INSTITUTE OF PHYSICS PUBLISHING INVERSE PROBLEMS

Inverse Problems **20** (2004) 1411–1421 PII: S0266-5611(04)77358-X

Convergence rates of convex variational regularization

Martin Burger[1] **and Stanley Osher**

Department of Mathematics, UCLA, 520 Portola Plaza, Los Angeles, CA 90095-1555, USA

E-mail: martin.burger@jku.at and sjo@math.ucla.edu

Received 10 March 2004, in final form 10 May 2004
Published 19 July 2004
Online at stacks.iop.org/IP/20/1411
doi:10.1088/0266-5611/20/5/005

Abstract
The aim of this paper is to provide quantitative estimates for the minimizers of non-quadratic regularization problems in terms of the regularization parameter, respectively the noise level. As usual for ill-posed inverse problems, these estimates can be obtained only under additional smoothness assumptions on the data, the so-called source conditions, which we identify with the existence of Lagrange multipliers for a limit problem. Under such a source condition, we shall prove a quantitative estimate for the Bregman distance induced by the regularization functional, which turns out to be the natural distance measure to use in this case. We put a special emphasis on the case of total variation regularization, which is probably the most important and prominent example in this class. We discuss the source condition for this case in detail and verify that it still allows discontinuities in the solution, while imposing some regularity on its level sets.

1. Introduction

General variational regularization methods with convex functionals have become of growing importance compared to classical Tikhonov regularization in the last decade. The most prominent example of *total variation regularization* (sometimes also called the *bounded variation regularization*), originally introduced as a technique for image denoising (cf [19]) has been used in several applied inverse problems and analysed by several authors over the last decade (cf [1, 5, 6, 8, 9, 15, 18]). Besides the specific properties of functions of bounded variation such as the possibility of discontinuities in the solution, the total variation regularization is of particular interest in regularization theory since it motivates the study of regularization methods in a non-reflexive Banach space.

While the theory of regularization methods, in particular for linear inverse problems, seems to be almost complete in Hilbert spaces (cf [11] for a detailed exposition), fewer results

[1] On leave from: Industrial Mathematics Institute, Johannes Kepler Universität Linz, Austria.

are available in Banach spaces, in particular non-reflexive ones. While the fundamental questions, such as the qualitative stability and convergence analysis, can be carried out in a similar way as long as weak-star convergence (or compactness in some other topology) is available (cf [1]), the derivation of quantitative estimates between the exact solution of the inverse problem and the reconstruction obtained by regularization is still open. Typically, under additional smoothness assumptions on the solution, the so-called source conditions, one may expect a quantitative estimate for a distance between the solution of the regularized problem and the exact solution. These results are well known for quadratic regularizers (cf [11, 12]), but so far hardly any results have been obtained for non-quadratic regularization (except for the special case of maximum entropy regularization, where the problem can be transferred to an equivalent one with quadratic minimization, cf [13]). In this paper we shall derive such quantitative estimates for a very special difference measure between the exact and regularized solution, namely the *Bregman distance* induced by the regularization functional.

The framework we consider will be the following: let $K : \mathcal{U} \to \mathcal{H}$ be a continuous linear operator between a Banach space $\mathcal{U}$ and a Hilbert space $\mathcal{H}$, which is also continuous in the possibly weaker topology $\mathcal{T}$ on $\mathcal{U}$. Moreover, let $J : \mathcal{U} \to \mathbb{R} \cup \{\infty\}$ be a convex functional such that

(J1) the functional J is lower semicontinuous in a topology $\mathcal{T}$ on $\mathcal{U}$;
(J2) the sub-level sets $M_\rho := \{J \leqslant \rho\}$ are compact in the topology $\mathcal{T}$, and nonempty for $\rho \geqslant 0$.

We start from the ill-posed operator equation

$$Ku = f, \tag{1.1}$$

where $f \in \mathcal{R}(K) \subset \mathcal{H}$ represents exact data. Since we are interested in the ill-posed case where $\mathcal{R}(K)$ is not closed, the problem does not necessarily have a solution if f is replaced by noisy data $g \in \mathcal{H}$, even if f and g are arbitrarily close in $\mathcal{H}$. Therefore, the problem has to be regularized in order to obtain a stable approximation of the solution. We shall mainly consider a standard variational regularization strategy using penalization with a convex functional J, i.e., the regularized problem then consists in minimizing

$$\frac{\lambda}{2}\|Ku - g\|_{\mathcal{H}}^2 + J(u) \to \min_{u \in \mathcal{U}}, \tag{1.2}$$

where $\lambda \in \mathbb{R}_+$ is a (large) Lagrange multiplier. In the classical set-up of regularization theory, $\alpha = 1/\lambda$ is called the regularization parameter (cf [11]). A minimizer of (1.2) is called *regularized solution* below and will be denoted by u^λ. We mention that under the above assumptions (J1) and (J2) one can easily prove the existence of a regularized solution and the standard convergence results as $\lambda \to \infty$ along the lines of [1]. Since the proofs are standard and our focus in this paper is rather to obtain convergence rates (i.e., quantitative estimates as the λ tends to infinity), they shall be omitted here.

