Electronic Transactions on Numerical Analysis.
Volume 38, pp. 44-68, 2011.

ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

ERROR ESTIMATES FOR GENERAL FIDELITIES*

MARTIN BENNING† AND MARTIN BURGER†

Abstract. Appropriate error estimation for regularization methods in imaging and inverse problems is of enormous importance for controlling approximation properties and understanding types of solutions that are particularly favoured. In the case of linear problems, i.e., variational methods with quadratic fidelity and quadratic regularization, the error estimation is well-understood under so-called source conditions. Significant progress for nonquadratic regularization functionals has been made recently after the introduction of the Bregman distance as an appropriate error measure. The other important generalization, namely for nonquadratic fidelities, has not been analyzed so far.

In this paper we develop a framework for the derivation of error estimates in the case of rather general fidelities and highlight the importance of duality for the shape of the estimates. We then specialize the approach for several important fidelities in imaging (L^1, Kullback-Leibler).

Key words. error estimation, Bregman distance, discrepancy principle, imaging, image processing, sparsity

AMS subject classifications. 47A52, 65J20, 49M30

1. Introduction.

Image processing and inversion with structural prior information (e.g., sparsity, sharp edges) are of growing importance in practical applications. Such prior information is often incorporated into variational models with appropriate penalty functionals used for regularization, e.g., total variation or ℓ^1-norms of coefficients in orthonormal bases. The error control for such models, which is of obvious relevance, is the subject of this paper.

Most imaging and inverse problems can be formulated as the computation of a function $\tilde{u} \in \mathcal{U}(\Omega)$ from the operator equation,

$$K\tilde{u} = g\,, \tag{1.1}$$

with given data $g \in \mathcal{V}(\Sigma)$. Here $\mathcal{U}(\Omega)$ and $\mathcal{V}(\Sigma)$ are Banach spaces of functions on bounded and compact sets Ω, respectively Σ, and K denotes a linear operator $K : \mathcal{U}(\Omega) \to \mathcal{V}(\Sigma)$. We shall also allow Σ to be discrete with point measures, which often corresponds to the situation encountered in practice. In the course of this work we want to call g the *exact data* and $\tilde{u}$ the *exact solution*.

Most inverse problems are ill-posed, i.e., K usually cannot be inverted continuously (due to compactness of the forward operator). Furthermore, in real-life applications the exact data g are usually not available. Hence, we face to solve the inverse problem,

$$Ku = f, \tag{1.2}$$

instead of (1.1), with $u \in \mathcal{U}(\Omega)$ and $f \in \mathcal{V}(\Sigma)$, while g and f differ from each other by a certain amount. This difference is referred to as being *noise* (or systematic and modelling errors, which we shall not consider here). Therefore, throughout this work we want to call f the *noisy data*. Although in general g is not available, nevertheless in many applications a maximum noise bound δ is given. This "data error" controls the maximum difference between g and f in some measure, depending on the type of noise. For instance, in the case of the standard L^2-fidelity, we have the noise bound,

$$\|g - f\|_{L^2(\Sigma)} \leq \delta\,.$$

*Received July 31, 2010. Accepted for publication January 26, 2011. Published online March 14, 2011. Recommended by R. Ramlau.

†Westfälische Wilhelms-Universität Münster, Institut für Numerische und Angewandte Mathematik, Einsteinstr. 62, D 48149 Münster, Germany (martin.{benning,burger}@wwu.de).

In order to obtain a robust approximation $\hat{u}$ of $\tilde{u}$ for (1.2) many regularization techniques have been proposed. Here we focus on the particularly important and popular class of convex variational regularization, which is of the form

$$\hat{u} \in \operatorname*{arg\,min}_{u \in \mathcal{W}(\Omega)} \left\{ H_f(Ku) + \alpha J(u) \right\} , \tag{1.3}$$

with $H_f : \mathcal{V}(\Sigma) \to \mathbb{R} \cup \{\infty\}$ and $J : \mathcal{W}(\Omega) \to \mathbb{R} \cup \{\infty\}$, $\mathcal{W}(\Omega) \subseteq \mathcal{U}(\Omega)$, being convex functionals. The scheme contains the general fidelity term $H_f(Ku)$, which controls the deviation from equality of (1.2), and the regularization term $\alpha J(u)$, with $\alpha > 0$ being the regularization parameter, which guarantees certain smoothness features of the solution. In the literature, schemes based on (1.3) are often referred to as *variational regularization schemes*. Throughout this paper we shall assume that J is chosen such that a minimizer of (1.3) exists, the proof of which is not an easy task for many important choices of H_f; cf., e.g., [1, 2]. Notice that if $H_f(Ku) + \alpha J(u)$ in (1.3) is strictly convex, the set of minimizers is indeed a singleton.

Variational regularization of inverse problems based on general, convex—and in many cases singular—energy functionals has been a field of growing interest and importance over the last decades. In comparison to classical Tikhonov regularization (cf. [13]) different regularization energies allow the preservation of certain features, e.g., preservation of edges with the use of Total Variation (TV) as a regularizer (see for instance the well-known ROF-model [29]) or sparsity with respect to some bases or dictionaries.

By regularizing the inverse problem (1.2), our goal is to obtain a solution $\hat{u}$ close to $\tilde{u}$ in a robust way with respect to noise. Hence, we are interested in error estimates that describe the behaviour of the "data error" δ and optimal choices for quadratic fitting; see [13]. A major step for error estimates in the case of regularization with singular energies has been the introduction of (generalized) Bregman distances (cf. [4, 20]) as an error measure; cf. [8]. The Bregman distance for general convex, not necessarily differentiable functionals, is defined as follows.

DEFINITION 1.1 (Bregman Distance). *Let $\mathcal{U}$ be a Banach space and $J : \mathcal{U} \to \mathbb{R} \cup \{\infty\}$ be a convex functional with non-empty subdifferential ∂J. Then, the Bregman distance is defined as*

$$D_J(u, v) := \{ J(u) - J(v) - \langle p, u - v \rangle_{\mathcal{U}} \mid p \in \partial J(v) \} .$$

The Bregman distance for a specific subgradient $\zeta \in \partial J(v)$, $v \in \mathcal{U}$, is defined as $D_J^\zeta : \mathcal{U} \times \mathcal{U} \to \mathbb{R}^+$ with

$$D_J^\zeta(u, v) := J(u) - J(v) - \langle \zeta, u - v \rangle_{\mathcal{U}} .$$

Since we are dealing with duality throughout this work, we are going to write

$$\langle a, b \rangle_{\mathcal{U}} := \langle a, b \rangle_{\mathcal{U}^* \times \mathcal{U}} = \langle b, a \rangle_{\mathcal{U} \times \mathcal{U}^*} ,$$

for $a \in \mathcal{U}^*$ and $b \in \mathcal{U}$, as the notation for the dual product, for the sake of simplicity.

The Bregman distance is no distance in the usual sense; at least $D_J^\zeta(u, u) = 0$ and $D_J^\zeta(u, v) \geq 0$ hold for all $\zeta \in \partial J(v)$, the latter due to convexity of J. If J is strictly convex, we even obtain $D_J^\zeta(u, v) > 0$ for $u \neq v$ and $\zeta \in \partial J(v)$. In general, no triangular inequality nor symmetry holds for the Bregman distance. The latter one can be achieved by introducing the so-called symmetric Bregman distance.

DEFINITION 1.2 (Symmetric Bregman Distance). *Let $\mathcal{U}$ be a Banach space and $J : \mathcal{U} \to \mathbb{R} \cup \{\infty\}$ be a convex functional with non-empty subdifferential ∂J. Then, a symmetric Bregman distance is defined as $\mathrm{D}_{\mathrm{J}}^{\mathrm{symm}} : \mathcal{U} \times \mathcal{U} \to \mathbb{R}^+$ with*

$$\mathrm{D}_{\mathrm{J}}^{\mathrm{symm}}(u_1, u_2) := D_J^{p_1}(u_2, u_1) + D_J^{p_2}(u_1, u_2) = \langle u_1 - u_2, p_1 - p_2 \rangle_{\mathcal{U}^*} ,$$

with

$$p_i \in \partial J(u_i) \quad \textit{for} \quad i \in \{1, 2\} .$$

Obviously, the symmetric Bregman distance depends on the specific selection of the subgradients p_i, which will be suppressed in the notation for simplicity throughout this work.

