, $J(u_k) \leq 5J(\tilde{u})$, and by equivalence of norms we obtain that

$$\|u_k\|_{BV} \leq C(J(u_k) + \|Ku_k\|),$$

whose right-hand side is uniformly bounded. The further assertions then follow by standard weak-* convergence techniques. $\square$

The above result is a typical convergence result for exact data. In the special case of denoising, it would mean that $f = \tilde{u}$ is of bounded variation; i.e., it does not include any noise that is not of bounded variation. For the specific models of denoising and deblurring considered above, this yields a rate of convergence.

COROLLARY 3.4. *Under the assumptions of Theorem* 3.3, *the following results hold:*

- *For denoising, i.e., for H given by* (3.5), *we have*

(3.12) $$\|f - u_k\|_{L^2} \leq \sqrt{\frac{J(f)}{\lambda k}} = \mathcal{O}\left(k^{-1/2}\right)$$

  if $J(f) < \infty$.
- *For deblurring, i.e., for H given by* (3.6), *we have*

(3.13) $$\|f - Au_k\|_{L^2} \leq \sqrt{\frac{J(\tilde{u})}{\lambda k}} = \mathcal{O}\left(k^{-1/2}\right)$$

  if $f = A\tilde{u}$ *and* $J(\tilde{u}) < \infty$.

Next, we consider the noisy case; i.e., we suppose that g is the true noise-free image and that $\tilde{u}$ is a minimizer of $H(\cdot, g)$ with $H(\tilde{u}, g) = 0$, which satisfies

(3.14) $$H(\tilde{u}, f) \leq \delta^2.$$

The positive number δ^2 can be considered as the noise level (or rather as an estimate for the noise level, which is easier to obtain in practice); we shall specify its meaning for the special cases below.

In medical imaging, for example, one often has a very good estimate of the noise induced by the imaging apparatus obtained by imaging known objects (phantoms). In general, a procedure we have found satisfactory to estimate δ is to restrict the image to a square region which is "quiet" and contains no edges and compute the standard deviation of this restriction of the image.

THEOREM 3.5 (noisy data). *Let* $\tilde{u}$, f, *and* g *be such that* $\tilde{u}$ *is a minimizer of* $H(\cdot, g)$ *and such that* (3.14) *holds. Then, as long as* $H(u_k, f) > \delta^2$ *(i.e., the residual lies above the noise level), the Bregman distance between* u_k *and* $\tilde{u}$ *is decreasing; more precisely,*

$$D^{p_k}(\tilde{u}, u_k) \leq D^{p_k}(\tilde{u}, u_k) + D^{p_{k-1}}(u_k, u_{k-1}) < D^{p_{k-1}}(\tilde{u}, u_{k-1}).$$

Proof. From (3.8) we obtain by inserting (3.14)

$$D^{p_k}(\tilde{u}, u_k) + D^{p_{k-1}}(u_k, u_{k-1}) + H(u_k, f) \leq \delta^2 + D^{p_{k-1}}(\tilde{u}, u_{k-1}).$$

Thus, for $H(u_k, f) > \delta^2$ we may conclude the decrease of $D^{p_k}(\tilde{u}, u_k)$.   □

Note that, due to Theorem 3.5, we obtain that if g is the noise-free image and $\tilde{u}$ is the true solution, iterations actually approach the true solution until the residual in the iteration drops below the noise level.

The result of Theorem 3.5 yields a natural stopping rule, the so-called *generalized discrepancy principle* (cf. [21]), which consists in stopping the iteration at the index $k_* = k_*(\delta, f)$ given by

$$(3.15) \qquad k_* = \max\{k \in \mathbb{N} \mid H(u_k, f) \geq \tau \delta^2\},$$

where $\tau > 1$. Note that due to the monotone decrease of $H(u_k, f)$, which is guaranteed by (3.7), the stopping index k_* is well defined. We also mention that the choice $\tau = 1$ that would seem obvious with respect to the noise estimate is too severe to guarantee the boundedness of $J(u_{k_*})$ and the semiconvergence of the regularization method, as we shall see below, but this statement is also true for other iterative regularization methods (cf. [21]).

If we sum the inequality in the proof of Theorem 3.5, we obtain

$$kH(u_k, f) \leq D^{p_k}(\tilde{u}, u_k) + \sum_{\nu=1}^{k} H(u_\nu, f) \leq \delta^2 k + J(\tilde{u}),$$

i.e.,

$$H(u_k, f) \leq \delta^2 + \frac{J(\tilde{u})}{k}.$$

As a consequence, $k_*(\delta)$ is finite for $\tau > 1$, and, since $H(u_{k_*(\delta)+1}, f) \leq \tau \delta^2$, we have

$$(3.16) \qquad \delta^2(k_*(\delta) + 1) \leq \frac{J(\tilde{u})}{\tau - 1}.$$

THEOREM 3.6 (semiconvergence for noisy data). *Let the assumptions of Theorem 3.5 be satisfied, and let the stopping index k_* be chosen according to (3.15). Moreover, let $K\tilde{u} = f$. Then $J(u_{k_*(\delta)})$ is uniformly bounded in δ, and hence, as $\delta \to 0$, there exists a weak-* convergent subsequence $(u_{k_*(\delta_\ell)})$ in $BV(\Omega)$. If the set $\{k_*(\delta)\}_{\delta \in \mathbb{R}^+}$ is unbounded, the limit of each weak-* convergent subsequence is a solution of $Ku = g$.*

Proof. By analogous reasoning, as in the proof of Theorem 3.3, we can derive an estimate of the form

$$J(u_k) \leq C(J(\tilde{u}) + k\delta^2)$$

for $k \leq k_*(\delta)$ and some positive constant C. From (3.16) we further obtain

$$J(u_{k_*(\delta)}) \leq \frac{\tau C}{\tau - 1},$$

and hence $J(u_{k_*(\delta)})$ is bounded. The existence of converging subsequences then follows from standard weak-* convergence techniques. In order to show that a weak-* limit u satisfies $Ku = g$, we use again the estimate

$$H(u_{k_*(\delta)}, f) \leq \delta^2 + \frac{J(\tilde{u})}{k_*(\delta)}$$

derived above. If $k_*(\delta_\ell) \to \infty$ for some subsequence δ_ℓ, then clearly $H(u_{k_*(\delta)}, f) \to 0$, and from the lower semicontinuity of H in this case we obtain $H(u, g) = 0$ for the limit; i.e., $Ku = g$ for the special H we consider. $\square$

We again consider this relation for the following special cases:

- *Denoising*: for H given by (3.5) we obviously have $\tilde{u} = g$, and hence (3.14) becomes

$$H(g, f) = \lambda \int (f - g)^2 \leq \delta^2.$$

  Thus, $\sigma = \sqrt{\frac{\delta^2}{\lambda}}$ is an estimate for the variance of the noise, which can be obtained from statistical tests in typical applications. The stopping rule consits in terminating the iteration when the residual $\|u - f\|$ drops below this variance estimate σ. For $k \leq k_*$ we actually have the stronger estimate

$$D(g, u_k) + D(u_k, u_{k-1}) + \lambda \left(1 - \frac{1}{\tau}\right) \|u_k - f\|_{L^2}^2 \leq D(g, u_{k-1}).$$

