

Error Estimation for Bregman Iterations and Inverse Scale Space Methods in Image Restoration

Martin Burger*, Elena Resmerita[†] and Lin He[†]

Abstract

In this paper we consider error estimation for image restoration problems based on generalized Bregman distances. This error estimation technique has been used to derive convergence rates of variational regularization schemes for linear and nonlinear inverse problems by the authors before (cf. [4, 21, 22]), but so far it was not applied to image restoration in a systematic way. Due to the flexibility of the Bregman distances, this approach is particularly attractive for imaging tasks, where often singular energies (nondifferentiable, not strictly convex) are used to achieve certain tasks such as preservation of edges.

Besides the discussion of the variational image restoration schemes, our main goal in this paper is to extend the error estimation approach to iterative regularization schemes (and time-continuous flows) that have emerged recently as multiscale restoration techniques and could improve some shortcomings of the variational schemes. We derive error estimates between the iterates and the exact image both in the case of clean and noisy data, the latter also giving indications on the choice of termination criteria.

The error estimates are applied to various image restoration approaches such as denoising and decomposition by total variation and wavelet methods. We shall see that interesting results for various restoration approaches can be deduced from our general results by just exploring the structure of subgradients.

Key words: Image restoration, error estimation, iterative regularization, Bregman distance, total variation, wavelets.

AMS Subject classification: Primary 47A52; Secondary 49M30, 94A08.

1 Introduction

Image restoration schemes based on penalization with strongly nonlinear and possibly singular energies received growing attention in the recent years due to various superior properties compared to classical schemes based on quadratic energies. Examples are *total variation methods* (cf., e.g. [9, 17]) originating from the work of Rudin, Osher, and Fatemi (cf. [23]) and wavelet shrinkage techniques as proposed by Donoho et al. (cf. [11]). In both cases the penalizing energies are nondifferentiable and positively one-homogeneous, which leads to various singular properties and creates enormous challenges in the convergence analysis of such schemes.

In the case of variational schemes, error estimates have recently been derived using the so-called *generalized Bregman distance* instead of norms for measuring distances (cf. [4, 21, 22]). Such Bregman distances are not only well suited for deriving error estimates, they can also be interpreted in rather natural ways, for instance with respect to convergence of edges in total variation methods.

Bregman distances have also been used to construct novel iterative or evolution schemes for the imaging tasks mentioned above (cf. [18]) and also generalized to nonlinear inverse problems (cf. [2]). From an optimization point of view, these algorithms can be interpreted as proximal point methods in rather general Banach spaces (allowing even non-reflexive ones such as ℓ^1 in

*Westfälische Wilhelms-Universität Münster, Institut für Numerische und Angewandte Mathematik, Einsteinstr. 62, D 48149 Münster, Germany (martin.burger@uni-muenster.de)

[†]Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Altenbergerstraße 69, 4040 Linz, Austria (elena.resmerita@ricam.oeaw.ac.at)

the wavelet case or the space of functions of bounded variation). It turned out that these novel schemes can remedy to great extent systematic errors of the variational schemes such as loss of contrast in the ROF-model. In the case of wavelet shrinkage, the application of the Bregman iteration to soft thresholding turned out to produce firm thresholding [13], while the inverse scale space methods yield hard thresholding as a natural extension.

It has been shown [18] that the iterative schemes in combination with a discrepancy type stopping rule (i.e., terminating in dependence of the noise variance) do regularize the ill-posed problem of recovering the true image from a noisy one. Despite their close connection to Bregman distances, error estimates for these schemes have not been given so far, and this is the main goal of this paper. Actually, measuring the errors with respect to Bregman distances comes naturally into play, as explained at the end of Section 4. To the best of our knowledge, these are also the first convergence rates in a non-hilbertian Banach space for a proximal method involving ill-posed operators. Since we are not only interested in the pure error estimation, we shall also derive some interpretations of the error estimates later.

2 Variational, Iterative, and Inverse Scale Space Methods for Singular Energies

In this section we shall introduce the basic notations and schemes under consideration, and recall some preliminary convergence results of the methods under discussion, as studied in [18, 5, 6].

First of all, let X be a Banach space embedded continuously in a Hilbert space Y (we think, e.g., of $X = BV(\Omega)$ and $Y = L^2(\Omega)$ for dimension less or equal two) such that X is a dual of some other Banach space (i.e., there exists a weak-* topology on X). Then we denote by

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

where $A : X \rightarrow Z$ is a bounded (typically compact) linear operator, which can be extended to a bounded linear operator on Y . Here we assume Z to be a Hilbert space for simplicity. In the case of image denoising with Gaussian noise, A is the identity on $L^2(\Omega)$, while for deblurring problems A would correspond to an integral operator. By $J : X \rightarrow \mathbb{R}$ we denote a regularization functional (e.g., the seminorm on the space $BV(\Omega)$) with domain $\mathcal{D}(J)$ and by $\tilde{u} \in X$ a solution of the equation

$$Au = f. \tag{1}$$

The existence of such a solution is actually a limiting condition on the data, which we call *exact data*, that is, f is generated by the real image \tilde{u} and not yet perturbed by noise. We shall later consider the (practically more relevant) case of noisy data as well. Then we think of the noisy output image indexed by the variance of the noise, i.e., we call the noisy data f^δ and assume

$$\|f - f^\delta\| \leq \delta. \tag{2}$$

The usual variational regularization scheme would consist in the solution of the problem

$$H(u, f^\delta) + J(u) \rightarrow \min_{u \in \mathcal{D}(J)}. \tag{3}$$

The minimizer is characterized by the first-order optimality condition

$$\lambda A^*(Au - f^\delta) + p = 0, \quad p \in \partial J(u),$$

where A^* is the Banach space adjoint operator of A and ∂J is the subgradient

$$\partial J(u) = \{p \in X^* \mid \langle v - u, p \rangle \leq J(v) - J(u), \text{ for any } v \in X\}. \tag{4}$$

A frequently used example of this setup are the ROF-model (cf. [23]) for image denoising

$$\frac{\lambda}{2} \int_{\Omega} (u - f^\delta)^2 + |u|_{BV} \rightarrow \min_{u \in BV(\Omega)} \tag{5}$$

with the BV-seminorm defined as

$$|u|_{BV} = \sup_{g \in C_0^\infty(\Omega), \|g\|_\infty \leq 1} \int_{\Omega} u \nabla \cdot g \, dx.$$

In this setup, $X = BV(\Omega)$ (embedded into $Y = L^2(\Omega)$ for dimension less or equal two), $Z = L^2(\Omega)$ and A is simply the embedding operator. This approach can easily be generalized to other imaging tasks such as deblurring by simply adjusting the fitting term H as above.

Another important example is the Wavelet shrinkage, where J is the norm in an appropriate Besov space, usually $B_{1,1}$. For the denoising case, i.e. A being the identity, an isometry between $L^2(\Omega)$ and $\ell^2(\mathbb{Z}^2)$ (the space of wavelet coefficients) can be used to write the resulting minimization problem in the form

$$\sum_i \sum_j \left(\frac{\lambda}{2} (u_{ij} - f_{ij}^\delta)^2 + w_{ij} |u_{ij}| \right) \rightarrow \min_{u_{ij} \in \ell^2(\mathbb{Z}^2) \cap \ell^1(\mathbb{Z}^2)}. \quad (6)$$

Here u_{ij} denote the wavelet coefficients of u , f_{ij}^δ the wavelet coefficients of f^δ , and w_{ij} are appropriate weight functions. Since the variational problems decompose in this case, one can separately minimize $e_{ij}(u) = \frac{\lambda}{2} (u - f_{ij}^\delta)^2 + w_{ij} |u|$ over $u \in \mathbb{R}$. The analytical solution of these problems can be computed [8, 27] as the *soft-shrinkage*

$$\hat{u} = S_\lambda(f_{ij}^\delta) = \text{sign}(f_{ij}^\delta) (|f_{ij}^\delta| - \lambda^{-1} w_{ij})_+, \quad (7)$$

where $a_+ = \max\{a, 0\}$ denotes the positive part of a real number a .

It is easy to see for $\delta = 0$ that the minimizer \hat{u} of (3) satisfies

$$J(\hat{u}) - J(\tilde{u}) \leq -H(\hat{u}, f) < 0,$$

that is, the functional J is smaller for the minimizer than for the real image. This systematic error causing an oversmoothing (also in the case of noisy data) can be a severe issue in imaging. One of the typical consequences is a loss of contrast, which has been analyzed in the case of total variation minimization by Meyer [17]. This loss of contrast can be interpreted as a deficiency of the scheme at larger scales, since usually in denoising one likes to eliminate only noise modeled as small scale features. In the wavelet soft shrinkage the loss is visible directly from the soft shrinkage operator, since the magnitude of every coefficient is decreased if the weights are positive. Since the coefficients are directly related to scales in wavelet techniques, the large scale deficiency can be corrected partly by choosing different weights (small or zero) corresponding to indices of large scale coefficients, but a systematic error remains for small scale features.

In order to cure the systematic error for general variational techniques, the following iterative procedure has been proposed in [18]: Starting from $u_0 = 0$ and $p_0 = 0 \in \partial J(u_0)$, a sequence of images is constructed via

$$u_k \in \arg \min_u \{J(u) + H(u, f^\delta) - \langle u, p_{k-1} \rangle\}, \quad (8)$$

for $k = 1, 2, \dots$, where $\langle \cdot, \cdot \rangle$ denotes the standard duality product and p_{k-1} is a subgradient of J at u_{k-1} . Clearly this procedure can be generalized to arbitrary initial values u_0 as long as there exists a subgradient $p_0 \in \partial J(u_0) \cap \mathcal{R}(A^*)$. It has been shown in [18] that the above iterative procedure is well-defined.

There are various alternative viewpoints of the iterative procedure (8).