The remainder of this paper is organized as follows: in section 2 we introduce some fundamental notions needed in the further analysis, in particular we discuss source conditions and the concept of Bregman distances. In section 3 we state the main results, namely quantitative estimates for the minimizers of (1.2) and two related regularization models. In section 4 we briefly discuss the application of our results to classical regularization approaches such as Tikhonov regularization in Hilbert spaces or maximum entropy regularization, where we obtain well-known convergence rate results as special cases of our analysis. In section 5 we return to the motivating example of the total variation regularization, which we apply the results to and discuss the interesting implications of the source condition for this case.

2. Basic notions and conditions

In the following we shall introduce some notation and the basic conditions needed for the results of this paper. We shall always denote the norm in $\mathcal{U}$ by $\|.\|$ and the Hilbert space norm in $\mathcal{H}$ by $\|.\|_{\mathcal{H}}$. In the same way we denote the standard duality product between $\mathcal{U}^*$ and $\mathcal{U}$ by $\langle ., . \rangle$ and the scalar product in the Hilbert space $\mathcal{H}$ by $\langle ., . \rangle_{\mathcal{H}}$.

2.1. Source conditions

In the following we discuss suitable source conditions for regularization with general convex functional. Before specifying a source condition, we introduce the notion of a J *minimizing solution:*

Definition 1. *An element $\tilde{u} \in \mathcal{U}$ is called a J minimizing solution of (1.1) if $K\tilde{u} = f$ and*

$$J(\tilde{u}) \leqslant J(v) \quad \forall\, v \in \mathcal{U}, Kv = f.$$

Note that a J minimizing solution is always equal to a classical solution if K has no nullspace.

Since the range of K is not closed in general, the J minimizing solution is not necessarily a saddle-point of the associated Lagrangian

$$L(u, w) = J(u) - \langle w, Ku - f \rangle_{\mathcal{H}}, \tag{2.1}$$

as it would be in the well-posed case. The existence of a Lagrange multiplier $\tilde{w} \in \mathcal{H}$ such that

$$L(\tilde{u}, w) \leqslant L(\tilde{u}, \tilde{w}) \leqslant L(u, \tilde{w}) \quad \forall\, u \in \mathcal{U}, w \in \mathcal{H}$$

is an additional regularity condition on the J minimizing solution $\tilde{u}$. In the context of ill-posed problems, this is usually called the *source condition*.

Using the subgradient, defined for convex functionals by

$$\partial J(u) := \{p \in \mathcal{U}^* \mid J(v) \geqslant J(u) + \langle p, v - u \rangle, \forall\, v \in \mathcal{U}\},$$

we can give a different characterization of the source condition. The existence of a Lagrange multiplier $\tilde{w}$ is equivalent to

$$\exists\, \tilde{w} \in \mathcal{H} : \quad K^*\tilde{w} \in \partial J(\tilde{u}), \tag{2.2}$$

since from the definition of $\partial J(\tilde{u})$ we have

$$J(\tilde{u}) + \langle K^*\tilde{w}, u - \tilde{u} \langle \leqslant J(u), \quad \forall\, u \in \mathcal{U}$$

if and only if (2.2) holds. This inequality can be rewritten as

$$L(\tilde{u}, \tilde{w}) \leqslant L(u, \tilde{w}) \quad \forall\, u \in \mathcal{U}$$

by adding $\langle \tilde{w}, f \rangle_{\mathcal{H}}$ on both sides. We shall below refer to (2.2) as the *source condition*.

We mention that in the case of classical Tikhonov regularization, i.e., $J(u) = \|u\|^2$ in a Hilbert space $\mathcal{U}$, the source condition takes the well-known form (cf [11])

$$\tilde{u} = K^*\tilde{w}.$$

For Tikhonov regularization, a source condition allows one to derive a quantitative estimate between the J minimizing solution $\tilde{u}$ and the regularized solution (cf [12]). In the non-quadratic case, in particular for the case of bounded variation regularization, the derivation of quantitative estimates is still open. Below we shall derive quantitative estimates for general convex regularizers for the so-called *Bregman distance* between the J minimizing solution and the regularized solution.