- *Deblurring*: for H given by (3.6) we have $A\tilde{u} = g$, and hence (3.14) is again

$$H(\tilde{u}, f) = \lambda \int (f - g)^2 \leq \delta^2,$$

  and $\sigma = \sqrt{\frac{\delta^2}{\lambda}}$ is an estimate for the variance of the noise in the output image. For $k \leq k_*$ we have

$$D(\tilde{u}, u_k) + D(u_k, u_{k-1}) + \lambda \left(1 - \frac{1}{\tau}\right) \|Au_k - f\|^2 \leq D(\tilde{u}, u_{k-1}).$$

  Note that in the particular case of the Osher–Solé–Vese model [35] in (1.12) we have $A = \nabla\Delta^{-1}$ and $f = \nabla\Delta^{-1} f_0$, where f_0 is the actual noisy image we obtain, and therefore the noise estimate is

$$\lambda \|\nabla\Delta^{-1}(f_0 - \tilde{u})\|_{L^2}^2 \leq \delta^2;$$

  i.e., we need an estimate of the variance of the noise in the H^{-1}-norm.

3.4. Further generalizations. In the following we discuss some possible generalizations of the above procedure with respect to the fitting functional H, the regularization functional J, and additional constraints.

We start with different regularization functionals J. The above analysis is not restricted to the space $BV(\Omega)$ and J being the BV seminorm. One can easily generalize the results to other locally bounded, convex, and nonnegative regularization functionals J defined on a Banach space $\mathcal{U} \subset L^2(\Omega)$. The conditions needed on J are that

- the level sets

$$\{u \in \mathcal{U} \mid J(u) \leq M\}$$

  are compact in $L^2(\Omega)$ (or any stronger topology than the one of $L^2(\Omega)$) for all $M \in \mathbb{R}$ and nonempty for $M > M_0 > 0$;
- J can be extended to a weakly lower semicontinuous functional from $L^2(\Omega)$ to $\mathbb{R} \cup \{+\infty\}$.

Under these conditions, then by similar reasoning as above there exists a minimizer of the functional Q_k, which is the minimal property we need for the well-definedness of the iterative procedure. If, in addition, $J + H(\cdot, f)$ is strictly convex, then this minimizer is unique, and we obtain a unique iterate u_k. From standard optimality theory for convex problems (cf. [20]), we may also conclude the decomposition (3.4) and the regularity $p_k \in L^2(\Omega) \subset \mathcal{U}^*$. The convergence analysis with the same stopping rule can be carried out as above, with the modification that the weak-* topology in BV has to be replaced by the topology in which the level sets of J are compact.

Possible generalizations of the regularization functional include the following:

- *Anisotropic total variation*: In order to obtain different minimizers such as indicator functions of squares as minimizers, one can use anisotropic regularization functionals of the form

$$J(u) = \int_\Omega G(\nabla u),$$

  with $G : \mathbb{R}^2 \to \mathbb{R}^+$ being a continuous one-homogeneous function (cf. [22]). An example of particular interest is $G(\nabla u) = |u_x| + |u_y|$. Of course, we can also use functions, which are not one-homogeneous, such as $G(\nabla u) = \|\nabla u\|^2$, thus including standard Tikhonov-type regularization techniques.

- *Approximations of total variation*: In several instances, one rather minimizes the smooth approximation

$$J_\epsilon(u) = \int_\Omega \sqrt{|\nabla u|^2 + \epsilon^2}$$

  for some $\epsilon > 0$ (cf., e.g., [19]). Such an approximation simplifies numerical computations due to the differentiability of J_ϵ and may help to avoid the staircasing effect in some cases. The analysis can be carried out in the same way as above, and due to the strict convexity of J_ϵ for $\epsilon > 0$ one even obtains that the Bregman distance is a strict distance.

- *Bounded variation norms*: Instead of taking the seminorm in $BV(\Omega)$, one might also use a full norm for the regularization, i.e.,

$$(3.17) \qquad J(u) = |u|_{BV} + \rho\|u\|^2_{L^2},$$

  for $\rho > 0$. In this case, the Bregman distance $D^p(u,v)$ is bounded below by $\rho\|u - v\|^2_{L^2}$, and hence convergence of the Bregman distance implies L^2-convergence, which is interesting, in particular, for deblurring and for more general fitting functionals as outlined below.

- *Derivatives of bounded variation*: Another obvious generalization considered by several authors (cf. [15, 27]) is to use the bounded variation of ∇u, i.e.,

$$J(u) = \int_\Omega |D^2 u|,$$

  where $D^2 u$ denotes the Hessian of u, or even more general functionals of the form

$$J(u) = \int_\Omega \varphi(u, \nabla u, D^2 u),$$

  with convex $\varphi : \mathbb{R} \times \mathbb{R}^2 \times \mathbb{R}^{2\times 2} \to \mathbb{R}_+$. The analysis can be carried out in the Banach space $\mathcal{U} = BV^2(\Omega)$ of functions with second-order bounded variation.

- *Finite-dimensional approximations*: By analogous reasoning, one can consider the discrete version of all the models introduced above and obtain the same type of convergence results.

For generalizations with respect to the fitting functional, the situation is more delicate. In general, even under rather strong assumptions on H, the compactness of level sets of the functional Q_k is not guaranteed, so that the iterates in Algorithm 1 are possibly not well defined. Moreover, we do not know any argument showing that the total variation of u_k remains bounded (even for exact data), so that the convergence analysis cannot be carried out as above.

Finally, generalizations to additional constraints would be of interest in practice. The iterative procedure then consists in minimizing Q_k subject to the additional constraints. This is of importance, e.g., for nonnegativity constraints or for multiplicative noise, where one wants to choose

$$H(u, f) = \int_\Omega \left(\frac{f}{u}\right)^2$$

subject to the constraint (cf. [37])

$$C(u) = -1 + \int_\Omega \frac{f}{u} = 0.$$

If the constraint set is not empty, the analysis of well-definedness of the iterates is of similar difficulty as in the unconstrained case, but the convergence analysis cannot be carried over easily to additional constraints; in particular the update formula (3.4) must involve additional terms corresponding to Lagrange multipliers of the constraints. Since preliminary numerical experiments demonstrate the success of the iterative regularization procedure also for multiplicative denoising, such an analysis seems to be an important task for future research.

3.5. Related work. In an interesting earlier work [43], the authors propose an iterative procedure also based on the ROF model. They also generate a sequence u_k which converges to the given image f. It is interesting to compare the two approaches.

To recall, our approach is to compute u_k as a minimizer of the convex functional

$$Q_k(u) = \frac{\lambda}{2} \int (u - f)^2 + J(u) - J(u_{k-1}) - \langle p_{k-1}, u - u_{k-1} \rangle$$

for $k = 1, 2, \dots$, with $u_0 = 0$, $p_0 = 0$, and to compute $p_k \, \varepsilon \, \{p_{k-1}\} - \lambda(u_k - f) \cap \partial J(u_k)$.

The Tadmor–Nezzar–Vese (TNV) approach is (in our language) as follows: set $u_0 = 0$ and compute u_k as a minimizer of the convex functional

$$\tilde{Q}_k(u) = \frac{\lambda}{2} 2^k \int (u - f)^2 + J(u - u_{k-1})$$

for $k = 1, 2, \dots$.