For instance, by adding constants, (8) can equivalently be rewritten as

$$u_k \in \arg \min_u \{H(u, f^\delta) + D_J^{p_{k-1}}(u, u_{k-1})\}, \quad (9)$$

with the *generalized Bregman distance*

$$D_J^p(v, u) = J(v) - J(u) - \langle v - u, p \rangle, \quad p \in \partial J(u). \quad (10)$$

Another equivalent version is

$$u_k \in \arg \min_u \{J(u) + H(u, f^\delta + v_{k-1})\}, \quad v_k = v_{k-1} - (Au_k - f^\delta) \quad (11)$$

where $v_0 = 0$. The equivalence of the last form is obtained with the setting $p_k = \lambda A^* v_k$ from the optimality condition of (8),

$$\lambda A^*(Au_k - f^\delta) + p_k - p_{k-1} = 0. \quad (12)$$

As an immediate consequence, the following equality is obtained:

$$p_k = \lambda \sum_{j=1}^k A^*(f^\delta - Au_j), \quad \text{respectively} \quad v_k = k f^\delta - \sum_{j=1}^k Au_j. \quad (13)$$

The smoothing effect in the iteration comes from an appropriate choice of the stopping index, as usual in iterative regularization of ill-posed problems (cf.). With decreasing noise variation $\delta \rightarrow 0$, the stopping index should increase to infinity.

In the limit $\lambda \downarrow 0$, the first-order optimality condition can be interpreted as a forward Euler discretization of the flow

$$\partial_t p = -A^*(Au - f^\delta), \quad p \in \partial J(u),$$

which has been termed *nonlinear inverse scale space method* (cf. [6, 5]) in analogy to previous work on inverse scale space by Scherzer and Groetsch [24].

The terminology inverse scale space method is due to the fact that this approach somehow behaves in an inverse way to the popular scale space methods. Scale space methods are only applicable for A being the identity on $Y = Z$ in the above setup, they consist of the gradient flow

$$\partial_t u(t) = -p(t) \in -\partial J(u(t)), \quad u(0) = f^\delta. \quad (14)$$

The smoothing effect happens again by appropriate stopping, which should be a small time in this case, since the long-time limit of the flow is just a minimizer of J . Below we shall see that a surprisingly analogous analysis as for the inverse scale space method is possible also for scale space methods.

We recall below several intermediate results as well as some of the main results shown in [18].

Proposition 2.1. *Under the above assumptions, the sequence $H(u_k, f^\delta)$ obtained from the iterates (8) is monotonically non-increasing, one even has*

$$H(u_k, f^\delta) \leq H(u_k, f^\delta) + D^{p_{k-1}}(u_k, u_{k-1}) \leq H(u_{k-1}, f^\delta). \quad (15)$$

Moreover, let u be such that $J(u) < \infty$, then one has

$$D^{p_k}(u, u_k) + D^{p_{k-1}}(u_k, u_{k-1}) + H(u_k, f^\delta) \leq H(u, f^\delta) + D^{p_{k-1}}(u, u_{k-1}). \quad (16)$$

Theorem 2.2 (Exact Data). *Assume that there exists a minimizer $\tilde{u} \in BV(\Omega)$ of $H(\cdot, f)$ such that $J(\tilde{u}) < \infty$. Then*

$$H(u_k, f) \leq H(\tilde{u}, f) + \frac{J(\tilde{u})}{k} \quad (17)$$

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 $Au = f$. If \tilde{u} is the unique solution of $Au = f$, then $u_k \rightarrow \tilde{u}$ in the weak*-topology in $BV(\Omega)$.

If f^δ is given such that (2) holds, then the optimality condition for (8) reads as follows¹:

$$q_k + p_k - p_{k-1} = 0, \quad q_k = \lambda A^*(Au_k - f^\delta). \quad (18)$$

¹To make notations short, we will also denote by u_k and p_k the iterates and their corresponding subgradients in the noisy data case.

Therefore,

$$p_k = \lambda \sum_{j=1}^k A^*(f^\delta - Au_j). \quad (19)$$

Other useful relationships were obtained, such as²:

$$H(u_k, f^\delta) \leq H(u_{k-1}, f^\delta), \quad \forall k \in \mathbb{N}^*. \quad (20)$$

$$D_J^{p_k}(\tilde{u}, u_k) + D_J^{p_{k-1}}(u_k, u_{k-1}) + H(u_k, f^\delta) \leq \delta^2 + D_J^{p_{k-1}}(\tilde{u}, u_{k-1}), \quad \forall k \in \mathbb{N}^*. \quad (21)$$

$$D_J^{p_k}(\tilde{u}, u_k) + \sum_{j=1}^k [D_J^{p_{j-1}}(u_j, u_{j-1}) + H(u_j, f^\delta)] \leq k\delta^2 + J(\tilde{u}). \quad (22)$$

3 Error Estimation for Variational Schemes

In the following we discuss the basic ideas needed for the error estimation with Bregman distances and recall some results previously derived for variational schemes as (3). At this instance we shall also highlight some properties and techniques that will be of importance for the novel error estimates below.

Obviously the regularization functional J is added to the energy to obtain a smoothing of the image with respect to a certain criterion, which can be understood from the functional J and its variations. In the case of a classical quadratic regularization term like $J(u) = \frac{1}{2} \int_{\Omega} |\nabla u|^2 dx$, one can employ the L^2 -adjoint $A^* : Z \rightarrow L^2(\Omega)$ of the operator A to obtain

$$-\Delta u = \lambda A^*(f^\delta - Au).$$

The smoothing therefore appears in two steps: First of all, the adjoint operator creates a smoothing depending on the restoration task under consideration, e.g., a convolution in the case of deblurring model. In the case of denoising ($A^* = A$ being the identity), this first smoothing step does not appear. As a second step, smoothing is obtained by inverting the Laplacian, which comes from the regularization functional only. For the general nonlinear variational scheme (3), the two-step smoothing has to be considered in a different way. The optimality condition then becomes

$$p = \lambda A^*(f^\delta - Au), \quad p \in \partial J(u).$$

This formulation highlights the nature of the first smoothing step - it is actually a dual one smoothing the subgradient p . The second step depends on the relationship between the primal variable u and the dual variable p . If J is strictly convex, then each $p \in X^*$ is the subgradient of at most one element $u \in X$. The situation becomes particularly interesting for singular energies like ℓ^1 -norm, L^1 -norm, or BV -seminorm, where there is no unique reconstruction of the primal from the dual variable. This can lead to a significantly different smoothing behaviour in the second step. For example, in the total variation case the dual variable is solely linked to properties of the level sets (mean curvature, see below), but not to the grey or color values on the level sets. This means that the smoothing creates rounded level sets, but does not prohibit discontinuous grey scales (as appearing in edges).

The question which functions, respectively subgradients can actually fulfill a dual smoothing property naturally leads to the so-called *source condition*

(SC) *There exists $\xi \in \partial J(\tilde{u})$ such that $\xi = A^*q$ for a source element $q \in Z$.*

It can be shown (see [4]) that the set of minimizers of (3) for arbitrary $f^\delta \in Z$ and $\lambda \in \mathbb{R}$ exactly coincides with the set of elements satisfying a source condition. In fact, (SC) can be interpreted as an abstract smoothness condition. In the classical case of deblurring, with A being an integral operator on $L^2(\Omega)$ and J the square of the norm in this space, the source condition implies that u

²By \mathbb{N}^* we mean the set of positive integers which are greater or equal 1.

is in the range of the integral operator A^* and thus smooth, with detailed smoothness properties obtained from the properties of the integral kernel. In the case of a more general regularization functional J , (SC) rather becomes a smoothness condition on the dual variable (subgradient), which can make a significant difference, as we shall see in the examples below.

The source condition has a direct implication for the Bregman distance as defined in (10). With the particular subgradient ξ we obtain

$$D_J^\xi(u, \tilde{u}) = J(u) - J(\tilde{u}) - \langle u - \tilde{u}, \xi \rangle = J(u) - J(\tilde{u}) - \langle q, Au - A\tilde{u} \rangle.$$

Thus, the Bregman distance can be related to the error in the regularization functionals ($J(u) - J(\tilde{u})$) and the output error ($Au - A\tilde{u}$). Since the output error can be controlled easily, it is not surprising that error estimates in the Bregman distance can be successful for general nonlinear variational schemes.

Before sketching the further analysis, we mention some further properties of Bregman distance. First of all, it is clear that D_J^p depends on the specific subgradient (if the subdifferential is not a singleton). This can be a complication on the one hand, but on the other hand it can be useful to derive fine estimates as we shall see below. If a source condition is satisfied by multiple subgradients, then one can aim at deriving an error estimate for each subgradient, which is clearly an advantage. One also observes that the Bregman distance is not symmetric in general, which can be remedied partly by using the *symmetric Bregman distance*

$$D_J^{symm}(u_1, u_2) = \langle u_1 - u_2, p_1 - p_2 \rangle = D_J^{p_1}(u_2, u_1) + D_J^{p_2}(u_1, u_2), \quad p_i \in \partial J(u_i).$$

Of course, the symmetric Bregman distance depends also on the specific selection of subgradients p_i , which we however suppress in the notation for simplicity.

It is also worth mentioning that, for positively one-homogeneous functionals J (as the total variation seminorm or the ℓ^1 -norm), the identity $J(u) = \langle p, u \rangle$ holds for all $p \in \partial J(u)$. The Bregman distance associated to such a functional J therefore simplifies to $D_J^p(v, u) = J(v) - \langle p, v \rangle$.