Many works deal with the analysis and error propagation by considering the Bregman distance between $\hat{u}$ satisfying the optimality condition of a variational regularization method and the exact solution $\tilde{u}$; cf. [7, 9, 17, 21, 22, 28]. The Bregman distance turned out to be an adequate error measure since it seems to control only those errors that can be distinguished by the regularization term. This point of view is supported by the need of so-called source conditions, which are needed to obtain error estimates in the Bregman distance setting. In the case of quadratic fitting we have the source condition,

$$\exists \, \xi \in \partial J(\tilde{u}), \, \exists \, q \in L^2(\Sigma) \; : \; \xi = K^* q ,$$

with K^* denoting the adjoint operator of K throughout this work. If, e.g., in the case of denoising with $K = Id$, the exact image $\tilde{u}$ contains features that are not elements of the subgradient of J, error estimates for the Bregman distance cannot be applied since the source condition is not fulfilled.

Furthermore, Bregman distances according to certain regularization functionals have widely been used to replace those regularization terms, which yield inverse scale space methods with improved solutions of inverse problems; cf. [5, 6, 26].

Most works deal with the case of quadratic fitting, with only few exceptions; see, e.g., [27]. However, in many applications, such as Positron Emission Tomography (PET), Microscopy, CCD cameras, or radar, different types of noise appear. Examples are Salt-and-Pepper noise, Poisson noise, additive Laplace noise, and different models of multiplicative noise.

In the next section, we present some general fidelities as recently used in various imaging applications. Next, we present basic error estimates for general, convex variational regularization methods, which we apply to the specific models. Then we illustrate these estimates and test their sharpness by computational results. We conclude with a brief outlook and formulate open questions. We would also like to mention the parallel development on error estimates for variational models with non-quadratic fidelity in [27], which yields the same results as our paper in the case of Laplacian noise. Since the analysis in [27] is based on fidelities that are powers of a metric instead of the noise models we use here, most approaches appear orthogonal. In particular, we base our analysis on convexity and duality and avoid the use of triangle inequalities, which can only be used for a metric.

2. Non-quadratic fidelities. In many applications different fidelities than the standard L^2-fidelity are considered, usually to incorporate different a priori knowledge on the distribution of noise. Exemplary applications are Synthetic Aperture Radar, Positron Emission Tomography or Optical Nanoscopy. In the following, we present three particular fidelities, for which we will derive specific estimates later on.

ETNA
Kent State University
http://etna.math.kent.edu

2.1. General norm fidelity. Typical non-quadratic fidelity terms are norms in general, i.e.,

$$H_f(Ku) := \|Ku - f\|_{\mathcal{V}(\Sigma)} \ ,$$

without taking a power of it. The corresponding variational problem is given via

$$\hat{u} \in \arg\min_{u \in \mathcal{W}(\Omega)} \left\{ \|Ku - f\|_{\mathcal{V}(\Sigma)} + \alpha J(u) \right\} . \tag{2.1}$$

The optimality condition of (2.1) can be computed as

$$K^*\hat{s} + \alpha\hat{p} = 0, \qquad \hat{s} \in \partial\|K\hat{u} - f\|_{\mathcal{V}(\Sigma)}, \quad \hat{p} \in \partial J(\hat{u})\,. \tag{2.2}$$

In the following we want to present two special cases of this general class of fidelity terms that have been investigated in several applications.

2.1.1. L^1 fidelity. A typical non-quadratic, nondifferentiable, fidelity term used in applications involving Laplace-distributed or impulsive noise (e.g., Salt'n'Pepper noise), is the L^1-fidelity; see for instance [10, 11, 31]. The related variational problem is given via

$$\hat{u} = \arg\min_{u \in \mathcal{W}(\Omega)} \left\{ \int_\Sigma |(Ku)(y) - f(y)| \ d\mu(y) + \alpha J(u) \right\} . \tag{2.3}$$

The optimality condition of (2.3) can easily be computed as

$$K^*\hat{s} + \alpha\hat{p} = 0, \qquad \hat{s} \in \mathrm{sign}(K\hat{u} - f), \quad \hat{p} \in \partial J(\hat{u})\,,$$

with $\mathrm{sign}(x)$ being the signum "function", i.e.,

$$\mathrm{sign}(x) = \begin{cases} 1 & \text{for } x > 0 \\ \in [-1, 1] & \text{for } x = 0 \\ -1 & \text{for } x < 0 \end{cases} .$$

2.1.2. BV* fidelity. In order to separate an image into texture and structure, in [23] Meyer proposed a modification of the ROF model via

$$F(u, v) := \|v\|_{BV^*(\Omega)} + \frac{1}{2\lambda} \sup_{\substack{q \in C_0^\infty(\Omega;\mathbb{R}^2) \\ \|q\|_\infty \le 1}} \int_\Omega u \ \mathrm{div} q \ dx$$

with respect to u (structure) and v (texture) for a given image $f = u + v$, and with $\|\cdot\|_{BV^*(\Omega)}$ defined as

$$\|w\|_{BV^*(\Omega)} := \inf_p \left\| \left(|p_1|^2 + |p_2|^2 \right)^{\frac{1}{2}} \right\|_{L^\infty(\Omega)} ,$$

subject to $\mathrm{div}\, p = w$. Initially the norm has been introduced as G-norm.

In this context, we are going to consider error estimates for the variational model

$$u \in \arg\min_{u \in \mathcal{W}(\Omega)} \left\{ \|Ku - f\|_{\mathrm{BV}^*(\Sigma)} + \alpha J(u) \right\}$$

with its corresponding optimality condition

$$K^*\hat{s} + \alpha\hat{p} = 0, \qquad \hat{s} \in \partial \left\|K\hat{u} - f\right\|_{\mathrm{BV}^*(\Sigma)} , \quad \hat{p} \in \partial J(\hat{u})\,.$$

ETNA
Kent State University
http://etna.math.kent.edu

2.2. Kullback-Leibler fidelity. In applications such as Positron Emission Tomography or Optical Nanoscopy, sampled data usually obey a Poisson process. For that reason, other fidelities than the L^2 fidelity have to be incorporated into the variational framework. The most popular fidelity in this context is the Kullback-Leibler divergence (cf. [24]), i.e.,

$$H_f(Ku) = \int_\Sigma \left[f(y) \ln \left(\frac{f(y)}{(Ku)(y)} \right) - f(y) + (Ku)(y) \right] \, d\mu(y) \, ,$$

Furthermore, due to the nature of the applications and their data, the function u usually represents a density that needs to be positive. The related variational minimization problem with an additional positivity constraint therefore reads as

$$\hat{u} \in \arg\min_{\substack{u \in \mathcal{W}(\Sigma) \\ u \geq 0}} \left\{ \int_\Sigma \left[f(y) \ln \left(\frac{f(y)}{(Ku)(y)} \right) - f(y) + (Ku)(y) \right] \, d\mu(y) + \alpha J(u) \right\} .$$

With the natural scaling assumption,

$$K^* \mathbf{1} = \mathbf{1} \, ,$$

we obtain the complementarity condition,

$$\begin{aligned} \hat{u} \geq 0, \qquad K^* \frac{f}{Ku} - \alpha \hat{p} \leq 1 \, , \\ \hat{u} \left(1 - K^* \frac{f}{K\hat{u}} + \alpha \hat{p} \right) = 0 \, , \qquad \hat{p} \in \partial J(\hat{u}) \, . \end{aligned} \tag{2.4}$$