We finally point out an interesting connection between the elements satisfying a source condition and the class of possible minimizers of the regularized problem (1.2):

Proposition 1. *Let $\lambda > 0$ be arbitrary, but fixed. Then the set of $\tilde{u} \in U$ satisfying the source condition (2.2) and the set of $\tilde{u} \in U$ being a minimizer of (1.2) for some $g \in \mathcal{H}$ are equal.*

Proof. First assume that $\tilde{u}$ is a minimizer of (1.2) for some $g \in \mathcal{H}$, then the necessary and sufficient first-order optimality condition implies

$$\lambda K^*(g - K\tilde{u}) \in \partial J(\tilde{u}).$$

Hence, $\tilde{u}$ satisfies (2.2) with $\tilde{w} := \lambda(g - K\tilde{u})$.

Vice versa, assume that $\tilde{u}$ satisfies (2.2), then it satisfies the first-order optimality condition for (1.2) with $g = \frac{1}{\lambda}\tilde{w} + K\tilde{u}$. Hence, it is a minimizer of (1.2). □

2.2. Generalized Bregman distances

We start with a short review of (generalized) Bregman distances related to a convex functional J. For the sake of simplicity we shall assume that $J(u) < \infty$ for all $u \in \mathcal{U}$ in the following. First assume that J is Fréchet-differentiable, which means that $\partial J(u)$ is a singleton for each $u \in U$, i.e., $\partial J(u) = \{\nabla J(u)\}$. Then the Bregman distance (cf [4]) $D_J : \mathcal{U} \times \mathcal{U} \to \mathbb{R}^+$ of two elements $u, v \in \mathcal{U}$ is defined by

$$D_J(u, v) := J(u) - J(v) - \langle \nabla J(v), u - v\rangle. \tag{2.3}$$

One obviously obtains $D_J(u, u) = 0$, and the convexity of J implies that D_J is really a distance, i.e., $D_J(u, v) \geqslant 0$. If in addition J is strictly convex, which is the standard case of application of Bregman distances, then we also have that $D_J(u, v) = 0$ if and only if $u = v$.

Kiwiel [14] generalizes the concept of Bregman distance to the nonsmoooth functionals, which are strictly convex. In this case one can still introduce a classical distance D_J which satisfies $D_J(u, v) > 0$ if $u \neq v$. Since our main motivation is regularization functionals like the total variation seminorm, which are neither continuously differentiable nor strictly convex, we introduce a further generalization of Bregman distance based on the subgradient. In this case, the subgradient is not a singleton, and we obtain a family of distances via

$$D_J(u, v) := \{J(u) - J(v) - \langle p, u - v\rangle \mid p \in \partial J(v)\}. \tag{2.4}$$

Consequently each element $d \in D_J(u, v)$ represents a distance between the elements u and v. However, for functionals that are not strictly convex, it is possible that $0 \in D_J(u, v)$ for $u \neq v$. Moreover, it is not guaranteed that $D_J(u, v)$ is nonempty, since $\partial J(v)$ is nonempty. As we shall see below, this is not an issue in the case of convex variational regularization, since we will always obtain elements in the subgradient of J at the regularized solution and the source condition implies the existence of an element in the subgradient of J at the J minimizing solution.

Obviously, the generalized notion of Bregman distance we use is a very weak measure for the difference between two elements in a Banach space. On the other hand it seems more natural to measure differences between regularized solutions and J minimizing solutions of an ill-posed problem in this distance than in the Banach space norm, since the whole regularization procedure does not have a connection to the norm or the norm topology (but the usually weaker topology $\mathcal{T}$ under the above assumptions (J1) and (J2)). Suppose, for example, that the regularization functional is invariant with respect to a subspace $\mathcal{V} \subset \mathcal{U}$, i.e.

$$J(u + v) = J(u) \qquad \forall\, u \in \mathcal{U}, v \in \mathcal{V},$$

which happens, e.g., for J being the total variation seminorm on the subspace $\mathcal{V}$ of constant functions. Then it is easy to show that

$$\langle p, v \rangle = 0 \qquad \forall u \in \mathcal{U},\ p \in \partial J(u),\ v \in \mathcal{V}$$

and hence $D_J(u + v, u) = \{0\}$, i.e., the generalized Bregman distance does not distinguish the elements u and v, whereas the difference $\|u + v - u\| = \|v\|$ can be arbitrarily large. This example supports the viewpoint that the Bregman distance controls only errors that can be distinguished by the regularization term. In order that the regularization approach makes sense, the errors that cannot be distinguished by the regularization term must be controlled by the fitting term anyway, for which one usually can derive a better quantitative estimate anyway.