For $J(u)$ homogeneous of degree one, as in the ROF model, this can be rewritten as follows: minimize

$$\tilde{\tilde{Q}}_k(u) = \frac{\lambda}{2} \int (u - f)^2 + J\left(\frac{u - u_{k-1}}{2^k}\right).$$

Thus we see the differences: (1) The TNV algorithm uses a hierarchical decomposition, where the difference in total variation between u and the previous iterate is computed. (2) A dyadic sequence of scales, $\lambda 2^k$, is used to obtain convergence.

The differences in performance can also be seen. If we define $f = \alpha x_R$ for $\alpha \lambda R \geq 1$ as in section 2, our algorithm recovers $u_k \equiv f$ for all $k \geq 2$. The TNV algorithm finds

$$u_k = \left(\alpha - \frac{1}{\lambda 2^{k-1} R} \right) \chi_R, \qquad k = 1, 2, \ldots.$$

Also, our algorithm has a denoising aspect to it. Theorem 3.5 indicates that our sequence of iterates u_k has the property that Bregman distance between u_k and $\tilde{u}$, the true restored solution, decreases until the discrepancy principle is satisfied. There is no such result in [43]. Finally, we mention that a similar approach as in [43], but without proofs, can also be found in the earlier paper [40].

4. Numerical results. In this section we present some of the results we have obtained using our iterative regularization procedure. We will concentrate on total variation denoising.

We use the following notation and formula: $f = u + v$ (denoise) or $f = Au + v$ (deblur/denoise). $v \sim \mathcal{N}(0, \delta^2)$ is Gaussian noise, and $\|v\|_{L^2} \approx \delta$.

$$SNR := 20 * \log_{10} \left(\frac{\|f - \bar{f}\|_{L^2}}{\|v - \bar{v}\|_{L^2}} \right)$$

is the signal-to-noise ratio, measured in decibels. $\bar{f}$ and $\bar{v}$ are the means of f and v over Ω. Knowledge of the noise level is useful for us only as a stopping criterion via our discrepancy principle. Our results are otherwise independent of the noise level. Variational methods applied to image processing often use noise estimates as constraints.

All solutions to the variational problem (1.1) were obtained using gradient descent in a standard fashion; see, e.g., [38]. This amounts to solving a parabolic equation to steady state with normal derivative zero imposed at the boundaries. The only nontrivial difficulty comes when $|\nabla u| \approx 0$. We fix this, as is usual, by perturbing $J(u) = \int |\nabla u|$ to $J(u) = \int \sqrt{|\nabla u|^2 + \epsilon}$, where ϵ is a small positive number; see, e.g., [19]. In our calculations we took $\epsilon = 1.0e - 12$. The initial guess for the first step was the noisy data f. For succeeding iterates in the Bregman procedure, we merely replace f by $f + v_{k-1}$ and proceed with ROF again, with the initial guess $f + v_{k-1}$ or the previous iterate u_{k-1}.

We first compare the results of our procedure with the constrained denoising used in the original ROF paper [38].

Example 1. In our first test we considered denoising a one-dimensional (1D) signal f. Figure 1(a) displays the original signal. Figure 1(b) displays the noisy signal with $\delta = 9.4544$. In Figure 1(c) we present the restored u obtained using ROF with the constraint $\|u - f\|_{L^2} \leq \delta$. We see the typical loss of accuracy in regions where there are narrow peaks and valleys. In Figures 1(d), 1(e), and 1(f), we present the results of our iterative regularization procedure with $\lambda = 0.005, 0.001,$ and 0.0005, respectively. We display the results for those values of $k = \bar{k}(\lambda, \delta)$ that are obtained from the stopping rule. All four of these results are more accurate than the result obtained with the single step ROF minimization subject to a constraint on the L^2-norm of the removed noise, especially near local minima and maxima of f. The results also confirm numerically the assertion that using a smaller λ (i.e., initially oversmoothing) requires more regularization steps to get the optimal restoration.

Example 2. We next considered another 1D signal. First, we exhibit the restoration of a clean signal. With $\lambda = 0.005$ we obtained u_1 in Figure 2(a) and u_2 in
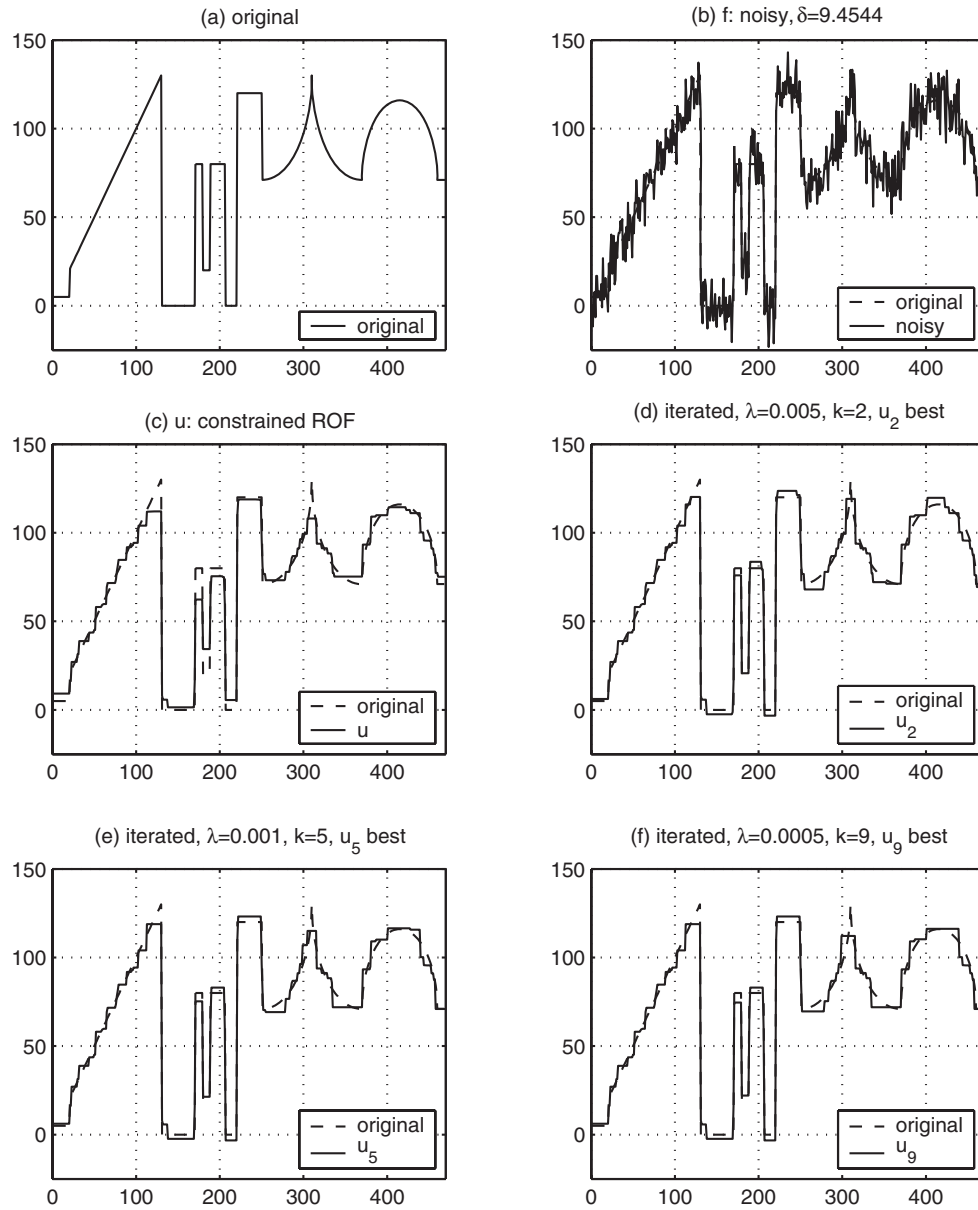
FIG. 1. *1D denoising comparison: constrained ROF vs. ROF with iterative regularization.*