Now assume that a solution \tilde{u} of $Au = f$ satisfies (SC). Then, for $\lambda \in \mathbb{R}$ and \hat{u} denoting the minimizer of (3), we obtain by adding ξ to the optimality condition

$$\hat{p} - \xi + \lambda A^*(A\hat{u} - f) = A^*(\lambda(f^\delta - f) - q), \quad \hat{p} \in \partial J(\hat{u}).$$

A duality product with $\hat{u} - \tilde{u}$ yields

$$D_J^{symm}(\hat{u}, \tilde{u}) + \lambda \langle A\hat{u} - f, A\hat{u} - f \rangle = \langle \lambda(f^\delta - f) - q, A\hat{u} - f \rangle.$$

An application of Young's inequality yields

$$D_J^{symm}(\hat{u}, \tilde{u}) + \frac{\lambda}{2} \|A\hat{u} - f\|^2 \leq \frac{\lambda}{2} \|f^\delta - f\|^2 + \frac{1}{2\lambda} \|q\|^2.$$

Finally, inserting (2) we obtain a variant of the estimates in [4]:

Theorem 3.1. *Let $\tilde{u} \in X$ satisfy $A\tilde{u} = f$ and let $f^\delta \in Z$ be noisy data satisfying (2). Then the minimizer $\hat{u} \in X$ of (3) verifies*

$$D_J^{symm}(\hat{u}, \tilde{u}) + \frac{\lambda}{2} \|A\hat{u} - f\|^2 \leq \frac{\lambda\delta^2}{2} + \frac{\|q\|^2}{2\lambda}. \quad (23)$$

We mention that stability estimates for the variational scheme can also be derived with an analogous technique. For u_i being the minimizer of (3) with right-hand side f_i , the estimate

$$D_J^{symm}(u_1, u_2) + \frac{\lambda}{2} \|Au_1 - Au_2\|^2 \leq \frac{\lambda}{2} \|f_1 - f_2\|^2$$

holds. Improved estimates can be derived under a stronger source condition of the form $\xi \in \mathcal{R}(A^*A)$ (cf. [21]), and the analysis can be generalized to certain nonlinear operators (cf. [22]).

4 Error Estimates for the Bregman Iteration

In this section we derive error estimates between a solution of the equation $Au = f$ and the iterates u_k produced by method (8). Although (semi)convergence properties for the sequence $\{u_k\}$ with respect to the weak*-topology are available (see Theorem 2.2), it is not clear how to quantify such a convergence in case that sourcewise representations of the solutions are assumed. We show below that the Bregman distances associated with the total variation functional can be employed in such a quantitative analysis. We start with the case of exact data.

Theorem 4.1 (Exact data). *Let $\tilde{u} \in X$ be a solution of $Au = f$ and assume that the source condition (SC) is satisfied. Then*

$$D_J^{p_k}(\tilde{u}, u_k) = O(1/k), \quad (24)$$

where u_k are the iterates of procedure (8). More precisely, the following error estimate holds:

$$D_J^{p_k}(\tilde{u}, u_k) \leq \frac{\|q\|^2}{2\lambda k}, \quad \forall k \in \mathbb{N}^*. \quad (25)$$

Proof. Let $x_k = \lambda \sum_{j=1}^k (f - Au_j)$. Fix $k \in \mathbb{N}^*$. From (13) and (SC), we get for any $j \in \mathbb{N}$, $0 < j \leq k$,

$$\begin{aligned} \lambda D_J^{p_j}(\tilde{u}, u_j) + \lambda D_J^\xi(u_j, \tilde{u}) &= \lambda \langle p_j - \xi, u_j - \tilde{u} \rangle = \langle A^*x_j - A^*q, \lambda(u_j - \tilde{u}) \rangle \\ &= \langle x_j - q, \lambda(Au_j - f) \rangle = \langle x_j - q, x_{j-1} - x_j \rangle \\ &= \frac{1}{2} \|x_{j-1} - q\|^2 - \frac{1}{2} \|x_j - q\|^2 - \frac{\lambda^2}{2} \|Au_j - f\|^2 \\ &\leq \frac{1}{2} \|x_{j-1} - q\|^2 - \frac{1}{2} \|x_j - q\|^2. \end{aligned}$$

By summing up the last inequality for $j = 1, k$, using the nonnegativity of $D_J^\xi(u_j, \tilde{u})$ and the fact that $D_J^{p_j}(\tilde{u}, u_j)$ is nonincreasing with respect to j (see Proposition 2.1), it follows that

$$D_J^{p_k}(\tilde{u}, u_k) \leq \frac{1}{k} \sum_{j=1}^k D_J^{p_j}(\tilde{u}, u_j) \leq \frac{1}{2\lambda k} \|q\|^2. \quad \square$$

Remark. Condition (SC) guarantees even convergence of the iterates to the solution in the sense of the Bregman distance associated with the functional J (see (25)).

Suppose that we are given noisy data f^δ that satisfy (2). The next result shows that the a priori stopping rule $k_*(\delta) \sim 1/\delta$ yields semiconvergence of the regularization method.

Proposition. *Let $\tilde{u} \in X$ verify $A\tilde{u} = f$, and assume that the source condition (SC) and inequality (2) are satisfied. Moreover, let the stopping index $k_*(\delta)$ be chosen of order $1/\delta$. Then, $\{J(u_{k_*(\delta)})\}_\delta$ is bounded and hence, as $\delta \rightarrow 0$, there exists a weak*-convergent subsequence $\{u_{k(\delta_n)}\}_n$ in $BV(\Omega)$ whose limit is a solution of $Au = f$. Moreover, if the solution of the equation is unique, then $u_{k_*(\delta)}$ converges in the weak*-topology to the solution as $\delta \rightarrow 0$.*

Proof. The proof follows the pattern of the proof for Theorem 2.2 for the exact data case, but it is provided here for the sake of completeness. From inequality (22) we have

$$\begin{aligned} J(\tilde{u}) + k\delta^2 &\geq \sum_{j=1}^k D_J^{p_{j-1}}(u_j, u_{j-1}) = J(u_k) - J(u_0) - \sum_{j=1}^k \langle p_{j-1}, u_j - u_{j-1} \rangle \\ &= J(u_k) - \sum_{j=1}^k \langle p_{j-1}, u_j - \tilde{u} \rangle - \sum_{j=1}^{k-1} \langle p_j, u_j - \tilde{u} \rangle \\ &= J(u_k) - \langle p_{k-1}, u_k - \tilde{u} \rangle + \sum_{j=1}^{k-1} \langle p_j - p_{j-1}, u_j - \tilde{u} \rangle \end{aligned}$$

and using (18):

$$\begin{aligned}
J(\tilde{u}) + k\delta^2 &= J(u_k) + \sum_{j=1}^{k-1} \langle q_j, u_k - \tilde{u} \rangle - \sum_{j=1}^{k-1} \langle q_j, u_j - \tilde{u} \rangle \\
&= J(u_k) + \lambda \sum_{j=1}^{k-1} \langle Au_j - f^\delta, Au_k - f \rangle - \lambda \sum_{j=1}^{k-1} \langle Au_j - f^\delta, Au_j - f \rangle.
\end{aligned}$$

Next we use Cauchy-Schwarz inequality, (22) and the inequality $ab \leq \frac{a^2}{2} + \frac{b^2}{2}$ to get

$$\begin{aligned}
J(\tilde{u}) + k\delta^2 &\geq J(u_k) + \lambda \sum_{j=1}^{k-1} \langle Au_j - f^\delta, Au_k - f \rangle - \lambda \sum_{j=1}^{k-1} \langle Au_j - f^\delta, Au_j - f \rangle \\
&= J(u_k) + \lambda \sum_{j=1}^{k-1} \langle Au_j - f^\delta, Au_k - f^\delta \rangle - \lambda \sum_{j=1}^{k-1} \langle Au_j - f^\delta, Au_j - f^\delta \rangle \\
&\geq J(u_k) - \frac{3k\lambda}{2} \lambda \sum_{j=1}^{k-1} \|Au_j - f^\delta\|^2 - \frac{k\lambda}{2} \|Au_k - f^\delta\|^2 \\
&\geq J(u_k) - 4J(\tilde{u}) - 4\lambda k\delta^2.
\end{aligned}$$

Consequently, we obtain the following estimate

$$J(u_k) \leq 5J(\tilde{u}) + 5\lambda k\delta^2, \quad (26)$$

which further implies that the sequence $\{J(u_{k_*(\delta)})\}_\delta$ is bounded for $\delta > 0$ sufficiently small and for $k_*(\delta) \sim 1/\delta$. This together with boundedness of the sequence $(Au_{k_*(\delta)})_\delta$ (which is guaranteed by (20)) further yields

$$\|u_{k_*(\delta)}\|_{BV} = \|Au_{k_*(\delta)}\| + J(u_{k_*(\delta)}) \leq M,$$

for any $\delta > 0$ and for some positive constant M . On one hand, since $BV(\Omega)$ is provided with a weak* - topology (see, e.g., [17, p. 24]), we conclude that there is a subsequence $\{u_{k_*(\delta_n)}\}_n$ which converges to a point \tilde{u} with respect to that topology. Due to the embedding of $(BV(\Omega), w^*)$ into $(L^2(\Omega), \|\cdot\|_2)$ for spatial dimension less or equal two (cf. [2, Thm 3.3]), this subsequence converges to \tilde{u} in the $L^2(\Omega)$ - norm. Now continuity of the operator A on $L^2(\Omega)$ applies and gives $\lim_{n \rightarrow \infty} Au_{k_*(\delta_n)} = A\tilde{u}$. On the other hand, we derive from (22) the inequality

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

By our special choice for $k_*(\delta_n)$ and by (2), we obtain $\lim_{n \rightarrow \infty} Au_{k_*(\delta_n)} = f$ and then, $A\tilde{u} = f$. \square

The error estimates for the noisy data case are established below:

Theorem 4.2 (Noisy data). *Let $\tilde{u} \in X$ verify $A\tilde{u} = f$, and assume that the source condition (SC) and inequality (2) are satisfied. Then, the following estimate holds:*

$$D_J^{p_k}(\tilde{u}, u_k) \leq \frac{k\delta^2}{2} + \frac{\|q\|^2 + 1 - (1 - \lambda\delta)^k}{2\lambda k(1 - \lambda\delta)^{k-1}}, \quad \forall k \in \mathbb{N}. \quad (27)$$