2.3. Multiplicative noise fidelity. In applications such as Synthetic Aperture Radar the data is supposed to be corrupted by multiplicative noise, i.e., $f = (Ku)\, v$, where v represents the noise following a certain probability law and $Ku \geq 0$ is assumed. In [1], Aubert and Aujol assumed v to follow a gamma law with mean one and derived the data fidelity,

$$H_f(Ku) = \int_\Sigma \left[\ln \left(\frac{(Ku)(y)}{f(y)} \right) + \frac{f(y)}{(Ku)(y)} - 1 \right] \, d\mu(y) \, .$$

Hence, the corresponding variational minimization problem reads as

$$\hat{u} \in \arg\min_{u \in \mathcal{W}(\Omega)} \left\{ \int_\Sigma \left[\ln \left(\frac{(Ku)(y)}{f(y)} \right) + \frac{f(y)}{(Ku)(y)} - 1 \right] \, d\mu(y) + \alpha J(u) \right\} \tag{2.5}$$

with the formal optimality condition

$$0 = K^* \left(\frac{(K\hat{u})(y) - f(y)}{((K\hat{u})(y))^2} \right) + \alpha \hat{p} \, , \qquad \hat{p} \in \partial J(\hat{u}) \, .$$

One main drawback of (2.5) is that the fidelity term is not globally convex and therefore will not allow unconditional use of the general error estimates which we are going to derive in Section 3. In order to convexify this speckle noise removal model, in [19] Huang et al.

suggested the substitution $z(y) := \ln((Ku)(y))$ to obtain the entirely convex optimization problem,

$$\hat{z} = \arg\min_{z \in \mathcal{W}(\Sigma)} \left\{ \int_{\Sigma} \left[z(y) + f(y) e^{-z(y)} - 1 - \ln(f(y)) \right] d\mu(y) + \alpha J(z) \right\} \tag{2.6}$$

with optimality condition

$$1 - f(y) e^{-\hat{z}(y)} + \alpha \hat{p} = 0 \tag{2.7}$$

for $\hat{p} \in \partial J(\hat{z})$. This model is a special case of the general multiplicative noise model presented in [30]. We mention that in the case of total variation regularization a contrast change as above is not harmful, since the structural properties (edges and piecewise constant regions) are preserved.

3. Results for general models. After introducing some frequently used non-quadratic variational schemes, we present general error estimates for (convex) variational schemes. These basic estimates allow us to derive specific error estimates for the models presented in Section 2. Furthermore, we explore duality and discover an error estimate dependent on the convex conjugates of the fidelity and regularization terms.

In order to derive estimates in the Bregman distance setting we need to introduce the so-called source condition,

$$\exists\, \xi \in \partial J(\tilde{u}),\ \exists\, q \in \mathcal{V}(\Sigma)^* : \quad \xi = K^* q\,. \tag{SC}$$

As described in Section 1, the source condition (SC) in some sense ensures that a solution $\tilde{u}$ contains features that can be distinguished by the regularization term J.

3.1. Basic estimates. In this section we derive basic error estimates in the Bregman distance measure for general variational regularization methods.

To find a suitable solution of the inverse problem (1.2) close to the unknown exact solution $\tilde{u}$ of (1.1), we consider methods of the form (1.3). We denote a solution of (1.3), which fulfills the optimality condition due to the Karush-Kuhn-Tucker conditions (KKT), by $\hat{u}$.

First of all, we derive a rather general estimate for the Bregman distance $D_J^{\xi}(\hat{u}, \tilde{u})$.

LEMMA 3.1. *Let $\tilde{u}$ denote the exact solution of the inverse problem (1.1) and let the source condition (SC) be fulfilled. Furthermore, let the functional $J : \mathcal{W}(\Omega) \to \mathbb{R} \cup \{\infty\}$ be convex. If there exists a solution $\hat{u}$ that satisfies (1.3) for $\alpha > 0$, then the error estimate*

$$H_f(K\hat{u}) + \alpha D_J^{\xi}(\hat{u}, \tilde{u}) \leq H_f(g) - \alpha \langle q, K\hat{u} - g \rangle_{\mathcal{V}(\Sigma)}$$

holds.

Proof. Since $\hat{u}$ is an existing minimal solution satisfying (1.3), we have

$$H_f(K\hat{u}) + \alpha J(\hat{u}) \leq H_f(\underbrace{K\tilde{u}}_{=g}) + \alpha J(\tilde{u})\,.$$

If we subtract $\alpha \left(J(\tilde{u}) + \langle \xi, \hat{u} - \tilde{u} \rangle_{\mathcal{U}(\Omega)} \right)$ on both sides we end up with

$$\begin{aligned} H_f(K\hat{u}) + \alpha \underbrace{\left(J(\hat{u}) - J(\tilde{u}) - \langle \xi, \hat{u} - \tilde{u} \rangle_{\mathcal{U}(\Omega)} \right)}_{= D_J^{\xi}(\hat{u}, \tilde{u})} &\leq H_f(g) - \alpha \underbrace{\langle \xi, \hat{u} - \tilde{u} \rangle_{\mathcal{U}(\Omega)}}_{= \langle K^* q, \hat{u} - \tilde{u} \rangle_{\mathcal{V}(\Sigma)}} \\ &= H_f(g) - \alpha \langle q, K\hat{u} - g \rangle_{\mathcal{V}(\Sigma)}\,. \end{aligned}$$

∎

ETNA
Kent State University
http://etna.math.kent.edu

Notice that J needs to be convex in order to guarantee the positivity of $D_J^\xi(\hat{u}, \tilde{u})$ and therefore to ensure a meaningful estimate. In contrast to that, the data fidelity H_f does not necessarily need to be convex, which makes Lemma 3.1 a very general estimate. Furthermore, the estimate also holds for any $\overline{u}$ for which we can guarantee

$$H_f(K\overline{u}) + \alpha J(\overline{u}) \leq H_f(K\tilde{u}) + \alpha J(\tilde{u})$$

(a property that obviously might be hard to prove for a specific $\overline{u}$), which might be useful to study non-optimal approximations to $\hat{u}$. Nevertheless, we are mainly going to deal with a specific class of convex variational problems that allows us to derive sharper estimates, similar to Lemma 3.1 but for $D_J^{\text{symm}}(\hat{u}, \tilde{u})$. Before we prove these estimates, we define the following class of problems that we further want to investigate:

DEFINITION 3.2. *We define the class* $\mathcal{C}(\Phi, \Psi, \Theta)$ *as follows:*
$(H, J, K) \in \mathcal{C}(\Phi, \Psi, \Theta)$ *if*

- $K : \Theta \to \Phi$ *is a linear operator between Banach spaces* Θ *and* Φ*,*
- $H : \Phi \to \mathbb{R} \cup \{\infty\}$ *is proper, convex and lower semi-continuous,*
- $J : \Psi \to \mathbb{R} \cup \{\infty\}$ *is proper, convex and lower semi-continuous,*
- *there exists a* u' *with* $Ku' \in$ *dom*(H) *and* $u' \in$ *dom*(J)*, such that* H *is continuous at* Ku'*.*

With this definition we assume more regularity to the considered functionals and are now able to derive the same estimate as in Lemma 3.1, but for $\mathrm{D}_\mathrm{J}^{\text{symm}}(\hat{u}, \tilde{u})$ instead of $D_J^\xi(\hat{u}, \tilde{u})$.

THEOREM 3.3 (Basic Estimate I). *Let* $(H_f, J, K) \in \mathcal{C}(\mathcal{V}(\Sigma), \mathcal{W}(\Omega), \mathcal{U}(\Omega))$*, for compact and bounded sets* Ω *and* Σ*. Then, if the source condition* (SC) *is fulfilled, the error estimate*

$$H_f(K\hat{u}) + \alpha\, \mathrm{D}_\mathrm{J}^{\text{symm}}(\hat{u}, \tilde{u}) \leq H_f(g) - \alpha\langle q, K\hat{u} - g\rangle_{\mathcal{V}(\Sigma)} \tag{3.1}$$

holds.