Figure 2(b). The improvement is obvious. Next, for the noisy signal in Figure 2(c) containing Gaussian noise with $\delta = 10$, we display the iterative regularization results u_1, u_2, and u_3 in Figures 2(d), 2(e), and 2(f), respectively. From u_1 to u_2, the results improved considerably. Succeeding u_k become noisy again for $k > 3$.

Example 3. We considered the noisy fingerprint image with $\delta = 10$ and $SNR = 14.8$ and applied the ROF model with $\lambda = 0.085$. This produced a restored image u with $\|f - u\|_{L^2} = 10.2 \approx \delta$. We can see that the removed noise component $v = f - u$ contains some signal. This is a common problem for the ROF model. Figures 3(a)–3(f)
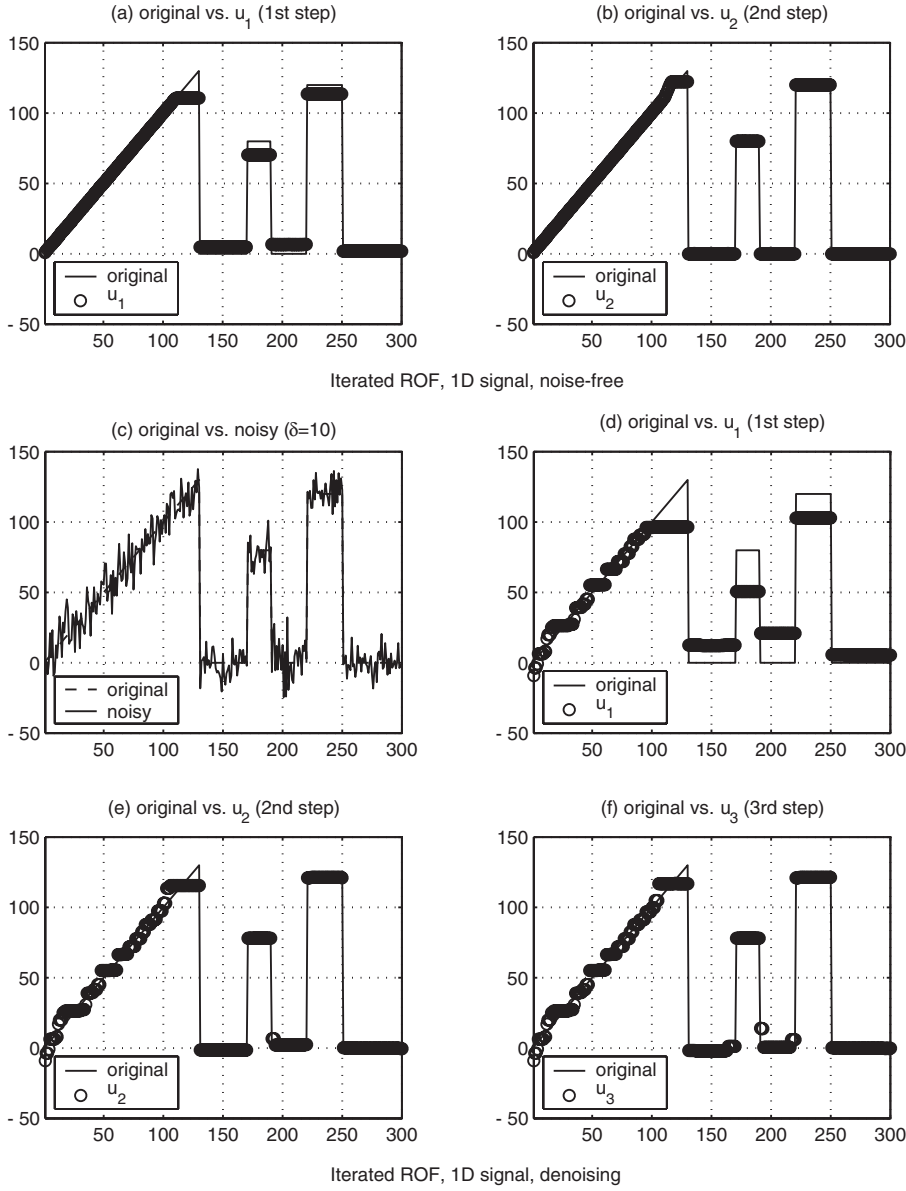
FIG. 2. *ROF and iterative regularization on a* 1*D signal (noise-free and noisy).*

display the original image, noisy image f, and noises η, u, $f - u$, and u-original, respectively. To make the small values of η, $f - u$, and u-original more visible, we added 128 to their values in Figures 3(c), 3(e), and 3(f), respectively. Next, we applied our iterative regularization procedure with $\lambda = 0.013$ to the noisy image in Figure 3(b). Notice that this value of λ is much smaller than the one used to obtain the one-step ROF results in Figures 3(a)–3(f). Figures 4(a)–4(k) display these results. u_1 is oversmoothed, and u_k improves steadily as k increases, with u_4 in Figure 4(g) giving the best restoration. Figure 4(j) shows the residual $f - u_4$, which contains only very little visible signal. In Figure 4(l) we plot $\|f - u_k\|_{L^2}$ as a function of
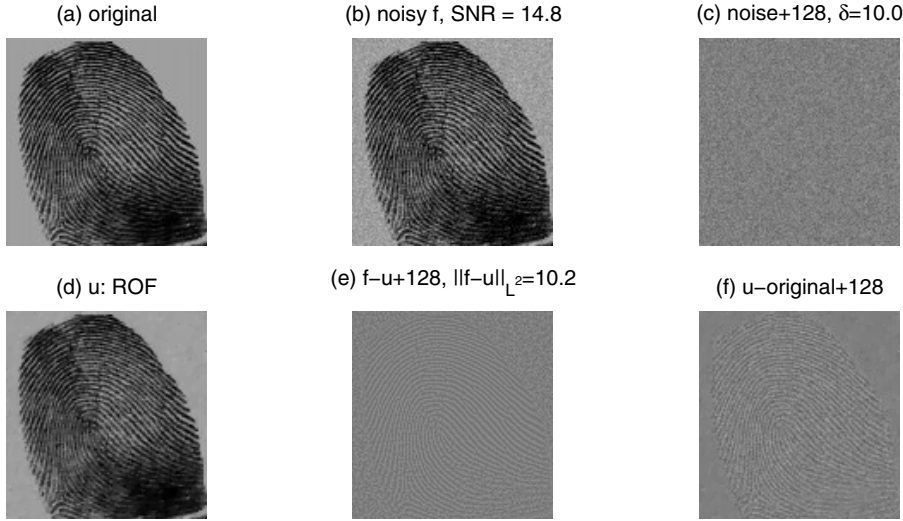
FIG. 3. *ROF with $\lambda = 0.085$, $\|f - u\|_{L^2} \approx \delta$, and signal contained in $v = f - u$.*

the iterate k. This plot shows that $\|f - u_k\|_{L^2}$ decreases monotonically with k, first dropping below δ at the optimal iterate $\bar{k} = 4$, hence validating Theorem 3.5.