Moreover, if the a priori choice $k_*(\delta) \sim \frac{1}{\delta}$ is made, then the following convergence rate is obtained

$$D_J^{p_{k_*(\delta)}}(\tilde{u}, u_{k_*(\delta)}) = O(\delta). \quad (28)$$

Proof. Let $x_k = \lambda \sum_{j=1}^k (f^\delta - Au_j)$. From (13) and (SC), we get for any positive $j \in \mathbb{N}$, $0 < j \leq k$,

$$\begin{aligned} \lambda D_J^{p_j}(\tilde{u}, u_j) + \lambda D_J^\xi(u_j, \tilde{u}) &= \lambda \langle p_j - \xi, u_j - \tilde{u} \rangle = \langle x_j - q, \lambda(Au_j - f) \rangle \\ &= \langle x_j - q, \lambda(Au_j - f^\delta) \rangle + \lambda \langle x_j - q, f^\delta - f \rangle \\ &= \langle x_j - q, x_{j-1} - x_j \rangle + \lambda \langle x_j - q, f^\delta - f \rangle \\ &\leq \langle x_j - q, x_{j-1} - x_j \rangle + \lambda \delta \|x_j - q\| \\ &\leq \frac{1}{2} \|x_{j-1} - q\|^2 - \frac{1 - \lambda \delta}{2} \|x_j - q\|^2 + \frac{\lambda \delta}{2}, \end{aligned}$$

where the last two terms were obtained by using the inequality $ab \leq \frac{a}{2} + \frac{ab^2}{2}$ for appropriate $a > 0$ and $b > 0$. Consequently,

$$2\lambda D_J^{p_j}(\tilde{u}, u_j) + (1 - \lambda \delta) \|x_j - q\|^2 \leq \|x_{j-1} - q\|^2 + \lambda \delta.$$

Fix a positive $k \in \mathbb{N}$. By summing up the last inequalities from 1 to k , we get

$$2\lambda \sum_{j=1}^k (1 - \lambda \delta)^{j-1} D_J^{p_j}(\tilde{u}, u_j) + (1 - \lambda \delta)^k \|x_k - q\|^2 \leq \|q\|^2 + \lambda \delta \sum_{j=1}^k (1 - \lambda \delta)^{j-1}, \quad (29)$$

and thus,

$$\sum_{j=1}^k D_J^{p_j}(\tilde{u}, u_j) \leq \frac{\|q\|^2}{2\lambda(1 - \lambda \delta)^{k-1}} + \frac{1 - (1 - \lambda \delta)^k}{2\lambda(1 - \lambda \delta)^{k-1}}. \quad (30)$$

Note that monotonicity of the sequence $D_J^{p_j}(\tilde{u}, u_j)$ is not guaranteed, as in the noise free case. This is why we are going to employ, instead, a monotonicity-like inequality derived from (21). More precisely, let $k > 1$ be an integer number and fix an integer j with $0 < j \leq k$. From inequality (21), we have

$$D_J^{p_{j+1}}(\tilde{u}, u_{j+1}) - D_J^{p_j}(\tilde{u}, u_j) \leq \delta^2,$$

Summing up these inequalities up to k yields

$$D_J^{p_k}(\tilde{u}, u_k) - D_J^{p_j}(\tilde{u}, u_j) \leq (k - j)\delta^2,$$

Summing up again with respect to j implies

$$k D_J^{p_k}(f\tilde{u}, u_k) - \sum_{j=1}^k D_J^{p_j}(\tilde{u}, u_j) \leq \left(k^2 - \sum_{j=1}^k j \right) \delta^2, \quad (31)$$

and then

$$k D_J^{p_k}(\tilde{u}, u_k) \leq \sum_{j=1}^k D_J^{p_j}(\tilde{u}, u_j) + \frac{k^2 \delta^2}{2}. \quad (32)$$

By combining (30) and (32), we get

$$k D_J^{p_k}(\tilde{u}, u_k) \leq \frac{k^2 \delta^2}{2} + \frac{\|q\|^2}{2\lambda(1 - \lambda \delta)^{k-1}} + \frac{1 - (1 - \lambda \delta)^k}{2\lambda(1 - \lambda \delta)^{k-1}}.$$

Inequality (27) follows then immediately. In order to obtain the convergence rate (28), note that $\lim_{\delta \rightarrow 0} (1 - \lambda \delta)^{k_*(\delta)-1} = e^{-\lambda}$, when $k_*(\delta) \sim 1/\delta$. This implies that $\frac{1}{(1 - \lambda \delta)^{k_*(\delta)-1}}$ is bounded for $\delta > 0$ sufficiently small. \square

We observe that the error estimates are somehow more restrictive than in the case of variational methods if ∂J is multivalued. In the variational case we obtained an error estimate for each

subgradient of J at \tilde{u} that satisfied the source condition (SC), but now we only have an error estimate for a very particular subgradient of J at u_k .

As mentioned in the Introduction, establishing error estimates in terms of Bregman distances is more than a consequence of proof techniques. In fact, there is a strong relationship between the weak*-convergence in $BV(\Omega)$ and the convergence with respect to D_J of the sequence $\{u_k\}_k$. Indeed, let us consider that equation (1) has a unique solution \tilde{u} . If the source condition (SC) holds and $k_*(\delta)\delta^2 \rightarrow 0$ as $\delta \rightarrow 0$, then $u_{k_*(\delta)} \rightarrow \tilde{u}$ in the weak*-topology when $\delta \rightarrow 0$. This implies that $u_{k_*(\delta)} \rightarrow \tilde{u}$ in the $L^2(\Omega)$ -norm and, consequently, $Au_{k_*(\delta)} \rightarrow f$ as $\delta \rightarrow 0$. Moreover, since J is lower semicontinuous with respect to the weak*-topology on $BV(\Omega)$, we also get $J(\tilde{u}) \leq \liminf_{\delta \rightarrow 0} J(u_{k_*(\delta)})$. Note that, by (13) and the definition of x_k , we have

$$D_J^{p_k}(\tilde{u}, u_k) = J(\tilde{u}) - J(u_k) - \langle f - Au_k, x_k \rangle, \quad \forall k \in \mathbb{N}^*.$$

Since $\{x_{k_*(\delta)}\}_\delta$ is bounded for δ small enough (cf. (29)), we obtain that $\langle f - Au_{k_*(\delta)}, x_{k_*(\delta)} \rangle \rightarrow 0$ when $\delta \rightarrow 0$. Hence,

$$0 \leq \liminf_{\delta \rightarrow 0} D_J^{p_{k_*(\delta)}}(\tilde{u}, u_{k_*(\delta)}) \leq \limsup_{\delta \rightarrow 0} D_J^{p_{k_*(\delta)}}(\tilde{u}, u_{k_*(\delta)}) = J(\tilde{u}) - \liminf_{\delta \rightarrow 0} J(u_{k_*(\delta)}) \leq 0,$$

which yields $\lim_{\delta \rightarrow 0} D_J^{p_{k_*(\delta)}}(\tilde{u}, u_{k_*(\delta)}) = 0$. However, we could not show the other implication between the weak*-convergence and the convergence in the D_J sense. We expect that it is not necessarily true.

5 Error Estimates for Inverse Scale Space Flows

The error estimates for the inverse scale space method can be derived in a completely analogous way as the ones for Bregman iteration in the previous section. One can even use the limiting behaviour λ of the iteration, which yields convergence to the inverse scale space method as shown in [5], to directly carry over the error estimates. For completeness we nonetheless give a direct derivation of error estimates also in this case, it also motivates an analogous treatment of scale space methods in the next section.

First of all we notice that by defining the integrated residual

$$\partial_t r(t) = f^\delta - Au(t), \quad r(0) = 0,$$

we obtain $p(t) = A^*r(t)$. Now consider the difference $\|r(t) - q\|$, which is finite due to the source condition $\xi = A^*q$, $q \in Z$. Then

$$\begin{aligned} \frac{d}{dt} \left(\frac{1}{2} \|r(t) - q\|^2 \right) &= \langle \partial_t r(t), r(t) - q \rangle = \langle f^\delta - Au(t), r(t) - q \rangle \\ &= \langle f^\delta - f, r(t) - q \rangle + \langle A\tilde{u} - Au(t), r(t) - q \rangle \\ &\leq \delta \|r(t) - q\| + \langle A\tilde{u} - Au(t), r(t) - q \rangle \\ &= \delta \|r(t) - q\| - \langle u - \tilde{u}, A^*(r(t) - q) \rangle \\ &= \delta \|r(t) - q\| - \langle u - \tilde{u}, p(t) - \xi \rangle. \end{aligned}$$

Since

$$0 \leq D_J^p(\tilde{u}, u(t)) \leq \langle u - \tilde{u}, p(t) - \xi \rangle$$

we have in particular

$$\frac{d}{dt} \left(\frac{1}{2} \|r(t) - q\|^2 \right) \leq \delta \|r(t) - q\|, \quad \text{hence} \quad \|r(t) - q\| \leq \|q\| + \delta t.$$

Inserting back we obtain

$$\frac{d}{dt} \left(\frac{1}{2} \|r(t) - q\|^2 \right) + D_J^p(\tilde{u}, u(t)) \leq \delta \|q\| + \delta^2 t$$

Hence, after integrating the estimate from 0 to t we conclude

$$\int_0^t D_J^{p(s)}(\tilde{u}, u(s)) ds + \frac{1}{2}\|r(t) - q\|^2 \leq \frac{1}{2}\|q\|^2 + \|q\|\delta t + \frac{1}{2}\delta^2 t^2. \quad (33)$$

For $\delta = 0$, we can use the decrease of the Bregman distance in time (cf. [5, 6]), i.e. $D_J^p(\tilde{u}, u(s)) \geq D_J^p(\tilde{u}, u(t))$ for $s \leq t$, to obtain:

Theorem 5.1 (Exact Data). *Let $\tilde{u} \in X$ satisfy $A\tilde{u} = f$ and (SC). Moreover, let u be the unique solution of the inverse scale space flow*

$$\partial_t p(t) = -A^*(Au(t) - f), \quad p(t) \in \partial J(u(t)).$$

Then the convergence rate $D_J^{p(t)}(\tilde{u}, u(t)) = \mathcal{O}(t^{-1})$ holds, more precisely

$$D_J^{p(t)}(\tilde{u}, u(t)) \leq \frac{\|q\|^2}{2t}.$$

In the noisy case some further effort is necessary, since the Bregman distance is not necessarily decreasing in time. Using standard manipulations and Young's inequality we obtain

$$\begin{aligned} D_J^{p(t)}(\tilde{u}, u(t)) &\leq D_J^{p(s)}(\tilde{u}, u(s)) + \int_s^t \langle \partial_t p(\tau), u(\tau) - \tilde{u} \rangle d\tau \\ &= D_J^{p(s)}(\tilde{u}, u(s)) - \int_s^t \langle A^*(Au(\tau) - f^\delta), u(\tau) - \tilde{u} \rangle d\tau \\ &= D_J^{p(s)}(\tilde{u}, u(s)) - \int_s^t \langle Au(\tau) - f, Au(\tau) - f \rangle d\tau - \int_s^t \langle f - f^\delta, Au(\tau) - f \rangle d\tau \\ &\leq D_J^{p(s)}(\tilde{u}, u(s)) + \frac{\delta^2}{4}(t - s). \end{aligned}$$

Now we integrate this inequality in time to obtain

$$D_J^{p(t)}(\tilde{u}, u(t)) \leq \frac{1}{t} \int_0^t D_J^{p(s)}(\tilde{u}, u(s)) ds + \frac{\delta^2}{8}t.$$

Now we can easily combine this estimate with (33) to derive the following result.

Theorem 5.2 (Noisy Data). *Let $\tilde{u} \in X$ satisfy $A\tilde{u} = f$ and (SC), and let f^δ be noisy data satisfying (2). Moreover, let u be the unique solution of the inverse scale space flow*

$$\partial_t p(t) = -A^*(Au(t) - f^\delta), \quad p(t) \in \partial J(u(t)).$$

Then the error estimate

$$D_J^{p(t)}(\tilde{u}, u(t)) \leq \frac{1}{2t} (\|q\| + \delta t)^2 + \frac{\delta^2 t}{8}.$$

In particular, for a choice $t_*(\delta) = \mathcal{O}(\delta^{-1})$ we obtain $D_J^{p(t_*)}(\tilde{u}, u(t_*)) = \mathcal{O}(\delta)$.

6 Error Estimates for Scale Space Flows

We now consider the scale space approach (14). The source condition (SC) in this case with A being the identity is just $\xi = q \in Y \cap \partial J(\tilde{u})$.

One observes that somehow the role of p and u is interchanged in the scale space compared to the inverse scale space flow. Therefore we try to derive an error estimate by analogous reasoning as in the previous section, but with changed roles of p and u . Hence, we compute the time derivative

$$\begin{aligned} \frac{d}{dt} \left(\frac{1}{2} \|u(t) - f\|^2 \right) &= \langle \partial_t u(t), u(t) - f \rangle = -\langle p(t), u(t) - f \rangle \\ &= -\langle p(t) - q, u(t) - f \rangle - \langle q, u(t) - f \rangle. \end{aligned}$$

Now we obtain in particular

$$\frac{d}{dt} \left(\frac{1}{2} \|u(t) - f\|^2 \right) \leq \|q\| \|u(t) - f\|$$

and consequently

$$\|u(t) - f\| \leq \|u(0) - f\| + \|q\|t = \|f^\delta - f\| + \|q\|t \leq \delta + \|q\|t.$$

Reinserting we conclude

$$\begin{aligned} \frac{d}{dt} \left(\frac{1}{2} \|u(t) - f\|^2 \right) + D_J^q(u(t), \tilde{u}) &\leq \frac{d}{dt} \left(\frac{1}{2} \|u(t) - f\|^2 \right) + \langle p(t) - q, u(t) - f \rangle \\ &\leq \|q\|^2 t + \|q\|\delta, \end{aligned}$$

and after time integration we have

$$\frac{1}{2} \|u(t) - f\|^2 + \int_0^t D_J^q(u(s), \tilde{u}) ds \leq \frac{\|q\|^2 t^2}{2} + \|q\|\delta t + \frac{\delta^2}{2},$$

where we have inserted (2) to estimate $\|u(0) - f\|$.

For the Bregman distance $D_J^q(u(t), \tilde{u})$ we compute

$$\begin{aligned} D_J^q(u(t), \tilde{u}) &= D_J^q(u(s), \tilde{u}) + \int_s^t \langle \partial_t u(\tau), p(\tau) - q \rangle d\tau \\ &= D_J^q(u(s), \tilde{u}) - \int_s^t \|p(\tau) - q\|^2 d\tau + \int_s^t \langle q, q - p(\tau) \rangle d\tau \\ &\leq D_J^q(u(s), \tilde{u}) + \frac{\|q\|^2}{4} (t - s). \end{aligned}$$

Hence

$$D_J^q(u(t), \tilde{u}) \leq \frac{1}{t} \int_0^t D_J^q(u(s), \tilde{u}) ds + \frac{\|q\|^2 t}{8}.$$

Finally, in combination with the above estimate we conclude an error estimate for scale space flows:

$$D_J^q(u(t), \tilde{u}) \leq \frac{t}{2} \left(\|q\| + \frac{\delta}{t} \right)^2 + \frac{\|q\|^2 t}{8}. \quad (34)$$

In particular the choice $t_* \sim \delta$ for the stopping time yields a rate $D_J^q(u(t_*), \tilde{u}) = \mathcal{O}(\delta)$, which is analogous to the methods discussed above.

7 Total Variation Methods

In order to apply the error estimation to total variation methods we recall some interpretation of the source condition (SC) from [4] and discuss some properties of subgradients. We assume that A is a bounded operator on $L^2(\Omega)$. The functional under consideration is the total variation or bounded variation seminorm

$$J(u) = |u|_{BV} := \sup_{g \in C_0^1(\Omega), \|g\|_\infty \leq 1} \int_\Omega u \nabla \cdot g dx. \quad (35)$$

Formally we can write (SC) as the nonlinear partial differential equation $-\nabla \cdot \left(\frac{\nabla \tilde{u}}{|\nabla \tilde{u}|} \right) = A^* \xi$. On smooth parts of \tilde{u} , the right-hand side is exactly the mean curvature on level set of \tilde{u} . Hence, the source condition implies that the level sets of the image intensity are smooth. As we shall see in the particularly important case of piecewise constant image intensities, the source condition does not exclude the possibility of discontinuities (edges), but rather enforces smoothness of the discontinuity set.

We start by recalling a well-known characterization of the subdifferential of the total variation (see, e.g., [17, p. 38]):

Lemma 7.1. *The subdifferential of J is given by*

$$\partial J(u) = \{\nabla \cdot g \mid g \in L^\infty(\Omega), \|g\|_\infty = 1, J(u) = \langle u, \nabla \cdot g \rangle\},$$

where $\nabla \cdot g$ denotes the divergence of g in a distributional sense.

This result can be specialized to explicitly compute some subgradients for piecewise constant images with smooth discontinuity sets. Let us first recall that the *signed distance function* of a set $\Omega \subset X$ is defined as

$$b(x) = \begin{cases} -d(x, \partial\Omega), & \text{if } x \in \Omega; \\ d(x, \partial\Omega), & \text{if } x \notin \Omega, \end{cases}$$

where $d(x, \partial\Omega)$ is the (metric) distance from x to the boundary of the set Ω .

Lemma 7.2. *Let $u \in BV(\Omega)$ be a piecewise constant image of the form $u = \sum_{k=1}^M c_k \mathbf{1}_{\Omega_k}$ with $\Omega_k \subset \Omega$ having C^1 -boundary and $d(\partial\Omega_j, \partial\Omega_k) \geq 2\epsilon_0 > 0$ for $j \neq k$. Moreover, let $h : \mathbb{R} \rightarrow [0, 1]$ be a continuous function with support contained in $(-\epsilon_0, \epsilon_0)$ and $h(0) = 1$. Then each function of the form*

$$p = \sum_{j=1}^M \text{sign}(c_j) \nabla \cdot (h(b_j) \nabla b_j),$$

with b_j denoting the signed distance function to $\partial\Omega_j$, is an element of $\partial J(u) \cap L^2(\Omega)$.

Proof. From Lemma 7.1 we know that we just have to check whether $g = \sum_{j=1}^M h(b_j) \nabla b_j$ is a vector field in $L^\infty(\Omega)$ with supremum norm equal to one. Since the support of $h(b_j)$ and $h(b_k)$ do not intersect for $j \neq k$ for the above construction we obtain

$$\|g\|_\infty = \max_j \|h(b_j) \nabla b_j\|_\infty \leq \max_j \|h(b_j)\|_\infty \leq 1,$$

where we have used the property $|\nabla b_j| = 1$ a.e. for signed distance functions. Since we have assumed that each Ω_j has C^1 -boundary, $|\nabla b_j| = 1$ holds indeed in a neighbourhood of $\partial\Omega_j$, i.e. for b_j sufficiently small. Moreover h is continuous with $h(0) = 1$, and hence the essential supremum of $h(b_j) \nabla b_j$ is indeed equal to one in such a neighbourhood.