3. Convergence rates for convex regularization methods in Banach spaces

Using the Bregman distance defined in the last section we are now able to derive the main results of this paper, quantitative estimates between a J minimizing solution $\tilde{u}$ and a regularized solution. We start with an estimate in the case of exact data.

Theorem 1 (exact data)**.** *Let $g = f$ and let $\tilde{u}$ be a J minimizing solution of (1.1). In addition, assume that the source condition (2.2) is satisfied. Then, for each minimizer u^λ of (1.2), there exists $d \in D_J(u^\lambda, \tilde{u})$ such that estimate*

$$d \leqslant \frac{\|\tilde{w}\|^2_{\mathcal{H}}}{2\lambda} = \mathcal{O}(\lambda^{-1}) \tag{3.1}$$

holds.

Proof. Let

$$d = J(u^\lambda) - J(\tilde{u}) - \langle K^*\tilde{w}, u^\lambda - \tilde{u}\rangle \in D_J(u^\lambda, \tilde{u}).$$

Since u^λ is a minimizer of the regularized problem and $K\tilde{u} = f$, we have

$$\frac{\lambda}{2}\|Ku^\lambda - f\|^2_{\mathcal{H}} + J(u^\lambda) \leqslant J(\tilde{u}).$$

Hence,

$$\frac{\lambda}{2}\|Ku^\lambda - f\|^2_{\mathcal{H}} + d + \langle \tilde{w}, Ku^\lambda - f\rangle_{\mathcal{H}} \leqslant 0,$$

and by adding $\frac{\|\tilde{w}\|^2_{\mathcal{H}}}{2\lambda}$ we obtain

$$\frac{\lambda}{2}\left\|Ku^\lambda - f + \frac{1}{\lambda}\tilde{w}\right\|^2_{\mathcal{H}} + d \leqslant \frac{\|\tilde{w}\|^2_{\mathcal{H}}}{2\lambda},$$

which yields (3.1). □

Now we turn our attention to the case of noisy data, i.e., we assume that $g \neq f$, but that a noise bound of the form

$$\|f - g\|_{\mathcal{H}} \leqslant \delta \tag{3.2}$$

is available. Then we can derive the following result.

Theorem 2 (noisy data)**.** *Let (3.2) hold and let $\tilde{u}$ be a J minimizing solution of (1.1). In addition, assume that the source condition (2.2) is satisfied. Then, for each minimizer u^λ of (1.2), there exists $d \in D_J(u^\lambda, \tilde{u})$ such that estimate*

$$d \leqslant \frac{\|\tilde{w}\|_{\mathcal{H}}^2}{2\lambda} + \frac{\lambda\delta^2}{2} \tag{3.3}$$

holds. In particular, if $\lambda \sim \delta^{-1}$, then

$$d = \mathcal{O}(\delta). \tag{3.4}$$

Proof. As in the proof of theorem 1 we define

$$d = J(u^\lambda) - J(\tilde{u}) - \langle K^*\tilde{w}, u^\lambda - \tilde{u}\rangle \in D_J(u^\lambda, \tilde{u})$$

and obtain using the noise bound (3.2)

$$\frac{\lambda}{2}\|Ku^\lambda - g\|_{\mathcal{H}}^2 + J(u^\lambda) \leqslant \frac{\lambda\delta^2}{2} + J(\tilde{u}).$$

Hence, we can derive the estimate

$$\frac{\lambda}{2}\left\|Ku^\lambda - g + \frac{1}{\lambda}\tilde{w}\right\|_{\mathcal{H}}^2 + d \leqslant \frac{\|\tilde{w}\|_{\mathcal{H}}^2}{2\lambda} + \frac{\lambda\delta^2}{2},$$

which implies (3.3) and finally (3.4) for the special choice $\lambda \sim \delta^{-1}$. □

Below, we shall discuss similar quantitative estimates for different regularization models.

3.1. Constrained model

A related regularization approach to (1.2) in the case of noisy data is the constrained minimization

$$J(u) \to \min_{u\in\mathcal{U}} \quad \text{subject to } \|Ku - g\|_{\mathcal{H}} \leqslant \delta. \tag{3.5}$$

One can show that there exists a Lagrange parameter $\lambda \in \mathbb{R}_+$ such that the minimizer of (3.5) and the regularized solution of (1.2) are the same. If one can show that $\lambda\delta \sim 1$, a quantitative estimate can be obtained from theorem 2. However, it seems easier to derive the quantitative estimate directly, which we will do in the following.