Example 4. For this example, we created an image containing various shapes and patterns to test our procedure. Figure 5(a) shows the original image and Figure 5(b) the noisy image with $\delta = 40$ and $SNR = 7.4$. First, we took a small $\lambda = 0.002$. As can be seen in Figures 5(d)–5(r) the results u_k improve until the fifth step, at which point the inequality $\|f - u_5\|_{L^2} < \delta$ first becomes satisfied. Then the succeeding u_k $(k \geq 6)$ become noisier, again validating Theorem 3.5. Next, to illustrate the relationship between λ and the optimal step $\bar{k}(\lambda)$, we chose different λ values and denoised the same noisy image, Figure 5(b). Figure 6 displays the results for $\lambda = 0.004$. The restoration u_3 is the best, $\bar{k}(0.004) = 3$. Figure 7 presents the results for $\lambda = 0.006$ and shows that u_2 is the best, $\bar{k}(0.006) = 2$. Moreover, as previously stated, $\bar{k}(0.002) = 5$. This verifies that $\bar{k}(\lambda)$ monotonically decreases as λ increases or, equivalently, as the amount of initial oversmoothing decreases. In this example we also showed the plots of $\|f - u_k\|_{L^2}$, $\|u - u_k\|_{L^2}$ and $D(u, u_k)$ versus k for u, the original clean image. As we can see $\|u - u_k\|_{L^2}$ also attains its minimum at $\bar{k}$. (This is not generally true for all of our examples. Often the smallest true L^2-error occurs for one or two more regularization iterations.) As predicted by Theorem 3.5, $D(u, u_k)$ is decreasing for $k < \bar{k}$. It is interesting that this quantity sometimes continues to decrease well after noise has returned to the iterate u_k; see Figure 5(r).

Example 5. In this example, we applied our approach to the denoising of a textured image. We used the same noisy data as in Lysaker, Osher, and Tai [29]: $\delta = 17.2$ and $SNR = 9.8$. As can be seen in Figure 8, for $\lambda = 0.02$, the results improved considerably from u_1 to u_2, and u_2 gave the best restoration.

Example 6. Here we considered denoising a satellite image. We added noise with $\delta = 40$ and $SNR = 6.3$. For $\lambda = 0.0055$, u_2 yields the best restoration, with almost all signal restored and very little visible signal in the residual. In u_3 and succeeding u_k, $k > 3$, some noise comes back. This is displayed in Figure 9.

Example 7. We have also used our procedure to restore blurry, noisy images. This will be described in greater detail in future joint work with L. He and A. Marquina,
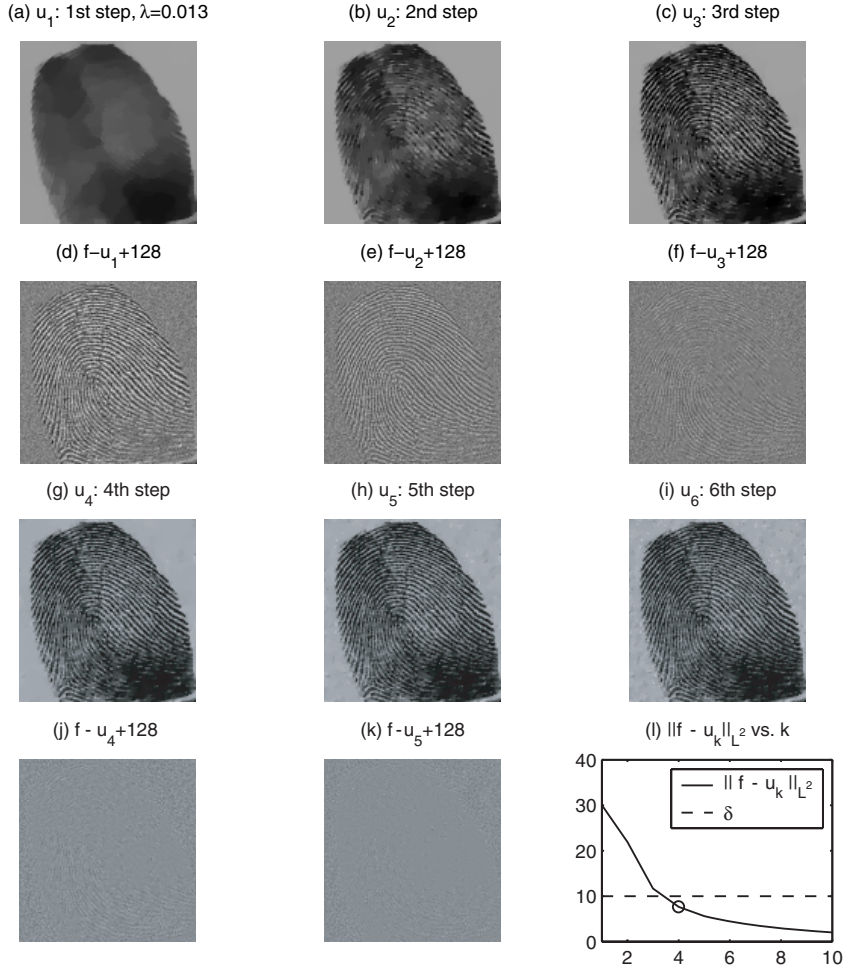
(a) u_1: 1st step, λ=0.013 | (b) u_2: 2nd step | (c) u_3: 3rd step