Proof. Since H_f and J are convex, the optimality condition of (1.3) is given via

$$0 \in \partial H_f(K\hat{u}) + \alpha\partial J(\hat{u})\,.$$

Since both H_f and J are proper, lower semi-continuous, and convex, and since there exists u' with $Ku' \in \mathrm{dom}(H_f)$ and $u' \in \mathrm{dom}(J)$, such that H_f is continuous at Ku', we have $\partial H_f(Ku) + \alpha\partial J(u) = \partial\left(H_f(Ku) + \alpha J(u)\right)$ for all $u \in \mathcal{W}(\Omega)$, due to [12, Chapter 1, Section 5, Proposition 5.6]. Due to the linear mapping properties of K, we furthermore have $\partial H_f(K\cdot)(u) = K^*\partial H_f(Ku)$ Hence, the equality

$$K^*\hat{\eta} + \alpha\hat{p} = 0$$

holds for $\hat{\eta} \in \partial H_f(K\hat{u})$ and $\hat{p} \in \partial J(\hat{u})$. If we subtract $\alpha\xi$, with ξ fulfilling (SC), and take the duality product with $\hat{u} - \tilde{u}$, we obtain

$$\langle K^*\hat{\eta}, \hat{u} - \tilde{u}\rangle_{\mathcal{U}(\Omega)} + \alpha\langle \hat{p} - \xi, \hat{u} - \tilde{u}\rangle_{\mathcal{U}(\Omega)} = -\alpha\langle \underbrace{\xi}_{=K^*q}, \hat{u} - \tilde{u}\rangle_{\mathcal{U}(\Omega)}\,,$$

which equals

$$\langle \hat{\eta}, K\hat{u} - \underbrace{K\tilde{u}}_{=g}\rangle_{\mathcal{V}(\Sigma)} + \alpha D_J^{\text{symm}}(\hat{u}, \tilde{u}) = -\alpha\langle q, K\hat{u} - \underbrace{K\tilde{u}}_{=g}\rangle_{\mathcal{V}(\Sigma)}\,.$$

ETNA
Kent State University
http://etna.math.kent.edu

Since H_f is convex, the Bregman distance $D^{\hat{\eta}}_{H_f}(g, K\hat{u})$ is non-negative, i.e.,

$$D^{\hat{\eta}}_{H_f}(g, K\hat{u}) = H_f(g) - H_f(K\hat{u}) - \langle \hat{\eta}, g - K\hat{u}\rangle_{\mathcal{V}(\Sigma)} \geq 0\,,$$

for $\hat{\eta} \in \partial H_f(K\hat{u})$. Hence, we obtain

$$\langle \hat{\eta}, K\hat{u} - g\rangle_{\mathcal{V}(\Sigma)} \geq H_f(K\hat{u}) - H_f(g)\,.$$

As a consequence, this yields (3.1). □

We can further generalize the estimate of Theorem 3.3 to obtain the second important general estimate in this work.

THEOREM 3.4 (Basic Estimate II). *Let $(H_f, J, K) \in \mathcal{C}(\mathcal{V}(\Sigma), \mathcal{W}(\Omega), \mathcal{U}(\Omega))$ for compact and bounded sets Ω and Σ. Then, if the source condition (SC) is fulfilled, the error estimate*

$$\begin{aligned}(1-c)H_f(K\hat{u}) + \alpha\, \mathrm{D}^{\mathrm{symm}}_{\mathrm{J}}(\hat{u}, \tilde{u}) \leq{}& (1+c)H_f(g)\\ &- \alpha\langle q, f - g\rangle_{\mathcal{V}(\Sigma)} - cH_f(g)\\ &+ \alpha\langle q, f - K\hat{u}\rangle_{\mathcal{V}(\Sigma)} - cH_f(K\hat{u})\end{aligned} \tag{3.2}$$

holds for $c \in]0,1[$.

Proof. Due to Theorem 3.3, we have

$$H_f(K\hat{u}) + \alpha D^{\mathrm{symm}}_J(\hat{u}, \tilde{u}) \leq H_f(g) - \alpha\langle q, K\hat{u} - g\rangle_{\mathcal{V}(\Sigma)}\,.$$

The left-hand side is equivalent to

$$(1-c)H_f(K\hat{u}) + \alpha D^{\mathrm{symm}}_J(\hat{u}, \tilde{u}) + cH_f(K\hat{u})\,,$$

while the right-hand side can be rewritten as

$$(1+c)H_f(g) - \alpha\langle q, K\hat{u} - g\rangle_{\mathcal{V}(\Sigma)} - cH_f(g)$$

for $c \in]0,1[$, without affecting the inequality. The dual product $\langle q, K\hat{u} - g\rangle_{\mathcal{V}(\Sigma)}$ is equivalent to $\langle q, f + K\hat{u} - g - f\rangle_{\mathcal{V}(\Sigma)}$ and hence we have

$$-\alpha\langle q, K\hat{u} - g\rangle_{\mathcal{V}(\Sigma)} = -\alpha\langle q, f - g\rangle_{\mathcal{V}(\Sigma)} + \alpha\langle q, f - K\hat{u}\rangle_{\mathcal{V}(\Sigma)}\,.$$

Subtracting $cH_f(K\hat{u})$ on both sides and replacing $-\alpha\langle q, K\hat{u} - g\rangle_{\mathcal{V}(\Sigma)}$ by $-\alpha\langle q, f - g\rangle_{\mathcal{V}(\Sigma)} + \alpha\langle q, f - K\hat{u}\rangle_{\mathcal{V}(\Sigma)}$ yields (3.2). □

In Section 4 these two basic estimates will allow us to easily derive specific error estimates for the noise models described in Section 2.

3.2. A dual perspective. In the following we provide a formal analysis in terms of Fenchel duality, which highlights a general way to obtain error estimates and provides further insights. In order to make the approach rigorous one needs to check detailed properties of all functionals allowing to pass to dual problems formally (cf. [12]), which is however not our goal here.

In order to formulate the dual approach we redefine the fidelity to $G_f(Ku - f) := H_f(Ku)$ and introduce the convex conjugates

$$G^*_f(q) = \sup_{v\in\mathcal{V}(\Sigma)} \big(\langle q, v\rangle_{\mathcal{V}(\Sigma)} - G_f(v)\big)\,, \qquad J^*(p) = \sup_{u\in\mathcal{W}(\Omega)} \big(\langle p, u\rangle_{\mathcal{U}(\Omega)} - J(u)\big)\,,$$

ETNA
Kent State University
http://etna.math.kent.edu

for $q \in \mathcal{V}(\Sigma)^*$ and $p \in \mathcal{U}(\Omega)^*$. Under appropriate conditions, the Fenchel duality theorem (cf. [12, Chapter 3, Section 4]) implies the primal-dual relation,

$$\min_{u \in \mathcal{W}(\Omega)} \left[\frac{1}{\alpha} G_f(Ku - f) + J(u) \right] = - \min_{q \in \mathcal{V}(\Sigma)^*} \left[J^*(K^* q) - \langle q, f \rangle_{\mathcal{V}(\Sigma)} + \frac{1}{\alpha} G_f^*(-\alpha q) \right],$$

as well as a relation between the minimizers $\hat{u}$ of the primal and $\hat{q}$ of the dual problem, namely,

$$K^* \hat{q} \in \partial J(\hat{u}), \qquad \hat{u} \in \partial J^*(K^* \hat{q}).$$

More precisely, the optimality condition for the dual problem becomes

$$K\hat{u} - f - r = 0, \qquad r \in \partial G_f^*(-\alpha \hat{q}).$$

If the exact solution $\tilde{u}$ satisfies a source condition with source element d (i.e. $K^* d \in \partial J(\tilde{u})$), then we can use the dual optimality condition and take the duality product with $\hat{q} - d$, which yields

$$\langle K(\hat{u} - \tilde{u}), \hat{q} - d \rangle_{\mathcal{V}(\Sigma)^*} = \frac{1}{\alpha} \langle r, (-\alpha d) - (-\alpha \hat{q}) \rangle_{\mathcal{V}(\Sigma)^*} + \langle f - g, \hat{q} - d \rangle_{\mathcal{V}(\Sigma)^*}.$$