Finally we compute

$$\begin{aligned} \langle u, \nabla \cdot g \rangle &= \sum_{j,k} c_k \text{sign}(c_j) \int_{\Omega_k} \nabla \cdot (h(b_j) \nabla b_j) \, dx = \sum_{j,k} c_k \text{sign}(c_j) \int_{\partial\Omega_k} n_k \cdot (h(b_j) \nabla b_j) \, d\sigma \\ &= \sum_k |c_k| \int_{\partial\Omega_k} n_k \cdot (h(b_k) \nabla b_k) \, d\sigma \end{aligned}$$

where we have used the fact that $|b_j| \geq d(\partial\Omega_j, \partial\Omega_k) \geq 2\epsilon_0$ to deduce that $h(b_j) = 0$ on $\partial\Omega_k$ for $k \neq j$. Finally we can insert the normal vector $n_k = \nabla b_k$, $|\nabla b_k| = 1$, and $h(b_k) = h(0) = 1$ on $\partial\Omega_k$ to deduce

$$\langle u, \nabla \cdot g \rangle = \sum_k |c_k| \int_{\partial\Omega_k} d\sigma = J(u).$$

Hence, $p \in \partial J(u)$. □

In this case we can relate the source condition directly to the properties of the signed distance function, because we can write (SC) as

$$\sum_j \text{sign}(c_j) \nabla \cdot (h(b_j) \nabla b_j) = \text{sign}(c_j) \sum_j (h'(b_j) + h(b_j) \Delta b_j) = \xi = A^* q.$$

Since we can make h as smooth as necessary (still preserving the properties required in Lemma 7.2), the source condition is mainly a smoothness condition on the signed distance functions and their

Laplacians. Since for smooth curves (surfaces) Δb_j evaluated on $\partial\Omega_j$ equals the (mean) curvature, we see that the source condition is in particular a curvature condition on the discontinuity set.

In the case of classical denoising this source condition can be analyzed even more closely, since A^* can be identified with the identity on $L^2(\Omega)$. Hence, the source condition is satisfied if

$$h'(b_j) + h(b_j)\Delta b_j \in L^2(\Omega) \quad \forall j = 1, \dots, M.$$

Noticing that h , h' , and b_j are bounded, this condition is equivalent to $\Delta b_j \in L^2(\Omega)$, i.e. the source condition is satisfied if the (extended) mean curvature is square integrable.

It is also worth considering an extension to denoising with H^{-1} -fitting term (cf. [19] and [10] for a related wavelet approach). The operator A in this case is the embedding operator from $L^2(\Omega)$ into $H^{-1}(\Omega)$ (the dual space of the Sobolev space $H^1(\Omega)$). If we restrict the images to mean zero (which can be achieved by simple rescaling), then the H^{-1} -norm can be written as

$$\|v\|_{H^{-1}} = \|\nabla(-\Delta)^{-1}v\|_{L^2},$$

where $(-\Delta)^{-1}$ is an abbreviation of solving the Poisson equation with right-hand side v , mean value zero, and homogeneous Neumann boundary conditions. The adjoint operator of the embedding can then be calculated easily as $A^* = (-\Delta)^{-1}$ which maps $H^{-1}(\Omega)$ into $H^1(\Omega)$. Hence the source condition is equivalent to $h'(b_j) + h(b_j)\Delta b_j \in H^1(\Omega)$. Besides the same square-integrability condition as in the L^2 -denoising case one obtains an additional gradient condition

$$h''(b_j)\nabla b_j + h'(b_j)\Delta b_j\nabla b_j + h(b_j)\nabla(\Delta b_j) \in L^2(\Omega).$$

If we choose again a smooth h and use the fact that ∇b_j is even globally bounded (supremum norm equal to one), the effective condition becomes $\nabla\Delta b_j \in L^2(\Omega)$. Noticing that the normal derivatives are trivial for operators of the signed distance function, we conclude that the tangential derivatives of the (mean) curvature need to be square integrable in this case, a very natural extension of the above condition for L^2 denoising.

By choosing localizing functions h one can also derive geometrically interpretable error estimates for piecewise constant images. Take for example $h_\epsilon = h(\frac{\cdot}{\epsilon_0})$ with $\epsilon \downarrow 0$ and let $\xi_\epsilon = \sum_j \text{sign}(c_j)\nabla \cdot (h_\epsilon(b_j)\nabla b_j)$. With $\Omega_\epsilon = \bigcup_j \{|b_j| < \epsilon\}$ we have

$$\begin{aligned} D_J^{\xi_\epsilon}(u, \tilde{u}) &= J(u) - \langle u, \xi_\epsilon \rangle = \sup_{\|g\|_\infty=1} \int_\Omega u \nabla \cdot \left(g - \sum_j \text{sign}(c_j) h_\epsilon(b_j) \nabla b_j \right) dx \\ &= \sup_{\|g\|_\infty=1} \left[\int_{\Omega \setminus \Omega_\epsilon} u \nabla \cdot g \, dx + \int_{\Omega_\epsilon} u \left(g - \sum_j \text{sign}(c_j) h_\epsilon(b_j) \nabla b_j \right) dx \right] \\ &\geq \sup_{\|g\|_\infty=1} \int_{\Omega \setminus \Omega_\epsilon} u \nabla \cdot g \, dx = |u|_{BV(\Omega \setminus \Omega_\epsilon)}. \end{aligned}$$

Hence, from the error estimates for the variational method and the scale space flow we can estimate the total variation of the reconstruction on each set $\Omega \setminus \Omega_\epsilon$. Evaluating such estimates one has to take care of the right-hand side. In the case of denoising the right-hand side depends on the L^2 -norm of ξ_ϵ , which is of order $\epsilon^{-1/2}$ with the above choice of h_ϵ . Hence, there is a need of detailed balancing between δ , λ , and ϵ , the resulting estimate is of the form

$$|u|_{BV(\Omega \setminus \Omega_\epsilon)} \leq \frac{\lambda\delta^2}{2} + \frac{C}{2\lambda\epsilon}.$$

As a consequence of this estimate we see that reasonable estimates about smallness of the total variation of u away from the discontinuity sets of \tilde{u} can only be given for $\epsilon \gg \frac{1}{\lambda}$.

Our numerical experiments are focused on two simple images, while numerically they are approximated on three different grids: 64×64 , 128×128 and 256×256 . The first image f_c is

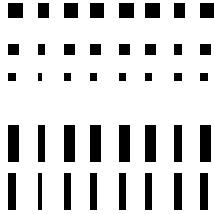


Figure 1: Block image

a centered circle $\chi_{B_0(R)}$ with radius $R = 0.25$ and the three approximating images are denoted as f_c^{64} , f_c^{128} and f_c^{256} . The second image f_b is a block image, shown in Figure 1 and the three approximating images are denoted as f_b^{64} , f_b^{128} and f_b^{256} . The whole domain is defined on $\Omega := [-0.5, 0.5] \times [-0.5, 0.5]$. By using procedure (11) where A is the identity operator, we will compare (between different grids) the sequence of the Bregman distances between the given image f and the iterative minimizers $\{u_k\}$. The use of different grids seems essential, since for a fixed grid every image should satisfy the source conditions somehow, but one would expect constants to deteriorate with finer grids if there is no source condition for the continuum limit.

The minimization problem (11) is usually solved by using the gradient descent method on the Euler-Lagrange equation. However, to calculate the subgradient of $J(u)$, which equals $-\nabla \cdot \frac{\nabla u}{|\nabla u|}$, it is necessary to approximate $|\nabla u|$ by $\sqrt{u_x^2 + u_y^2} + \epsilon$, where ϵ is very small. To avoid the inaccuracy brought by ϵ , we choose to solve the dual problem [7]

$$V_1 = \arg \min \left\| \frac{\nabla \cdot V}{\lambda} - f \right\|_2^2, \text{ s.t.}, \| |V| \|_\infty \leq 1. \quad (36)$$

Thus we obtain the residual image $v_1 = \frac{\nabla \cdot V_1}{\lambda}$ and the desired image $u_1 = f - v_1$. Further, we carry out the following equivalent iterative procedure of (8)

$$V_k = \arg \min \left\| \frac{\nabla \cdot V}{\lambda} - (f + v_{k-1}) \right\|_2^2, \text{ s.t.}, \| |V| \|_\infty \leq 1, \quad (37)$$

where $v_{k-1} = \frac{\nabla \cdot V_{k-1}}{\lambda}$, and the desired sequence of images $u_k = f + v_{k-1} - v_k$.

Following [7], we solve the dual problem (36) by the following semi-gradient descent (or fixed point) algorithm. We let $V_{i,j}^0 = (0, 0)$, and for any $n \geq 0$, we update V by

$$V_{i,j}^{n+1} = \frac{V_{i,j}^n + dt(\nabla(\nabla \cdot V^n - \lambda f))_{i,j}}{1 + dt|(\nabla(\nabla \cdot V^n - \lambda f))_{i,j}|}, \quad (38)$$

where $dt > 0$ is the time step, which is chosen as $\frac{1}{4}dx^2$, where dx is the spatial step. According to Theorem 3.1 from [7], this time step guarantees the convergence of the solution. Different from [7], we discretize ∇u and $\nabla \cdot V$ by the central scheme with the usual Neumann boundary condition for u and the periodic boundary condition for V . Furthermore, $V \cdot \bar{n} = 0$ is required on $\partial\Omega$, see [7, 1, 15].