(d) $f-u_1$+128 | (e) $f-u_2$+128 | (f) $f-u_3$+128

(g) u_4: 4th step | (h) u_5: 5th step | (i) u_6: 6th step

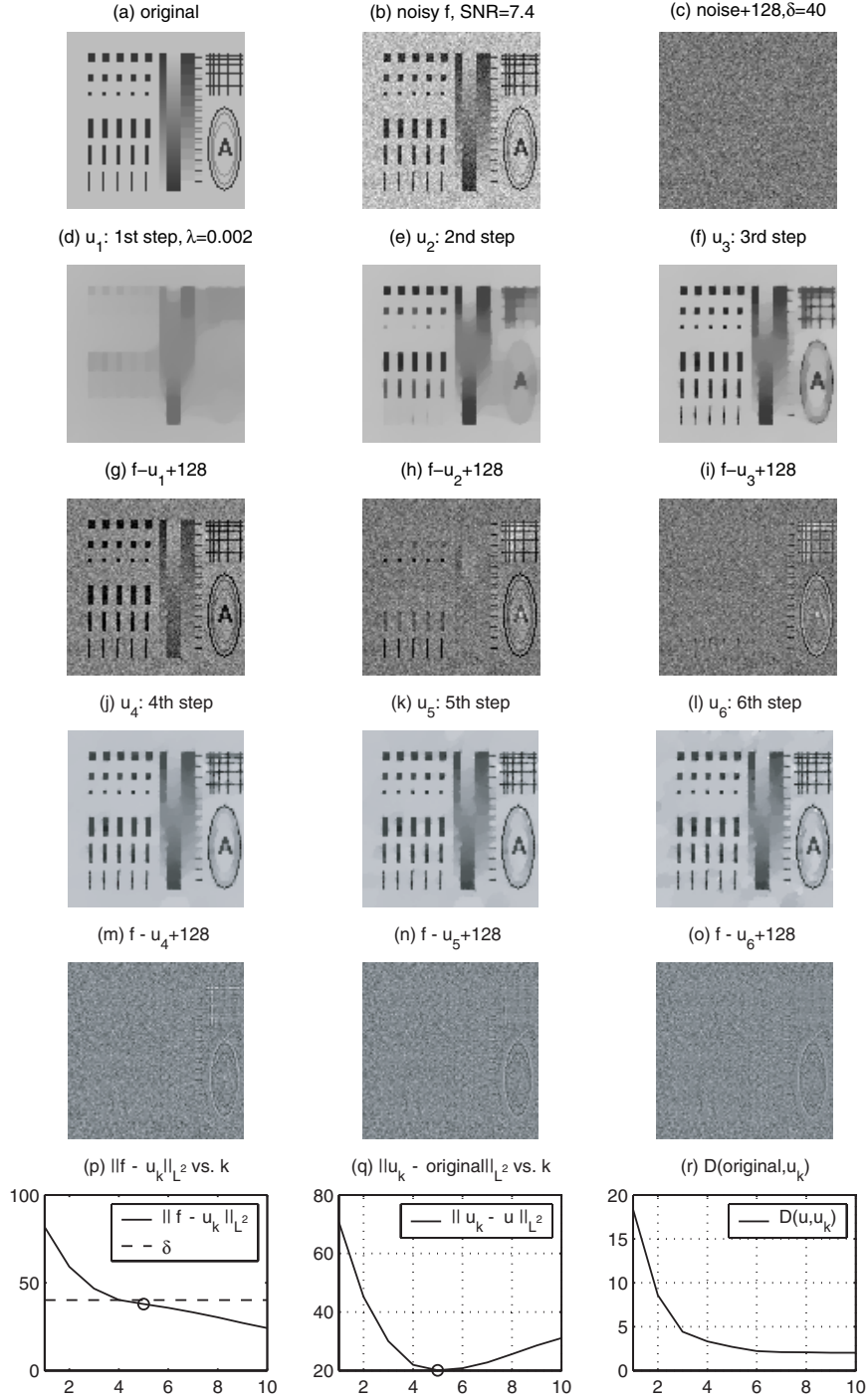(j) $f - u_4$+128 | (k) $f-u_5$+128 | (l) $\|f - u_k\|_{L^2}$ vs. k

FIG. 4. *ROF with iterative regularization with* $\delta = 10$ *and* $\lambda = 0.013$: *best restoration obtained when* $\|f - u_k\|_{L^2}$ *drops below* δ *at* $k = 4$. *Noise returns in succeeding* $u_5, u_6, \ldots$.

and we thank them for their help in obtaining these results. Here we illustrate the use of our approach on a satellite image blurred by a Gaussian kernel with added noise ($\delta = 10$). Figure 10 shows the results for $\lambda = 0.1$. With iterative regularization, u_2 recovers more signal than u_1, especially small details. The restored image u_2 has the least noise, but u_3 appears to be sharper. Succeeding iterations ($k \geq 4$) become noisier.

FIG. 5. *ROF with iterative regularization with* $\delta = 40$ *and* $\lambda = 0.002$*: best restoration obtained when* $\|f - u_k\|_{L^2}$ *drops below* δ *at* $k = 5$*. Noise returns in succeeding* $u_6, u_7, \ldots$
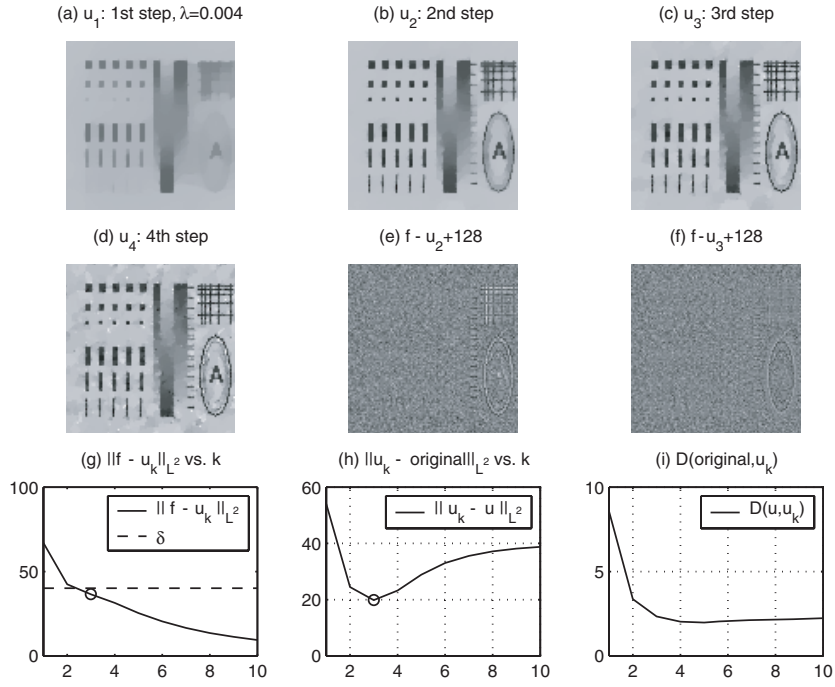
(a) u_1: 1st step, λ=0.004

(b) u_2: 2nd step

(c) u_3: 3rd step

(d) u_4: 4th step

(e) $f - u_2$+128

(f) $f - u_3$+128

(g) $\|f - u_k\|_{L^2}$ vs. k

(h) $\|u_k - \text{original}\|_{L^2}$ vs. k

(i) $D(\text{original}, u_k)$

FIG. 6. *ROF with iterative regularization with $\delta = 40$ and $\lambda = 0.004$: best restoration obtained at u_3.*