One observes that the left-hand side equals

$$\mathrm{D}_{\mathrm{J}}^{\mathrm{symm}}(\hat{u}, \tilde{u}) = \langle \hat{u} - \tilde{u}, K^*(\hat{q} - d) \rangle_{\mathcal{U}(\Omega)^*},$$

i.e., the Bregman distance we want to estimate. Using $r \in \partial G_f^*(-\alpha \hat{q})$, we find

$$\langle r, (-\alpha d) - (-\alpha \hat{q}) \rangle_{\mathcal{V}(\Sigma)^*} \leq G_f^*(-\alpha d) - G_f^*(-\alpha \hat{q}).$$

Under the standard assumption $G_f(0) = 0$, we find that G_f^* is nonnegative and hence in the noise-free case ($f = g$), we end up with the estimate,

$$\mathrm{D}_{\mathrm{J}}^{\mathrm{symm}}(\hat{u}, \tilde{u}) \leq \frac{1}{\alpha} G_f^*(-\alpha d).$$

Hence the error in terms of α is determined by the properties of the convex conjugate of G_f. For typical smooth fidelities G_f, we have $G_f^*(0) = 0$ and $(G_f^*)'(0) = 0$. Hence $\frac{1}{\alpha} G_f^*(-\alpha d)$ will at least grow linearly for small α, as confirmed by our results below.

In the applications to specific noise models our strategy will be to estimate the terms on the right-hand side of (3.2) by quantities like $G_f^*(-\alpha d)$ and then work out the detailed dependence on α.

4. Application to specific noise models. We want to use the basic error estimates derived in Section 3 to derive specific error estimates for the noise models presented in Section 2. In the following it is assumed that the operator K satisfies the conditions of Theorem 3.3 and Theorem 3.4.

4.1. General norm fidelity. With the use of Theorem 3.3 we can—in analogy to the error estimates for the exact penalization model in [8]—obtain the following estimate for $H_f(Ku) := \|Ku - f\|_{\mathcal{V}(\Sigma)}$ with $\hat{u}$ satisfying the optimality condition (2.2) and $\tilde{u}$ being the exact solution of (1.1).

THEOREM 4.1. *Let $\hat{u}$ satisfy the optimality condition (2.2) and let $\tilde{u}$ denote the exact solution of (1.1). Furthermore, the difference between exact data g and noisy data f is*

ETNA
Kent State University
http://etna.math.kent.edu

bounded in the $\mathcal{V}$-norm, i.e. $\|f-g\|_{\mathcal{V}(\Sigma)} \le \delta$ and (SC) holds. Then, for the symmetric Bregman distance $\mathrm{D}_{\mathrm{J}}^{\mathrm{symm}}(\hat{u},\tilde{u})$ for a specific regularization functional J, such that

$$(H_f, J, K) \in \mathcal{C}(\mathcal{V}(\Sigma), \mathcal{W}(\Omega), \mathcal{U}(\Omega))$$

is satisfied, the estimate

$$\left(1-\alpha \|q\|_{\mathcal{V}(\Sigma)^*}\right) H_f(K\hat{u}) + \alpha\, \mathrm{D}_{\mathrm{J}}^{\mathrm{symm}}(\hat{u},\tilde{u}) \le \left(1+\alpha\|q\|_{\mathcal{V}(\Sigma)^*}\right)\delta \tag{4.1}$$

holds. Furthermore, for $\alpha < 1/\|q\|_{\mathcal{V}(\Sigma)^*}$, we obtain

$$\mathrm{D}_{\mathrm{J}}^{\mathrm{symm}}(\hat{u},\tilde{u}) \le \delta\left(\frac{1}{\alpha} + \|q\|_{\mathcal{V}(\Sigma)^*}\right). \tag{4.2}$$

Proof. Since we have $(H_f, J, K) \in \mathcal{C}(\mathcal{V}(\Sigma), \mathcal{W}(\Omega), \mathcal{U}(\Omega))$, we obtain (due to Theorem 3.3)

$$\begin{aligned}
H_f(K\hat{u}) + \alpha\, \mathrm{D}_{\mathrm{J}}^{\mathrm{symm}}(\hat{u},\tilde{u}) &\le \underbrace{H_f(g)}_{\le\delta} - \alpha\langle q, K\hat{u}-g\rangle_{\mathcal{V}(\Sigma)} \\
&\le \delta - \alpha\langle q, K\hat{u}-f+f-g\rangle_{\mathcal{V}(\Sigma)} = \delta - \alpha\left(\langle q, K\hat{u}-f\rangle_{\mathcal{V}(\Sigma)} + \langle q, f-g\rangle_{\mathcal{V}(\Sigma)}\right) \\
&\le \delta + \alpha\|q\|_{\mathcal{V}(\Sigma)^*}\left(\|K\hat{u}-f\|_{\mathcal{V}(\Sigma)} + \|f-g\|_{\mathcal{V}(\Sigma)}\right) \\
&\le \delta + \alpha\|q\|_{\mathcal{V}(\Sigma)^*}\left(\|K\hat{u}-f\|_{\mathcal{V}(\Sigma)} + \delta\right),
\end{aligned}$$

which leads us to (4.1). If we insert $H_f(K\hat{u}) = \|K\hat{u}-f\|_{\mathcal{V}(\Sigma)}$ and set $\alpha < 1/\|q\|_{\mathcal{V}(\Sigma)^*}$, then we can subtract $\|K\hat{u}-f\|_{\mathcal{V}(\Sigma)}$ on both sides. If we divide by α, then we obtain (4.2). ∎

As expected from the dual perspective above, we obtain in the case of exact data ($\delta = 0$) for α sufficiently small

$$\mathrm{D}_{\mathrm{J}}^{\mathrm{symm}}(\hat{u},\tilde{u}) = 0, \qquad H_g(K\hat{u}) = 0.$$

For larger α no useful estimate is obtained. In the noisy case we can choose α small but independent of δ and hence obtain

$$\mathrm{D}_{\mathrm{J}}^{\mathrm{symm}}(\hat{u},\tilde{u}) = \mathcal{O}(\delta).$$

We remark on the necessity of the source condition (SC). In usual converse results one proves that a source condition needs to hold if the distance between the reconstruction and the exact solution satisfies a certain asymptotic in δ; cf. [25]. Such results so far exist only for quadratic fidelity and special regularizations and cannot be expected for general Bregman distance estimates—even less with non-quadratic fidelity models. We shall therefore only look on the asymptotics of H_f in the noise-free case and argue that for this asymptotic the source condition is necessary (at least in some sense). In the case of a general norm fidelity this is particularly simple due to the asymptotic exactness for α small. The optimality condition $K^*\hat{s} + \alpha\hat{p} = 0$ can be rewritten as

$$\hat{p} = K^* q, \qquad \hat{p} \in \partial J(\hat{u}), q \in \mathcal{V}(\Sigma)^*,$$

with $q = -\frac{1}{\alpha}\hat{s}$. Since $\hat{u}$ is a solution minimizing J for α sufficiently small, we see that if the asymptotic in α holds, there exists a solution of $Ku = g$ with minimal J satisfying (SC).

4.2. Poisson noise. In the case of Poisson noise the source condition can be written as

$$\exists \xi \in \partial J(\tilde{u}),\, \exists q \in L^\infty(\Sigma): \quad \xi = K^* q\,, \tag{SCL^1}$$

and we have the Kullback-Leibler fidelity,

$$H_f(Ku) = \int_\Sigma \left[f(y) \ln\left(\frac{f(y)}{(Ku)(y)}\right) - f(y) + (Ku)(y) \right] \, d\mu(y)\,,$$

and a positivity constraint $u \geq 0$. Theorem 3.4 will allow us to derive an error estimate of the same order as known for quadratic fidelities. Before that, we have to prove the following lemma.