Note that there does not exist any $q_b \in L^2(\Omega) \cap \partial J(f_b)$ due to the unbounded curvature. On the other hand, there exists a $q_c \in L^2(\Omega) \cap \partial J(f_c)$ (see [17, p. 37] and [4, Proposition 1]). In fact we can construct a subgradient

$$q_c(d) = \begin{cases} \frac{2}{R}, & \text{if } -R \leq d \leq 0; \\ -\frac{2R}{R_1^2 - R^2}, & \text{if } 0 < d \leq R_1 - R; \\ 0, & \text{else,} \end{cases} \quad (39)$$

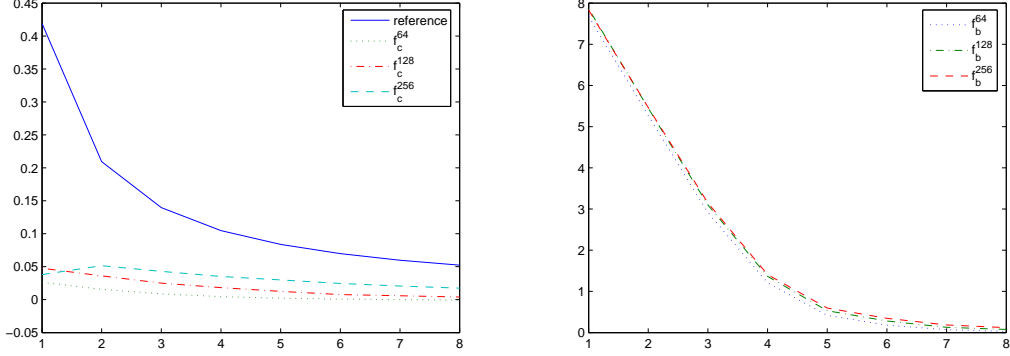


Figure 2: Results from (37) for the circle image (left) and the block image (right): Solid line represents $\frac{\|q_i\|^2}{2\lambda^k}$, dot line, dashdot line and dashed line represent the Bregman distance $D(f, u^k)$ for $f = f_i^{64}$, $f = f_i^{128}$ and $f = f_i^{256}$ correspondingly, $k = 1, \dots, 8$, $\lambda = 20$.

where R_1 is chosen such that $\chi_{B_0(R_1)} \subset \Omega$. For example, $R_1 = 0.5$. (Remark: for such a q_c , the continuous function $h : \mathbb{R} \rightarrow [0, 1]$ in Lemma 7.2. is defined as

$$h(d) = \begin{cases} \frac{d+R}{R}, & \text{if } -R \leq d \leq 0; \\ \frac{c_1}{d+R} - c_2(d+R), & \text{if } 0 < d \leq R_1 - R; \\ 0, & \text{else,} \end{cases} \quad (40)$$

where $c_1 = \frac{R_1^2 R}{R_1^2 - R^2}$ and $c_2 = \frac{R}{R_1^2 - R^2}$.

Figure 2 (left) illustrates that the Bregman distances between the circle images f_c^i and the sequence of minimizers $\{u^k\}$ from the iterative procedure (37) are bounded by $\frac{\|q\|^2}{2\lambda^k}$, where $\lambda = 20$. However, Figure 2 (right) indicates that the Bregman distances $D_J(f_b^i, u_k)$ can still be bounded independent of the grid size, even though no subgradient of $J(f_b)$ lies in $L^2(\Omega)$. This may be surprising, but does not contradict our results, where (SC) is only a sufficient condition. In addition, we have also tested the case of noisy data (Theorem 4.2) numerically on a noisy image $f_c^\delta = f_c^{64} + n^\delta$, where n^δ is a Gaussian noise of scale $\delta = 0.1$, also confirming our theoretical results.

8 Wavelet Shrinkage

In the following we consider the application of the error estimation techniques to wavelet methods. For the sake of simpler notation and without restriction of generality, we set up the problem directly on $\ell^2(\mathbb{Z}^2)$, the space of wavelet coefficients. In particular we interpret A as an operator on $\ell^2(\mathbb{Z}^2)$, which is no restriction since every linear operator can be defined via its realization on a basis of the underlying Hilbert space. The classical penalization functional in this case is

$$J(u) = \sum_{i,j \in \mathbb{Z}} w_{ij} |u_{ij}|, \quad (41)$$

assuming that $(w_{ij}) \in \ell^\infty(\mathbb{Z}^2)$ are given positive weights. It is easy to characterize subgradients:

Lemma 8.1. *Let $u \in \ell^1(\mathbb{Z}^2)$, then $p \in \partial J(u)$ if and only if*

$$p_{ij} \begin{cases} = w_{ij} & \text{if } u_{ij} > 0 \\ = -w_{ij} & \text{if } u_{ij} < 0 \\ \in [-w_{ij}, w_{ij}] & \text{if } u_{ij} = 0. \end{cases}$$

With these subgradients we can compute the Bregman distance as

$$\begin{aligned} D_J^p(v, u) &= J(v) - \langle p, v \rangle = \sum_{j,k \in \mathbb{Z}} (w_{ij} \operatorname{sign}(v_{ij}) - p_{ij}) v_{ij} \\ &= \sum_{v_{ij} u_{ij} < 0} 2w_{ij} |v_{ij}| + \sum_{u_{ij}=0, v_{ij} \neq 0} w_{ij} |v_{ij}| \left(1 - \frac{p_{ij}}{w_{ij}} \operatorname{sign}(v_{ij})\right). \end{aligned}$$

Note that for the particular subgradient p satisfying $p_{ij} = -\operatorname{sign}(v_{ij})w_{ij}$ for $u_{ij} = 0$ we obtain

$$D_J^p(v, u) = 2 \sum_{\operatorname{sign}(v_{ij}) \neq \operatorname{sign}(u_{ij})} |v_{ij}|.$$

With the analogous choice of a subgradient of v we can obtain the symmetric Bregman distance

$$D_J^p(v, u) = 2 \sum_{\operatorname{sign}(v_{ij}) \neq \operatorname{sign}(u_{ij})} (|u_{ij}| + |v_{ij}|) = 2 \sum_{\operatorname{sign}(v_{ij}) \neq \operatorname{sign}(u_{ij})} |u_{ij} - v_{ij}|$$

The Bregman distance therefore measures the deviation only for wavelet coefficients that have a different sign.

We can also interpret the source condition, at least if we bound the weights away from zero:

Proposition 8.2. *Let w_{ij} be weights as above such that $w_{ij} \geq \underline{w} > 0$ for all $i, j \in \mathbb{Z}$. Then, for any bounded linear operator $A : \ell^2(\mathbb{Z}^2) \rightarrow Z$, the source condition (SC) is satisfied only if u is sparse, i.e., there exists a finite set $I \subset \mathbb{Z}^2$ such that $u_{ij} = 0$ for $j, k \notin I$.*

Proof. Since A is a bounded operator on $\ell^2(\mathbb{Z}^2)$ we obtain $\mathcal{R}(A^*) \subset \ell^2(\mathbb{Z}^2)$. Since for $u_{ij} \neq 0$ we have $p_{ij}^2 \geq \underline{w}^2$, the sum can only converge if there is a finite number of such p_{ij} , i.e., if $u_{ij} = 0$ except for a finite number of indices. \square

Hence, the source condition restricts the admissible to sparse ones in the respective wavelet basis. In the denoising case, i.e., for A being the identity, the source condition is satisfied if and only if u is sparse. In the case $A = I$ and for weights $w_{ij} = 1$ the constant in the error estimate ($\|p\|$) is exactly the number of nonzero elements.

For the following numerical experiments we use $J(u) = \sum_{i,j \in \mathbb{Z}} |u_{ij}|$ and we minimize (6) by the soft shrinkage algorithm [11] with a threshold $\frac{1}{\lambda}$. Furthermore, by following the same approach in [18], we can derive the following Bregman iteration scheme for (6):

$$u^k = \arg \min \left\{ J(u) + \frac{\lambda}{2} \sum_{i,j \in \mathbb{Z}} |(f_{ij} + v_{ij}^{k-1}) - u_{ij}|^2 \right\}, \quad (42)$$

where $v_{ij}^k := f_{ij} + v_{ij}^{k-1} - u_{ij}^k$ and $v_{ij}^0 := 0$. As proved in [27], the above iterative procedure gives firm shrinkage [13].

We illustrate again the sequence of Bregman distances between the observed image f and the iterative minimizers $\{u^k\}$ for the circle image and the block image on different grids. Since we use the Haar wavelet as the basis, the block image has a sparse decomposition. A subgradient of $J(f_b^i)$ (for $i = 64, 128, 256$) approximately satisfies $\|q_b^{64}\|^2 \simeq 4\|q_b^{128}\|^2 \simeq 16\|q_b^{256}\|^2$. This can be seen from the decrease ratio of the Bregman distance sequences between different grids in Figure 3 (right). In contrast, the decrease ratio for the circle image is only around 2, see Figure 3 (left). This is due to the fact that the circle image is not sparse when decomposed by the Haar wavelet. Hence, in this case we have indeed found an example where the non-satisfied source condition at the continuum level indeed has an impact on finite grids: The convergence rate deteriorates with the grid refinement.

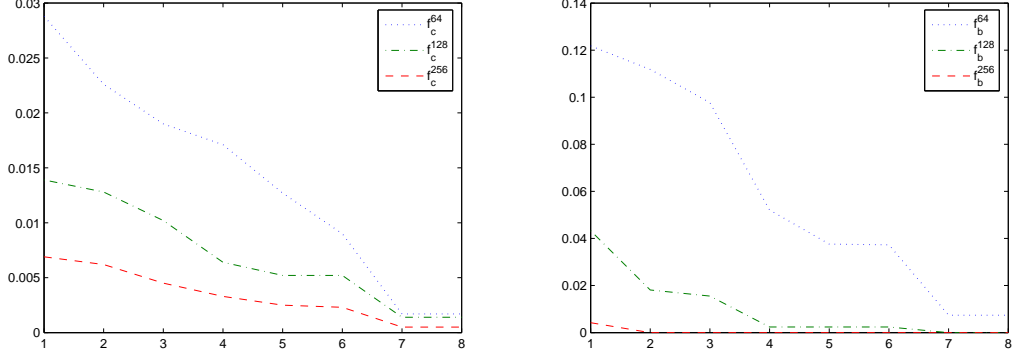


Figure 3: Results from (42) for the circle image (left) and block image (right): Dot line, dashdot line and dashed line represent the Bregman distance $D(f, u^k)$ for $f = f_i^{64}$, $f = f_i^{128}$ and $f = f_i^{256}$ correspondingly, $k = 1, \dots, 8$, $\lambda = 0.3$.