Theorem 3. *Let (3.2) hold and let $\tilde{u}$ be a J minimizing solution of (1.1). In addition, assume that the source condition (2.2) is satisfied. Then, for each minimizer u^δ of the constrained problem (3.5), there exists $d \in D_J(u^\delta, \tilde{u})$ such that the estimate*

$$d \leqslant 2\delta\|\tilde{w}\|_{\mathcal{H}} \tag{3.6}$$

holds.

Proof. Let

$$d = J(u^\delta) - J(\tilde{u}) - \langle K^*\tilde{w}, u^\delta - \tilde{u}\rangle \in D_J(u^\delta, \tilde{u}).$$

Since u^δ is a minimizer of (3.5) and $\tilde{u}$ is an element of the feasible set we have

$$d \leqslant -\langle K^*\tilde{w}, u^\delta - \tilde{u}\rangle = \langle \tilde{w}, g - f + f - Ku^\delta\rangle_{\mathcal{H}} \leqslant \|\tilde{w}\|_{\mathcal{H}}(\|Ku^\delta - g\|_{\mathcal{H}} + \delta),$$

which implies (3.6) because of the constraint $\|Ku^\delta - g\|_{\mathcal{H}} \leqslant \delta$. □

3.2. Exact penalization model

An alternative approach to regularization, with a particular emphasis on total variation denoising has been introduced recently in [7]:

$$\lambda\|Ku - g\|_{\mathcal{H}} + J(u) \to \min_{u\in\mathcal{U}}. \tag{3.7}$$

In the nomenclature of nonlinear programming, this model could be called the *exact penalization* of the constraint $Ku = g$. Indeed, the convergence properties for (3.7) are different from those for the models discussed above since in the case of exact data we obtain the J minimizing solution for finite λ.

Theorem 4 (exact data)**.** *Let $g = f$ and let $\tilde{u}$ be a J minimizing solution of (1.1). In addition, assume that the source condition (2.2) is satisfied and that $\lambda > \|\tilde{w}\|_{\mathcal{H}}$. Then, each minimizer u^λ of (3.7) is a J minimizing solution of $Ku = f$ and*

$$0 \in D_J(u^\lambda, \tilde{u}).$$

If in addition J is strictly convex on the nullspace of K, then $u^\lambda = \tilde{u}$.

Proof. Since u^λ is a minimizer of (3.7) and $g = f$ we have

$$J(u^\lambda) \leqslant \lambda\|Ku^\lambda - f\|_{\mathcal{H}} + J(u^\lambda) \leqslant J(\tilde{u})$$

and, with d as above,

$$\lambda\|Ku^\lambda - f\|_{\mathcal{H}} + d \leqslant -\langle \tilde{w}, Ku^\lambda - f\rangle_{\mathcal{H}} \leqslant \|\tilde{w}\|_{\mathcal{H}}\|Ku^\lambda - f\|_{\mathcal{H}}.$$

For $\lambda > \|\tilde{w}\|_{\mathcal{H}}$ we may conclude $Ku^\lambda = f$ and $d = 0$. Since $J(u^\lambda) \leqslant J(\tilde{u})$ we further obtain that u^λ is a J minimizing solution and therefore $J(u^\lambda) = J(\tilde{u})$. From the latter we may also conclude that $u^\lambda = \tilde{u}$ if J is strictly convex on the nullspace of K. □

The behaviour in the noisy case is similar, one does not need to take λ depending on δ in order to obtain a Bregman distance of order δ.

Theorem 5 (noisy data)**.** *Let (3.2) hold and let $\tilde{u}$ be a J minimizing solution of (1.1). In addition, assume that the source condition (2.2) is satisfied and that $\lambda > \|\tilde{w}\|_{\mathcal{H}}$. Then, for each minimizer u^λ of (3.7), there exists $d \in D_J(u^\lambda, \tilde{u})$ such that estimate*

$$d \leqslant (\lambda + \|\tilde{w}\|_{\mathcal{H}})\delta = \mathcal{O}(\delta) \tag{3.8}$$

holds.