LEMMA 4.2. *Let α and φ be positive, real numbers, i.e., $\alpha, \varphi \in \mathbb{R}^+$. Furthermore, let $\gamma \in \mathbb{R}$ be a real number and $c \in]0,1[$. Then, the family of functions*

$$h_n(x) := (-1)^n \alpha\gamma(\varphi - x) - c\left(\varphi \ln\left(\frac{\varphi}{x}\right) - \varphi + x\right)\,,$$

for $x > 0$ and $n \in \mathbb{N}$, are strictly concave and have their unique maxima at

$$\overline{x}_n^h = \frac{\varphi}{1 + (-1)^n \frac{\alpha}{c}\gamma}.$$

They are therefore bounded by

$$h_n(x) < h_n(\overline{x}_n^h) = (-1)^n \alpha\gamma\varphi - c\varphi \ln\left(1 + (-1)^n \frac{\alpha}{c}\gamma\right)$$

for $\frac{\alpha}{c}|\gamma| < 1$ and $x \neq \overline{x}_n^h$.

Proof. It is easy to see that $h_n^{''}(x) = -c\frac{\varphi}{x^2} < 0$ and, hence, h_n is strictly concave for all $n \in \mathbb{N}$. The unique maxima $\overline{x}_n^h$ can be computed via $h_n^{'}(\overline{x}_n^h) = 0$. Finally, since h_n is strictly concave for all $n \in \mathbb{N}$, $h_n(\overline{x}_n^h)$ has to be a global maximum. ∎

Furthermore, we have to ensure the existence of $u' \geq 0$ with $Ku' \in \text{dom}(H_f)$ and $u' \in \text{dom}(J)$, such that H_f is continuous at Ku'. If, e.g., $\text{dom}(J) = \text{BV}(\Omega)$, we do not obtain continuity of H_f at Ku' if K maps to, e.g., $L^1(\Sigma)$. Therefore, we restrict K to map to $L^\infty(\Sigma)$. However, we still keep (SCL^1) to derive the error estimates, which corresponds to an interpretation of K mapping to L^1. This implies more regularity than needed, since one usually uses q in the dual of the image space, which would mean $q \in L^\infty(\Sigma)^*$. For the latter we are not able to derive the same estimates. Note, however, that the assumption of K mapping to $L^\infty(\Sigma)$ is used only to deal with the positivity of K. With the help of Lemma 4.2 and the restriction to K we are able to prove the following error estimate.

THEOREM 4.3. *Let $\hat{u}$ satisfy the optimality condition (2.4) with $K : \mathcal{U}(\Omega) \to L^\infty(\Sigma)$ satisfying $\mathcal{N}(K) = \{0\}$, let $\tilde{u}$ denote the exact solution of (1.1), and let f be a probability density measure, i.e., $\int_\Sigma f\, d\mu(y) = 1$. Assume that the difference between noisy data f and exact data g is bounded in the Kullback-Leibler measure, i.e.,*

$$\int_\Sigma \left[f \ln\left(\frac{f}{g}\right) - f + g \right] \, d\mu(y) \leq \delta$$

and that (SCL^1) holds. Then, for $c \in]0,1[$ and $\alpha < \frac{c}{\|q\|_{L^\infty(\Sigma)}}$, the symmetric Bregman distance $\text{D}_{\text{J}}^{\text{symm}}(\hat{u}, \tilde{u})$ for a specific regularization functional J, such that

$$(H_f, J, K) \in \mathcal{C}(L^1(\Sigma), \mathcal{W}(\Omega), \mathcal{U}(\Omega))$$

is satisfied, is bounded via

$$(1-c)H_f(K\hat{u}) + \alpha\, \mathrm{D}_{\mathrm{J}}^{\mathrm{symm}}(\hat{u},\tilde{u}) \leq (1+c)\delta - c\ln\left(1 - \frac{\alpha^2}{c^2}\,\|q\|^2_{L^\infty(\Sigma)}\right). \tag{4.3}$$

Proof. We have $(H_f, J, K) \in \mathcal{C}(L^1(\Sigma), \mathcal{W}(\Omega), \mathcal{U}(\Omega))$. Using an analogous proof as in Theorem 3.4 with the non-negativity of $\hat{u}$ being incorporated in a variational inequality, we can still derive (3.2) in this case. Hence, we have to investigate $-\alpha\langle q, f-g\rangle_{L^1(\Sigma)} - cH_f(g)$ and $\alpha\langle q, f - K\hat{u}\rangle_{L^1(\Sigma)} - cH_f(K\hat{u})$. If we consider both functionals pointwise and force $\alpha^2 < \left(\frac{c}{q}\right)^2$, then we can use Lemma 4.2 to estimate

$$-\alpha\langle q, f-g\rangle_{L^1(\Sigma)} - cH_f(g) \leq \int\limits_\Sigma f\left(-\alpha q - c\ln\left(1 - \frac{\alpha}{c}q\right)\right)\, d\mu(y)$$

and

$$\alpha\langle q, f-K\hat{u}\rangle_{L^1(\Sigma)} - cH_f(K\hat{u}) \leq \int\limits_\Sigma f\left(\alpha q - c\ln\left(1 + \frac{\alpha}{c}q\right)\right)\, d\mu(y)\,.$$

Adding these terms together yields the estimate

$$(1-c)H_f(K\hat{u}) + \alpha\, \mathrm{D}_{\mathrm{J}}^{\mathrm{symm}}(\hat{u},\tilde{u}) \leq (1+c)\underbrace{H_f(g)}_{\leq\delta} + \int\limits_\Sigma f\left(-c\ln\left(1 - \frac{\alpha^2}{c^2}q^2\right)\right)d\mu(y)\,.$$

It is easy to see that for $\alpha < \frac{c}{\|q\|_{L^\infty(\Sigma)}}$ we have

$$-\ln\left(1 - \frac{\alpha^2}{c^2}q^2\right) \leq -\ln\left(1 - \frac{\alpha^2}{c^2}\,\|q\|^2_{L^\infty(\Sigma)}\right).$$

Hence, for positive f we obtain

$$\begin{aligned}(1-c)H_f(K\hat{u}) + \alpha\, \mathrm{D}_{\mathrm{J}}^{\mathrm{symm}}(\hat{u},\tilde{u}) \leq& (1+c)\delta + \int\limits_\Sigma f\left(-c\ln\left(1 - \frac{\alpha^2}{c^2}\,\|q\|^2_{L^\infty(\Sigma)}\right)\right)d\mu(y)\\ =&(1+c)\delta - c\ln\left(1 - \frac{\alpha^2}{c^2}\,\|q\|^2_{L^\infty(\Sigma)}\right)\underbrace{\int\limits_\Sigma f\, d\mu(y)}_{=1}\end{aligned}$$

and, hence, (4.3) holds. ☐

One observes from a Taylor approximation of the second term on the right-hand side of (4.3) around $\alpha = 0$ that

$$H_f(K\hat{u}) = \mathcal{O}\left(\delta + \alpha^2\right), \qquad \mathrm{D}_{\mathrm{J}}^{\mathrm{symm}}(\hat{u},\tilde{u}) = \mathcal{O}\left(\frac{\delta}{\alpha} + \alpha\right)$$

for small α, which is analogous to the quadratic case.

REMARK 4.4. The assumption $\mathcal{N}(K) = \{0\}$ is very strict. If $\mathcal{N}(K)$ is larger, the error estimate is still satisfied since H_f is convex (no longer strictly convex) and the terms $-\alpha\langle q, f-g\rangle_{L^1(\Sigma)} - cH_f(g)$ and $\alpha\langle q, f-K\hat{u}\rangle_{L^1(\Sigma)} - cH_f(K\hat{u})$ are concave (instead of being strictly concave). Hence, Lemma 4.2 can still be applied to find an upper estimate, the only difference is that there can be more than just one maximum.