(a) u_1: 1st step, λ=0.006

(b) u_2: 2nd step

(c) u_3: 3rd step

(d) $f - u_1$+128

(e) $f - u_2$+128

(f) $f - u_3$+128

(g) $\|f - u_k\|_{L^2}$ vs. k

(h) $\|u_k - \text{original}\|_{L^2}$ vs. k
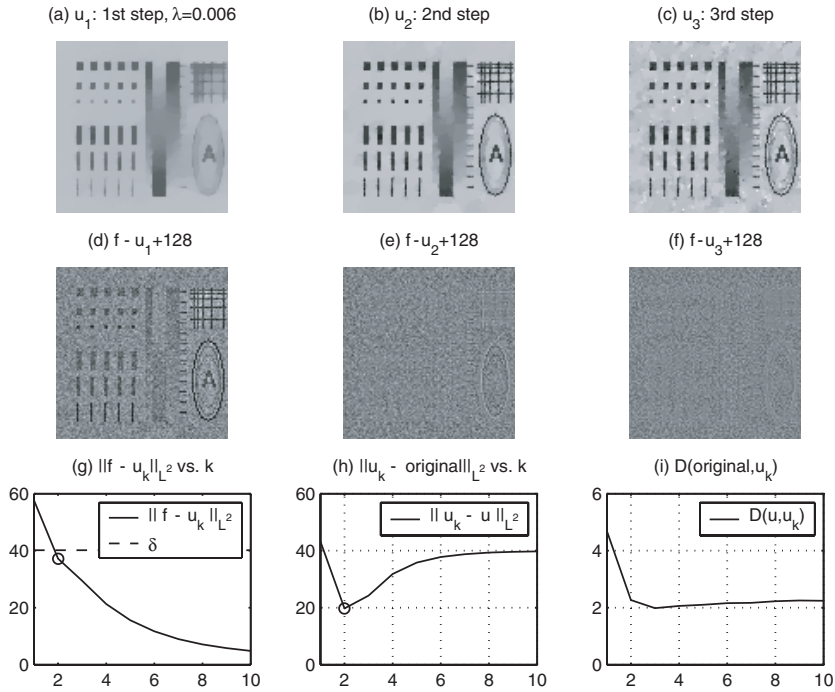
(i) $D(\text{original}, u_k)$

FIG. 7. *ROF with iterative regularization with $\delta = 40$ and $\lambda = 0.006$: best restoration obtained at u_2.*
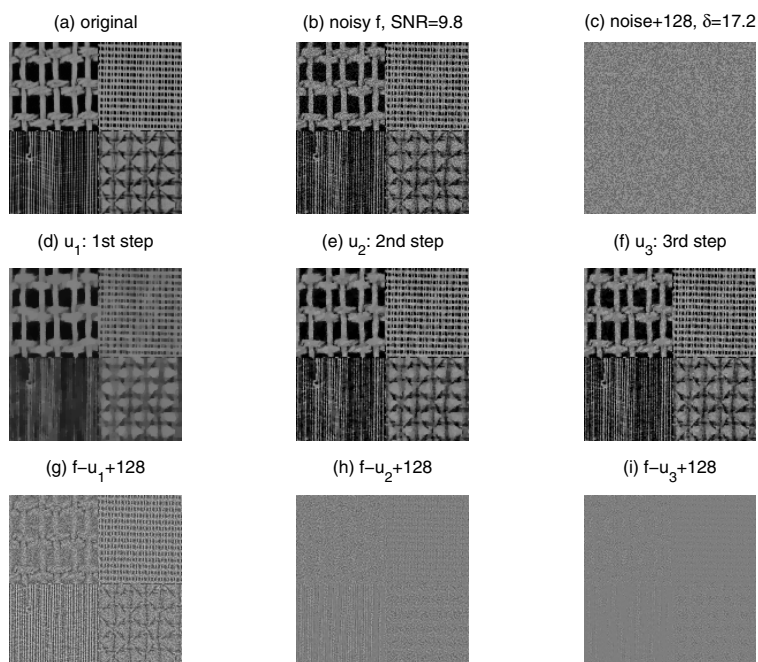
Fɪɢ. 8. *ROF with iterative regularization with $\delta = 17$ and $\lambda = 0.02$, textured image, and best restoration obtained at u_2.*
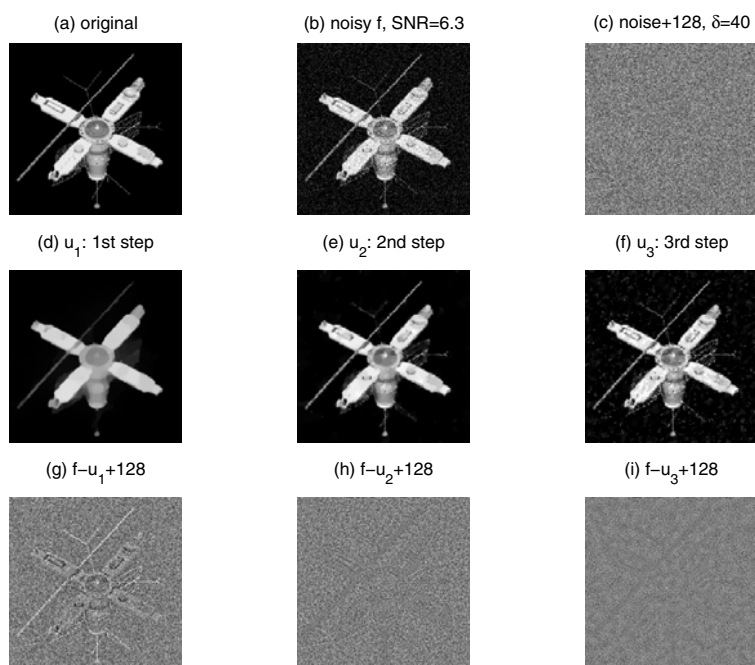

Fɪɢ. 9. *ROF with iterative regularization with $\delta = 40$ and $\lambda = 0.0055$, satellite image, and best restoration obtained at u_2.*
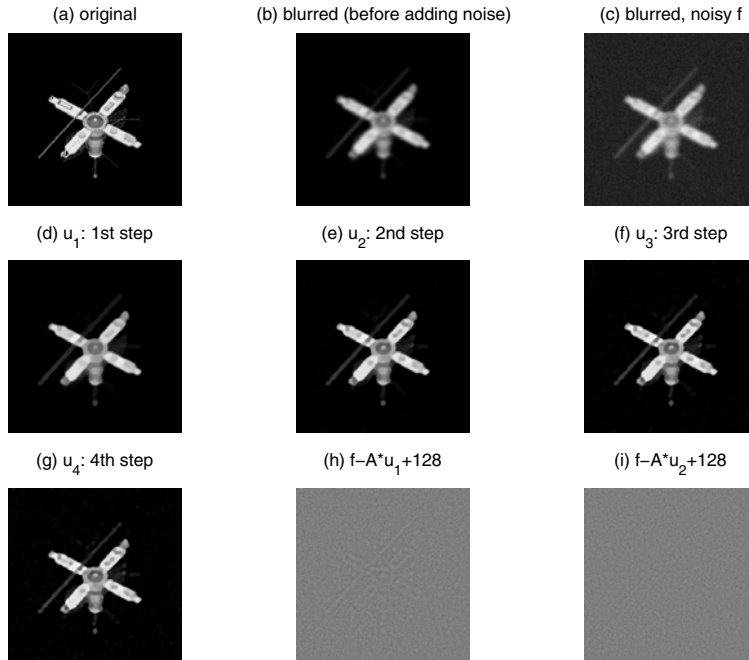
Fig. 10. *ROF with iterative regularization, deblur + denoise, Gaussian blur/noise, $\delta = 10$, and $\lambda = 0.1$.*

REFERENCES

[1] R. Acar and C. R. Vogel, *Analysis of total variation penalty methods*, Inverse Problems, 10 (1994), pp. 1217–1229.