Proof. With d as above we obtain

$$\begin{aligned}\lambda\|Ku^\lambda - f\|_{\mathcal{H}} + d &\leqslant \lambda\delta - \langle \tilde{w}, Ku^\lambda - g + g - f\rangle_{\mathcal{H}} \\ &\leqslant \lambda\delta + \|\tilde{w}\|_{\mathcal{H}}(\|Ku^\lambda - g\|_{\mathcal{H}} + \delta),\end{aligned}$$

which implies (3.8) if $\lambda > \|\tilde{w}\|_{\mathcal{H}}$. □

3.3. Possible generalizations

In the following we briefly discuss some obvious generalizations of the above analysis and state the corresponding results without giving theorems and proofs explicitly.

- *Non-attainable data.* One can easily derive analogous results for f not being in the closure of the range of K, i.e., $K\tilde{u}$ being equal to the projection of f onto the closure of the range. All estimates of theorems 1–5 still hold, but with different constants.
- *Additional constraints.* One can easily verify that all assertions of theorems 1–5 remain true if we add an additional constraint $u \in \mathcal{C}$ for some set $\mathcal{C}$ being closed in the topology $\mathcal{T}$, and if $\tilde{u} \in \mathcal{C}$.

- *Nonlinear problems*. The theory can easily be extended to nonlinear problems satisfying a nonlinearity condition of the form

$$\langle K(u) - K(v) - K'(v)(u - v), \tilde{w} \rangle_{\mathcal{H}} \leqslant \eta \|K(u) - K(v)\|_{\mathcal{H}} \|\tilde{w}\|_{\mathcal{H}}$$

with the source condition

$$\exists\, \tilde{w} \in \mathcal{H} : \quad K'(\tilde{u})^* \tilde{w} \in \partial J(u).$$

The proofs of theorems 1–3 can then be carried out in a similar way, it turns out that the constants on the right-hand side just become multiplied by $(1 + \eta)^2$. The assertions of theorems 4 and 5 hold in this case if $\lambda > (1 + \eta)\|\tilde{w}\|_{\mathcal{H}}$.

4. Application to classical regularization models

In the following we concretize the above theory for some classical regularization models, where convergence rates have been obtained by different techniques in the past.

4.1. Tikhonov regularization

For *Tikhonov regularization* $\mathcal{U}$ is a Hilbert space itself, $\mathcal{T}$ is the weak topology on $\mathcal{U}$, and

$$J(u) = \tfrac{1}{2}\|u\|^2, \qquad \partial J(u) = \{u\}.$$

Hence, the source condition has the well-known form $\tilde{u} = K^* \tilde{w}$, and the Bregman distance is given by

$$D_J(u, v) = \tfrac{1}{2}\|u\|^2 - \tfrac{1}{2}\|v\|^2 - \langle u - v, v \rangle = \tfrac{1}{2}\|u - v\|^2.$$

Hence, we obtain the well-known convergence rate (cf [12])

$$\|u^\lambda - \tilde{u}\| = \mathcal{O}(\sqrt{\delta}),$$

for $\lambda\delta \sim 1$ or for the constrained model.

4.2. Maximum entropy regularization

Another important example is *maximum entropy regularization* (cf [2, 10, 13]), where the regularization functional is the negative Shannon entropy

$$J(u) = \int_\Omega (u \ln u - u)\, \mathrm{d}x,$$

which can be defined on $L^1_+(\Omega)$ or in a generalized sense on a space of bounded measures. In both cases we have to add a non-negativity constraint, which does not change the results as noted above. For u positive and regular we have

$$\partial J(u) = \{\ln u\},$$

and hence, the source condition is of the form $\tilde{u} = \mathrm{e}^{K^* \tilde{w}}$. The Bregman distance induced by this functional is the so-called *Kullback–Leibler divergence* given by

$$D_J(u, v) = \int_\Omega u \ln\left(\frac{u}{v}\right) + v - u \,\mathrm{d}x.$$

Hence, the above theory implies a quantitative estimate in the Kullback–Leibler divergence, which is a natural distance measure in information theory.

5. Application to total variation denoising

Now consider the case of total variation denoising via the ROF-model (cf [19]), i.e., we have the space

$$\mathcal{U} = BV(\Omega), \qquad \mathcal{H} = L^2(\Omega),$$

for some domain $\Omega \subset \mathbb{R}^N$, $N \leqslant 3$ and the operator $K = I$ on $\mathcal{H}$. Moreover, the regularization functional is given by

$$J(u) = |u|_{BV} = \sup_{p \in C_0^\infty(\Omega)^N, \|p\|_\infty \leqslant 1} \int_\Omega u \operatorname{div} p \, \mathrm{d}x.$$

The above assumptions (J1) and (J2) are then satisfied if $\mathcal{T}$ is the weak-* topology on $BV(\Omega)$.