ETNA
Kent State University
http://etna.math.kent.edu

4.3. Multiplicative noise. In the case of multiplicative noise we are going to examine model (2.6) instead of (2.5), since (2.6) is convex for all z and therefore allows the application of Theorem 3.4. The source condition differs slightly, since there is no operator in that type of model. Therefore, we get

$$(z\mathrm{SC}L^1) \qquad \exists\, \xi \in \partial J(\tilde{z}),\, \exists\, q \in L^\infty(\Sigma): \quad \xi = q\,.$$

In analogy to the Poisson case, we have to prove the following lemma first, to derive qualitative and quantitative error estimates in the case of multiplicative noise.

LEMMA 4.5. *Let α and φ be positive, real numbers, i.e., $\alpha, \varphi \in \mathbb{R}^+$. Furthermore, let $\gamma \in \mathbb{R}$ be a real number and $c \in]0,1[$. Then, the family of functions,*

$$k_n(x) := (-1)^n \alpha\gamma(\varphi - x) - c(x + \varphi e^{-x} - 1 - \ln(\varphi))$$

for $x > 0$ and $n \in \mathbb{N}$, are strictly concave and have their unique maxima at

$$\overline{x}_n^k = -\ln\left(\frac{c + (-1)^n \alpha\gamma}{c\varphi}\right)$$

for $\frac{\alpha}{c}|\gamma| < 1$. Hence, k_n is bounded via

$$k_n(x) < k_n(\overline{x}_n^k) = \alpha\gamma\left((-1)^n\left(\varphi + \ln\left(\frac{c + (-1)^n\alpha\gamma}{c\varphi}\right)\right) - 1\right) + c\ln\left(\frac{c + (-1)^n\alpha\gamma}{c}\right),$$

for $x \neq \overline{x}_n^k$.

Proof. Similarly to Lemma 4.2, it can easily be shown that $k_n''(x) = -c\varphi e^{-x} < 0$ for all $x \in \mathbb{R}^+$ and hence, the k_n are strictly concave for all $n \in \mathbb{N}$. The arguments $\overline{x}_n^k$ are computed to satisfy $k_n'(\overline{x}_n^k) = 0$. Since the k_n are strictly concave, $k_n(\overline{x}_n^k)$ has to be a global maximum for all $n \in \mathbb{N}$. ☐

With the help of Lemma 4.5, we are able to prove the following error estimate.

THEOREM 4.6. *Let $\hat{z}$ satisfy the optimality condition (2.7) and let $\tilde{z}$ denote the solution of $\tilde{z} = \ln(K\tilde{u}) = \ln(g)$, with $\tilde{u}$ being the exact solution of (1.1). Assume that the difference between noisy data f and exact data g is bounded in the measure of (2.5), i.e.,*

$$\int_\Sigma \ln\left(\frac{g}{f}\right) + \frac{f}{g} - 1 \, d\mu(y) \leq \delta$$

and that ($z\mathrm{SC}L^1$) holds. Then, for $c \in]0,1[$ and $\alpha < c/\left\|q\right\|_{L^\infty(\Sigma)}$, the symmetric Bregman distance $\mathrm{D}_{\mathrm{J}}^{\mathrm{symm}}(\hat{z},\tilde{z})$ for a specific regularization functional J such that

$$(H_f, J, Id) \in \mathcal{C}(L^1(\Sigma), \mathcal{W}(\Sigma), \mathcal{U}(\Sigma))$$

is satisfied, is bounded via

$$(4.4)\qquad (1-c)H_f(\hat{z}) + \alpha\,\mathrm{D}_{\mathrm{J}}^{\mathrm{symm}}(\hat{z},\tilde{z}) \leq (1+c)\delta + \alpha\left|\Sigma\right|\left\|q\right\|_{L^\infty(\Sigma)} \ln\left(\frac{c + \alpha\left\|q\right\|_{L^\infty(\Sigma)}}{c - \alpha\left\|q\right\|_{L^\infty(\Sigma)}}\right).$$

Proof. First of all, we have $H_f \in \mathcal{C}(L^1(\Sigma), \mathcal{W}(\Sigma), \mathcal{U}(\Sigma))$. Furthermore, there exists u' with $Ku' \in \mathrm{dom}(H_f)$ and $u' \in \mathrm{dom}(J)$, such that H_f is continuous at Ku'. Hence,

ETNA
Kent State University
http://etna.math.kent.edu

we can apply Theorem 3.4 to obtain (3.2). Therefore, we have to consider the functionals $-\alpha\langle q, f-g\rangle_{L^1(\Sigma)} - cH_f(g)$ and $\alpha\langle q, f-\hat{z}\rangle_{L^1(\Sigma)} - cH_f(\hat{z})$ pointwise. Due to Lemma 4.5 we have

$$\begin{aligned}
&-\alpha\langle q, f-g\rangle_{L^1(\Sigma)} - cH_f(g) + \alpha\langle q, f-\hat{z}\rangle_{L^1(\Sigma)} - cH_f(\hat{z}) \\
&\leq \int_\Sigma \alpha q\left(1 - f - \ln\left(\frac{c-\alpha q}{cf}\right)\right) + c\ln\left(\frac{c-\alpha q}{c}\right) d\mu(y) \\
&+ \int_\Sigma \alpha q\left(f + \ln\left(\frac{c-\alpha q}{cf}\right) - 1\right) + c\ln\left(\frac{c+\alpha q}{c}\right) d\mu(y) \\
&= \alpha \int_\Sigma q \underbrace{\left(\ln\left(\frac{c+\alpha q}{cf}\right) - \ln\left(\frac{c-\alpha q}{cf}\right)\right)}_{=\ln\left(\frac{c+\alpha q}{c-\alpha q}\right)} d\mu(y) \\
&+ c\int_\Sigma \underbrace{\left(\ln\left(\frac{c+\alpha q}{c}\right) + \ln\left(\frac{c-\alpha q}{c}\right)\right)}_{=\ln\left(1-\frac{\alpha^2}{c^2}q^2\right)} d\mu(y)\,,
\end{aligned}$$

for $\alpha < c/q$. It is easy to see that $q\ln\left(\frac{c+\alpha q}{c-\alpha q}\right) \leq \|q\|_{L^\infty(\Sigma)} \ln\left(\frac{c+\alpha\|q\|_{L^\infty(\Sigma)}}{c-\alpha\|q\|_{L^\infty(\Sigma)}}\right)$. Furthermore, it also easily can be verified that the function $l(x) := \ln\left(1 - \frac{\alpha^2}{c^2}x^2\right)$ is strictly concave and has its unique global maximum $l(\overline{x}) = 0$ at $\overline{x} = 0$. Hence, if we consider $\ln\left(1 - \frac{\alpha^2}{c^2}q^2\right)$ pointwise, $c\int_\Sigma \ln\left(1 - \frac{\alpha^2}{c^2}q^2\right) d\mu(y) \leq 0$ has to hold. Inserting these estimates into (3.2) yields (4.4). ☐

Again a Taylor approximation of the second term on the right-hand side of (4.4) around $\alpha = 0$ yields the asymptotic behaviour,

$$H_f(K\hat{u}) = \mathcal{O}\left(\delta + \alpha^2\right), \qquad \mathrm{D}_{\mathrm{J}}^{\mathrm{symm}}(\hat{u}, \tilde{u}) = \mathcal{O}\left(\frac{\delta}{\alpha} + \alpha\right).$$