[2] J.-F. Aujol, G. Aubert, L. Blanc-Feraud, and A. Chambolle, *Decomposing an image: Application to textured images and SAR images*, J. Math. Imaging Vision, to appear.

[3] J.-F. Aujol and A. Chambolle, *Dual norms and image decomposition models*, Internat. J. Comput. Vision, 63 (2005), pp. 85–104.

[4] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, *Image inpainting*, in Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, ACM, New York, 2000, pp. 417–424.

[5] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, *Simultaneous structure and texture inpainting*, IEEE Trans. Image Process., 12 (2003), pp. 882–889.

[6] L. Bregman, *The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming*, U.S.S.R. Comput. Math. and Math. Phys., 7 (1967), pp. 200–217.

[7] M. Burger and S. Osher, *Convergence rates of convex variational regularization*, Inverse Problems, 20 (2004), pp. 1411–1422.

[8] E. Casas, K. Kunisch, and C. Pola, *Regularization by functions of bounded variation and applications to image enhancement*, Appl. Math. Optim., 40 (1999), pp. 229–257.

[9] E. Cetin, *Reconstruction of signals from Fourier transform samples*, Signal Process., 16 (1989), pp. 129–148.

[10] E. Cetin, *An iterative algorithm for signal reconstruction from bispectrum*, IEEE Trans. Signal Process., 39 (1991), pp. 2621–2628.

[11] A. Chambolle, *An algorithm for total variation regularization and denoising*, J. Math. Imaging Vision, 20 (2004), pp. 89–97.

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

[13] T. F. Chan and S. Esedoglu, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math., to appear.

[14] T. F. CHAN, G. H. GOLUB, AND P. MULET, *A nonlinear primal-dual method for total variation-based image restoration*, SIAM J. Sci. Comput., 20 (1999), pp. 1964–1977.

[15] T. CHAN, A. MARQUINA, AND P. MULET, *High-order total variation-based image restoration*, SIAM J. Sci. Comput., 22 (2000), pp. 503–516.

[16] T. F. CHAN AND C. K. WONG, *Total variation blind deconvolution*, IEEE Trans. Image Process., 7 (1998), pp. 370–375.

[17] G. CHAVENT AND K. KUNISCH, *Regularization of linear least squares problems by total bounded variation*, ESAIM Control Optim. Calc. Var., 2 (1997), pp. 359–376.

[18] G. CHEN AND M. TEBOULLE, *Convergence analysis of a proximal-like minimization algorithm using Bregman functions*, SIAM J. Optim., 3 (1993), pp. 538–543.

[19] D. DOBSON AND O. SCHERZER, *Analysis of regularized total variation penalty methods for denoising*, Inverse Problems, 12 (1996), pp. 601–617.

[20] I. EKELAND AND R. TEMAM, *Convex Analysis and Variational Problems*, Classics Appl. Math. 28, SIAM, Philadelphia, 1999.

[21] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[22] S. ESEDOGLU AND S. OSHER, *Decomposition of images by the anisotropic Rudin-Osher-Fatemi model*, Comm. Pure Appl. Math., 57 (2004), pp. 1609–1626.

[23] D. GOLDFARB AND W. YIN, *Second-order cone programming methods for total variation-based image restoration*, SIAM J. Sci. Comput., to appear.

[24] C. GROETSCH AND O. SCHERZER, *Nonstationary iterated Tikhonov-Morozov method and third order differential equations for the evaluation of unbounded operators*, Math. Methods Appl. Sci., 23 (2000), pp. 1287–1300.

[25] M. HANKE, *A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems*, Inverse Problems, 13 (1997), pp. 79–95.

[26] M. HANKE AND C. W. GROETSCH, *Nonstationary iterated Tikhonov regularization*, J. Optim. Theory Appl., 98 (1998), pp. 37–53.

[27] W. HINTERBERGER AND O. SCHERZER, *Variational Methods on the Space of Functions of Bounded Hessian for Convexification and Denoising*, preprint, University of Innsbruck, Innsbruck, Austria, 2003.

[28] K. C. KIWIEL, *Proximal minimization methods with generalized Bregman functions*, SIAM J. Control Optim., 35 (1997), pp. 1142–1168.

[29] M. LYSAKER, S. OSHER, AND X.-C. TAI, *Noise removal using smoothed normals and surface fitting*, IEEE Trans. Image Process., 13 (2004), pp. 1345–1357.

[30] Y. MEYER, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, AMS, Providence, RI, 2001.

[31] M. NIKOLOVA, *Local strong homogeneity of a regularized estimator*, SIAM J. Appl. Math., 61 (2000), pp. 633–658.

[32] M. NIKOLOVA, *Estimées localement fortement homogènes*, C. R. Acad. Sci. Paris Sér. I Math., 325 (1997), pp. 665–670.

[33] M. NIKOLOVA, *Weakly constrained minimization. Application to the estimation of images and signals involving constant regions*, J. Math. Imaging Vision, 21 (2004), pp. 155–175.

[34] 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.

[35] S. OSHER, A. SOLÉ, AND L. VESE, *Image decomposition and restoration using total variation minimization and the H^{-1} norm*, Multiscale Model. Simul., 1 (2003), pp. 349–370.

[36] W. RING, *Structural properties of solutions of total variation regularization problems*, M2AN Math. Model. Numer. Anal., 34 (2000), pp. 799–810.

[37] L. RUDIN, P. L. LIONS, AND S. OSHER, *Multiplicative denoising and deblurring: Theory and algorithms*, in Geometric Level Set Methods in Imaging Vision and Graphics, S. Osher and N. Paragios, eds., Springer, New York, 2003, pp. 103–121.

[38] L. I. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.

[39] L. RUDIN AND S. OSHER, *Total variation based image restoration with free local constraints*, in Proceedings of the IEEE International Conference on Image Processing, Austin, TX, 1994, pp. 31–35.

[40] O. SCHERZER AND C. GROETSCH, *Inverse scale space theory for inverse problems*, in Scale-Space and Morphology in Computer Vision, Lecture Notes in Comput. Sci. 2106, M. Kerckhove, ed., Springer, New York, 2001, pp. 317–325.

[41] J. L. STARCK, M. ELAD, AND D. DONAHO, *Image decomposition via the combination of sparse representations and a variational approach*, IEEE Trans. Image Process., to appear.

[42] D. STRONG AND T. CHAN, *Edge-preserving and scale-dependent properties of total variation*

regularization, Inverse Problems, 19 (2003), pp. S165–S187.

[43] E. Tadmor, S. Nezzar, and L. Vese, *A multiscale image representation using hierarchical* (BV, L^2) *decompositions*, Multiscale Model. Simul., 2 (2004), pp. 554–579.

[44] T. Tasdizen, R. Whitaker, P. Burchard, and S. Osher, *Geometric processing via normal maps*, ACM Trans. Graphics, 22 (2003), pp. 1012–1033.

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

[46] L. A. Vese and S. J. Osher, *Numerical methods for p-harmonic flows and applications to image processing*, SIAM J. Numer. Anal., 40 (2002), pp. 2085–2104.

[47] C. R. Vogel, *Computational Methods for Inverse Problems*, Frontiers Appl. Math. 23, SIAM, Philadelphia, 2002.