If u is a smooth function with $|\nabla u| > 0$, then the subgradient is a singleton given by

$$\partial J(u) = \{\kappa(u)\}, \qquad \kappa(u) = \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right).$$

The element $\kappa(u)$ has a geometric meaning, it represents the mean curvature of the level sets of u. Since K^* is the identity on $L^2(\Omega)$, the source condition in this case becomes

$$\kappa(u) \in L^2(\Omega),$$

i.e., it is a rather weak regularity condition on the level sets of u.

Motivated by this observation, we may conjecture that the source condition is also satisfied for discontinuous functions in $BV(\Omega)$ as long as their level sets and their discontinuity set have bounded curvature. An example of a discontinuous function that satisfies the source condition is the indicator function of a ball, i.e.,

$$\tilde{u}(x) = \begin{cases} 1 & \text{if } |x| < R \\ 0, & \text{else.} \end{cases}$$

Meyer [16] showed that in this case $\tilde{u}$ is the minimizer of (1.2) for some appropriate f and due to proposition 1 we may conclude that $\tilde{u}$ satisfies the source condition (2.2). The analysis in [16] and proposition 1 also provide an example of a function which does not satisfy (2.2), namely the indicator function of a square, which cannot be a minimizer of (1.2) for any $\lambda > 0$ and $g \in \mathcal{H}$. This confirms again that (2.2) in the case of total variation denoising is not a condition on the smoothness of $\tilde{u}$ (it can be discontinuous), but a condition on the smoothness of the discontinuity set and the level sets of $\tilde{u}$.

We also mention that along the lines of the proof of theorem 1 we can obtain an estimate of the form

$$m(u^\lambda, u^*) = \mathcal{O}(\delta),$$

for the metric

$$m(u, v) = \int_\Omega (u - v)^2 \, \mathrm{d}x + \left| \int_\Omega (|\nabla u| - |\nabla v|) \, \mathrm{d}x \right|,$$

which has been used, e.g., for the numerical analysis of problems involving total variation (cf [3]).

The estimates on the Bregman distance can be used to derive further estimates on the fine structure of minimizers of the regularized problem in the presence of noise. For example, consider again $f = \tilde{u}$ to be the indicator function of the ball $B_R(0)$ and let g be noisy data satisfying (3.2). Then, with (r, θ) denoting polar coordinates, the element

$$\tilde{w} = -\operatorname{div} \tilde{p}, \qquad \tilde{p} = q(|r - R|) \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$$

is a element of $\partial J(\tilde{u}) \cap L^2(\Omega)$ for q given by

$$q(s) = \max\left\{1 - \frac{s}{\epsilon}, 0\right\}.$$

For this particular element $\tilde{w}$ we obtain, due to the local support of $\tilde{w}$, that

$$\begin{aligned} d &= J(u^\lambda) - J(\tilde{u}) - \langle \tilde{w}, u^\lambda - \tilde{u}\rangle_{\mathcal{H}} \\ &= J(u^\lambda) - \int_\Omega \tilde{w} u^\lambda \, \mathrm{d}x \\ &\geqslant TV(u^\lambda|_{\Omega^\epsilon}), \end{aligned}$$

where the last term denotes the total variation of the function u^λ restricted to the set

$$\Omega^\epsilon = \{x \in \Omega \mid d(x, \partial B_R(0)) \geqslant \epsilon\} = (\Omega \backslash B_{R+\epsilon}(0)) \cup B_{R-\epsilon}(0).$$

Since $\|\tilde{w}\|_{\mathcal{H}} \leqslant \frac{C}{\sqrt{\epsilon}}$ for some constant $C \in \mathbb{R}_+$, we obtain by choosing $\lambda = \frac{\|\tilde{w}\|_{\mathcal{H}}}{\delta}$ in (3.3) the estimate

$$TV(u^\lambda|_{\Omega^\epsilon}) \leqslant C\frac{\delta}{\sqrt{\epsilon}}.$$

Together with the estimate on m we can also derive an estimate of the form

$$TV(u^\lambda|_{\Omega\backslash\Omega^\epsilon}) \geqslant TV(\tilde{u}) - C\left(\delta + \frac{\delta}{\sqrt{\epsilon}}\right) = 2R\pi - C\left(\delta + \frac{\delta}{\sqrt{\epsilon}}\right).$$

This shows that most of the variation of u^λ is concentrated around the set of discontinuity of $\tilde{u}$ for small noise.