9 Nonnegative Image Restoration

An interesting problem in imaging is the deblurring with positivity constraint, which is of particular importance for density images as appearing in medical imaging. In this case the model for the penalization functional is $J(u) = J_0(u) + \chi_+(u)$, where J_0 is a standard regularization functional (like the ones above) and χ_+ is the indicator function of the nonnegativity constraint, i.e.

$$\chi_+(u) = \begin{cases} 0 & \text{if } u \geq 0 \text{ a.e.} \\ \infty & \text{else} \end{cases} \quad (43)$$

We assume for simplicity that J_0 is regular enough to guarantee that $\partial J(u) = \partial J_0(u) + \partial \chi_+(u)$. For instance, a sufficient condition in this respect is the following (cf. [12, Proposition 5.6]):

If h_1 and h_2 are two convex functions on X such that there is a point x_0 in the domain of h_1 and h_2 where h_1 is continuous, then

$$\partial(h_1 + h_2)(x) = \partial h_1(x) + \partial h_2(x), \quad \forall x \in X.$$

Then we can separately compute subgradients of χ_+ , which is a reasonably simple task:

Lemma 9.1. *Let $\chi_+ : L^r(\Omega) \rightarrow \mathbb{R} \cup \{\infty\}$ be defined by (43), and let $1 \leq r < \infty$. Then $p \in L^{r^*}(\Omega)$, for $r^* = \frac{r}{r-1}$, is a subgradient $p \in \partial \chi_+(u)$ for $u \in L^r(\Omega)$ with $\chi_+(u) = 0$, if and only if*

$$p = 0 \text{ a.e. on } \text{supp}(u), \quad p \leq 0 \text{ a.e. on } \Omega \setminus \text{supp}(u). \quad (44)$$

Proof. We start with the sufficiency of (44). We obviously need to verify the subgradient property

$$\langle v - u, p \rangle \leq \chi_+(v) - \chi_+(u) = \chi_+(v), \quad \forall v \in L^p(\Omega)$$

only for $\chi_+(v) = 0$, since it is trivial for $\chi_+(v) = \infty$. Let $v \in L^p(\Omega)$ be nonnegative almost everywhere. Then with p as above, we obtain

$$\langle p, v - u \rangle = \int_{\Omega \setminus \text{supp}(u)} p v \, dx \leq 0,$$

since $v \geq 0$ and $p \leq 0$ a.e. on $\Omega \setminus \text{supp}(u)$, which implies that p is indeed a subgradient.

On the other hand let $p \in \partial J(u)$. Then we can choose $v = u + \epsilon w$ for w being any bounded function supported in $\{u > \alpha\}$ for arbitrary $\alpha > 0$, and if $|\epsilon|$ is sufficiently small we have $v \geq 0$. Hence,

$$0 \geq \langle v - u, p \rangle = \epsilon \int_{\{u > \alpha\}} w p \, dx.$$

Since we can choose ϵ both positive and negative, we obtain $\int_{\{u>\alpha\}} wp \, dx = 0$. Since $\alpha > 0$ and w are arbitrary we conclude $p = 0$ on $\text{supp}(u)$. If we choose $v = u + w$, with an arbitrary nonnegative bounded function w , then v is still nonnegative. Hence

$$0 \geq \langle p, v - u \rangle = \int_{\{u=0\}} wp \, dx,$$

which implies $p \leq 0$ a.e. on $\{u = 0\}$ due to the arbitrariness of $w \geq 0$. \square

Now we can also compute the Bregman distance for a subgradient $q = p_0 + p \in \partial J(u)$ with $p_0 \in \partial J_0(u)$ and $p \in \partial \chi_+(u)$ as

$$D_J^q(v, u) = D_{J_0}^{p_0}(v, u) + \chi_+(v) + \int_{\{u=0\}} (-p)v \, dx = D_{J_0}^{p_0}(v, u) + \chi_+(v) + \int_{\{u=0\}} |p| |v| \, dx.$$

Hence, the Bregman distance adds a part that measures v on the active set $\{u = 0\}$. Since $|p|$ is arbitrary on $\{u = 0\}$ we can obtain estimates of L^s -norms of v for appropriate subgradients, e.g. by choosing $p = |v|^{s-1}$.

In order to get some further insight into the source condition, we consider the case $J_0(u) = \frac{1}{2}\|u\|_{L^2}^2$ and $A : L^2(\Omega) \rightarrow Z$ being a bounded linear operator. Since $\partial J_0(u) = \{u\}$ the source condition becomes $A^*q = \xi = u + p$. Hence on the subset $\text{supp}(u)$ we obtain $A^*q > 0$, since $p = 0$. On the complement set $\{u = 0\}$ we have $A^*q = p \leq 0$. Thus, the source condition can be written equivalently as $u(x) = ((A^*q)(x))_+$. This means that u only needs to be smooth on its support (if A^* is a smoothing operator, which happens in deblurring), but there can be cusps at the boundary of the support.

Acknowledgements

A major part of this work has been carried out when M.B. was with the Industrial Mathematics Institute, Johannes Kepler University Linz, and the Johann Radon Institute for Computational and Applied Mathematics. The authors thank Ronny Ramlau and Massimo Fornasier (both RICAM Linz) for providing some insight and links to literature for wavelet shrinkage. Financial support is acknowledged to the Austrian Science Fund FWF through project SFB F 013 / 08.

References

- [1] G.Aubert, J.F.Aujol, *Modeling very oscillating signals, application to image processing*, Appl. Math. Optim. **51** (2005), 163–182.
- [2] M.Bachmayr, *Iterative Total Variation Methods for Nonlinear Inverse Problems*, Master Thesis (Johannes Kepler University, Linz, 2007).
- [3] L.M.Bregman, *The relaxation method for finding the common point of convex sets and its application to the solution of problems in convex programming*, USSR Comp. Math. Math. Phys. **7** (1967), 200-217.
- [4] M.Burger, S.Osher, *Convergence rates for convex variational regularization*, Inverse Problems **20** (2004), 1411-1421.
- [5] M.Burger, K.Frick, S.Osher, O.Scherzer, *Inverse total variation flow*, SIAM Multiscale Mod. Simul. (2007), to appear.
- [6] M.Burger, G.Gilboa, S.Osher, J.Xu, *Nonlinear inverse scale space methods*, Comm. Math. Sci. **4** (2006), 179-212.

- [7] A.Chambolle, *An algorithm for total variation regularization and denoising*, J. Math. Imaging Vision **20** (2004), 89–97.
- [8] A.Chambolle, R.DeVore, N.Y.Lee, B.Lucier, *Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Trans. Image Proc., **7** (1998), 319–335.
- [9] T.Chan, J.Shen, *Image Processing and Analysis* (SIAM, Philadelphia, 2005).
- [10] I.Daubechies, G.Teschke, *Wavelet-based image decompositions by variational functionals*, in: F.Truchetet, *Wavelet Applications in Industrial Processing, Proc. SPIE 5266* (2004), 94-105.
- [11] D.Donoho, I.Johnstone, *Ideal spatial adaptation via wavelet shrinkage*, Biometrika **81** (1994), 425–455.
- [12] I.Ekeland, R.Temam, *Convex analysis and variational problems*, Corrected Reprint Edition, SIAM, Philadelphia, 1999.
- [13] H.Y. Gao, A.G. Bruce, *WaveShrink with firm shrinkage*, Statist. Sinica **7** (1997), 855–874.
- [14] L.He, T.C.Chung, S.Osher, T.Fang, P.Speier, *MR Image Reconstruction by Using the Iterative Refinement Method and Nonlinear Inverse Scale Space Methods*, UCLA, CAM 06-35.
- [15] S.Kindermann, S.Osher, J.Xu, *Denoising by BV-duality*, J. Sci. Comp. **28** (2006), 411-444.
- [16] J.Lie, J.M. Nordbotten, *Inverse scale spaces for nonlinear regularization*, J. Math. Imaging Vision (2006), to appear.
- [17] Y.Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, AMS, Providence, 2001.
- [18] S.Osher, M.Burger, D.Goldfarb, J.Xu, W.Yin, *An iterative regularization method for total variation-based image restoration*, SIAM Multiscale Model. Simul. **4** (2005), 460-489.
- [19] S.J.Osher, A.Sole, L.Vese, *Image decomposition and restoration using total variation minimization and the H^{-1} norm*, SIAM Multiscale Model. Simul. **1** (2003), 349–370.
- [20] S.Osher, J.Xu, *Iterative regularization and nonlinear inverse scale space applied to wavelet based denoising*, IEEE Trans. Image Proc., to appear.
- [21] E.Resmerita, *Regularization of ill-posed problems in Banach spaces: convergence rates*, Inverse Problems **21** (2005) 1303-1314.
- [22] E.Resmerita, O.Scherzer, *Error estimates for non-quadratic regularization and the relation to enhancing*, Inverse Problems **22** (2006) 801-814.
- [23] L.I.Rudin, S.J.Osher, E.Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D **60** (1992), 259–268.
- [24] O.Scherzer, C.Groetsch, *Inverse scale space theory for inverse problems*, in: M.Kerckhove, ed., *Scale-space and Morphology in Computer Vision.Proceedings of the third International Conference Scale-space 2001* (Springer, Berlin, 2001), 317-325.
- [25] F.Schoepfer, A.K.Louis, T.Schuster, *Nonlinear iterative methods for linear ill-posed problems in Banach spaces* Inverse Problems **22** (2006), 311-329.
- [26] E.Tadmor, S.Nezzar, L.Vese, *A multiscale image representation using hierarchical (BV; L_2) decompositions*, Multiscale Model. Simul., **2**, 554-579, 2004.
- [27] J.Xu, S.Osher, *Iterative regularization and nonlinear inverse scale space applied to wavelet based denoising*, IEEE Trans. Image Proc. (2007), to appear.