We finally investigate an important case of an additional constraint, namely

$$\mathcal{C} = \{u \in BV(\Omega) \mid u(x) \in \{0, 1\} \text{ a.e.}\},$$

which is of particular importance for the study of binary images. In this case we can identify each function of $\mathcal{C}$ with the Lebesgue-measurable set

$$\Sigma_u = \{x \in \Omega \mid u(x) = 1\},$$

which has a Hausdorff-measurable boundary (cf [17]), since the Hausdorff measure of $\partial\Sigma$ (the so-called *perimeter*) is equal to the total variation seminorm of u. If f is also an indicator function, the minimization of (1.2) with parameter 2λ and of (3.7) with parameter λ are equivalent, they can be rewritten as the purely geometric variational problem

$$\lambda\mathcal{L}^N(\Sigma_u \Delta \Sigma_f) + \mathcal{H}^{N-1}(\partial\Sigma u) \to \min_{\Sigma_u},$$

where $\mathcal{L}^d$ denotes the Lebesgue measure, and $\mathcal{H}^{N-1}$ the $d-1$-dimensional Hausdorff measure. Thus, with the above techniques, we can even obtain quantitative estimates for geometric variational problems. In particular, due to the equivalence of (1.2) and (3.7) we obtain the J minimizing solution for exact data if λ is sufficiently large in both cases. A more detailed study of Bregman distances on classes of indicator functions is certainly an interesting topic for future research, it might allow us to derive further insight into the convergence of certain geometric features.

Acknowledgments

The authors thank Donald Goldfarb (Columbia University) and Elena Resmerita (Radon Institute for Computational and Applied Mathematics, Linz) for useful and stimulating discussions and for introducing them to the notion of Bregman distances. Financial support is acknowledged by the Austrian national science foundation FWF through grant SFB F 013/08, the ONR through grants N00014-02-1-0720, N00014-03-1-0071, the NIH through grant P20 MH65166, the NSF through grant DMS-0312222, and the NSF-funded Institute of Pure and Applied Mathematics (IPAM) at UCLA.

References

[1] Acar R and Vogel C R 1994 Analysis of total variation penalty methods *Inverse Problems* **10** 1217–29
[2] Amato U and Hughes W 1991 Maximum entropy regularization of Fredholm integral equations of the first kind *Inverse Problems* **7** 793–808
[3] Belik P and Luskin M 2003 Approximation by piecewise constant functions in a BV metric *Math. Models Methods Appl. Sci.* **13** 373–93
[4] Bregman L M 1967 The relaxation of finding the common points of convex sets and its application to the solution of problems in convex programming *USSR Comput. Math. Math. Phys.* **7** 200–17
[5] Casas E, Kunisch K and Pola C 1999 Regularization by functions of bounded variation and applications to image enhancement *Appl. Math. Optim.* **40** 229–57
[6] Chambolle A and Lions P L 1997 Image recovery via total variational minimization and related problems *Numer. Math.* **76** 167–88
[7] Chan T F and Esedoglu S 2004 Aspects of total variation regularized L^1 function approximation *CAM-Report 04-07* (Los Angeles: UCLA)
[8] Chavent G and Kunisch K 1997 Regularization of linear least squares problems by total bounded variation *ESAIM Control Optim. Calc. Var.* **2** 359–76
[9] Dobson D and Scherzer O 1996 Analysis of regularized total variation penalty methods for denoising *Inverse Problems* **12** 601–17
[10] Eggermont P B B 1993 Maximum entropy regularization for Fredholm integral equations of the first kind *SIAM J. Math. Anal.* **24** 1557–76
[11] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer)
[12] Engl H W, Kunisch K and Neubauer A 1989 Convergence rates for Tikhonov regularisation of nonlinear ill-posed problems *Inverse Problems* **5** 523–40
[13] Engl H W and Landl G 1993 Convergence rates for maximum entropy regularization *SIAM J. Numer. Anal.* **30** 1509–36
[14] Kiwiel K C 1997 Proximal minimization methods with generalized Bregman functions *SIAM J. Control Optim.* **35** 1142–68
[15] Luce R and Perez S 1999 Parameter identification for an elliptic partial differential equation with distributed noisy data *Inverse Problems* **15** 291–307
[16] Meyer Y 2001 *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations* (Providence, RI: American Mathematical Society)
[17] Morel J M and Solimini S 1994 *Variational Methods in Image Segmentation with 7 Image Processing Experiments* (Basel: Birkhäuser)
[18] Nashed M Z and Scherzer O 1998 Least squares and bounded variation regularization with nondifferentiable functionals *Numer. Funct. Anal. Optim.* **19** 873–901
[19] Rudin L I, Osher S J and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica* D **60** 259–68