5. A-posteriori parameter choice. Before we start discussing computational aspects and examples we want to briefly consider a-posteriori parameter choices for variational problems. A typical a-posteriori parameter choice rule is the discrepancy principle. For a general norm fidelity $\|Ku - f\|_{\mathcal{V}(\Sigma)}$ the discrepancy principle states that for a given noise bound $\|f - g\|_{\mathcal{V}(\Sigma)} \leq \delta$ the solution $\hat{u}$ to a regularized variational problem should satisfy $\|K\hat{u} - f\|_{\mathcal{V}(\Sigma)} \leq \delta$, i.e.,

$$\hat{u} \in \arg\min_{u\in\mathcal{W}(\Omega)} \left\{J(u)\right\}, \tag{5.1}$$

subject to

$$\|Ku - f\|_{\mathcal{V}(\Sigma)} \leq \delta\,. \tag{5.2}$$

We can reformulate this problem as

$$\hat{u} \in \arg\min_{u\in\mathcal{W}(\Omega)} \left\{\mathcal{X}_\delta\left(\|Ku - f\|_{\mathcal{V}(\Sigma)}\right) + J(u)\right\} \tag{5.3}$$

ETNA
Kent State University
http://etna.math.kent.edu

with $\mathcal{X}_\delta$ being the characteristic function

$$\mathcal{X}_\delta(v) := \begin{cases} 0 & \text{if } v \leq \delta \\ +\infty & \text{else} \end{cases}.$$

With the use of the triangular inequality of the norm and the monotonicity and convexity of the characteristic function, it easily can be seen that $\mathcal{X}_\delta\left(\|Ku - f\|_{\mathcal{V}(\Sigma)}\right)$ is convex, and by setting $H_f(Ku) = \mathcal{X}_\delta\left(\|Ku - f\|_{\mathcal{V}(\Sigma)}\right)$, we can apply Lemma 3.1 to obtain the following theorem.

THEOREM 5.1. *Let $\tilde{u}$ denote the exact solution of (1.1) and let the source condition (SC) be fulfilled. If there exists a minimal solution $\hat{u}$ satisfying (5.1) subject to (5.2) and if $\|f - g\|_{\mathcal{V}(\Sigma)}$ is also bounded by δ, the error estimate*

$$D_J^\xi(\hat{u}, \tilde{u}) \leq 2\delta \|q\|_{\mathcal{V}(\Sigma)^*}$$

holds.

Proof. If we apply Lemma 3.1 to the variational problem (5.3), we obtain

$$\begin{aligned}
\underbrace{\mathcal{X}_\delta\left(\underbrace{\|K\hat{u} - f\|_{\mathcal{V}(\Sigma)}}_{\leq\delta}\right)}_{=0} + D_J^\xi(\hat{u}, \tilde{u}) &\leq \underbrace{\mathcal{X}_\delta\left(\underbrace{\|f - g\|_{\mathcal{V}(\Sigma)}}_{\leq\delta}\right)}_{=0} - \langle q, K\hat{u} - g\rangle_{\mathcal{V}(\Sigma)} \\
&= -\left(\langle q, K\hat{u} - f\rangle_{\mathcal{V}(\Sigma)} + \langle q, f - g\rangle_{\mathcal{V}(\Sigma)}\right) \\
&\leq \|q\|_{\mathcal{V}^*(\Sigma)}\left(\underbrace{\|K\hat{u} - f\|_{\mathcal{V}(\Sigma)}}_{\leq\delta} + \underbrace{\|f - g\|_{\mathcal{V}(\Sigma)}}_{\leq\delta}\right) \\
&= 2\delta \|q\|_{\mathcal{V}(\Sigma)^*} \qquad \square
\end{aligned}$$

REMARK 5.2. Obviously a discrepancy principle also can be considered for general fidelities, not only for norm fidelities, i.e., we may replace $\|Ku - f\|_{\mathcal{V}(\Sigma)}$ in (5.2) by a general fidelity $H_f(Ku)$. In that case we can apply the same convex-combination trick as in Theorem 3.4 to obtain—with a subsequent computation of estimates for $-\alpha\langle q, f - g\rangle_{\mathcal{V}(\Sigma)} - cH_f(g)$ and $\alpha\langle q, f - K\hat{u}\rangle_{\mathcal{V}(\Sigma)} - cH_f(K\hat{u})$ (as in Lemma 4.2 and Lemma 4.5)—error estimates for $D_J^\xi(\hat{u}, \tilde{u})$.

6. Computational tests. In the following we present some numerical results for validating the practical applicability of the error estimates as well as their sharpness. We shall focus on two models, namely, the L^1-fidelity (due to the interesting asymptotic exactness) and the Kullback-Leibler fidelity (due to the practical importance).

6.1. Laplace noise. In order to validate the asymptotic exactness or non-exactness in the case of Laplacian noise we investigate a denoising approach with quadratic regularization, i.e., the minimization

$$\int_\Omega |u - f| \, dx + \frac{\alpha}{2} \int_\Omega (|\nabla u|^2 + u^2) \, dx \to \min_{u \in H^1(\Omega)}, \tag{6.1}$$

whose optimality condition is

$$-\alpha\Delta u + \alpha u + s = 0, \qquad s \in \text{sign}\,(u - f).$$

ETNA
Kent State University
http://etna.math.kent.edu

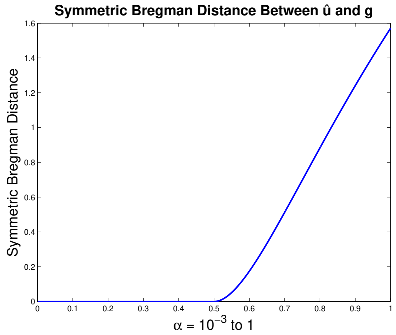


FIG. 6.1. *The Bregman distance error between* $\hat{u}$ *and* g *for* $\alpha \in [10^{-3}, 1]$. *As soon as* $\alpha \leq \frac{1}{2}$, *the error equals zero.*

A common approach to the numerical minimization of functionals such as (6.1) is a smooth approximation of the L^1-norm, e.g., by replacing $|u - f|$ with $\sqrt{(u-f)^2 + \epsilon^2}$ for small ϵ. Such an approximation will however alter the asymptotic properties and destroy the possibility to have asymptotic exactness. Hence we shall use a dual approach as an alternative, which we derive from the dual characterization of the one-norm,

$$\begin{aligned} &\inf_u \left[\int_\Omega |u - f| \, dx + \frac{\alpha}{2} \int_\Omega (|\nabla u|^2 + u^2) \, dx \right] \\ &\qquad = \inf_u \sup_{|s| \leq 1} \left[\int_\Omega (u - f)s \, dx + \frac{\alpha}{2} \int_\Omega (|\nabla u|^2 + u^2) \, dx \right] \\ &\qquad = \sup_{|s| \leq 1} \inf_u \left[\int_\Omega (u - f)s \, dx + \frac{\alpha}{2} \int_\Omega (|\nabla u|^2 + u^2) \, dx \right]. \end{aligned}$$

Exchanging infimum and supremum in the last formula can easily be justified with standard methods in convex analysis; cf. [12]. The infimum can be calculated exactly from solving $-\alpha \Delta u + \alpha u = -s$ with homogeneous Neumann boundary conditions, and hence we obtain after simple manipulation the dual problem (with the notation $A := (-\Delta \cdot + \cdot)$)

$$\frac{1}{2} \int_\Omega s(A^{-1}s) \, dx + \alpha \int_\Omega fs \, dx \to \min_{s \in L^\infty(\Omega), \|s\|_\infty \leq 1}.$$

This bound-constrained quadratic problem can be solved with efficient methods; we simply use a projected gradient approach, i.e.,

$$s^{k+1} = \mathbf{P}_1 \left(s^k - \tau \left(A^{-1} s^k + \alpha f \right) \right),$$

where $\tau > 0$ is a damping parameter and $\mathbf{P}_1$ is the pointwise projection operator,

$$\mathbf{P}_1(v)(x) = \begin{cases} v(x) & \text{if } |v(x)| \leq 1 \\ \frac{v(x)}{|v(x)|} & \text{else} \end{cases}.$$

Due to the quadratic H^1 regularization, we obtain

$$D_{H^1}^{\text{symm}}(\hat{u}, g) = 2 D_{H^1}(\hat{u}, g) = \|\hat{u} - g\|_{H^1(\Omega)}^2, \tag{6.2}$$

ETNA
Kent State University
http://etna.math.kent.edu

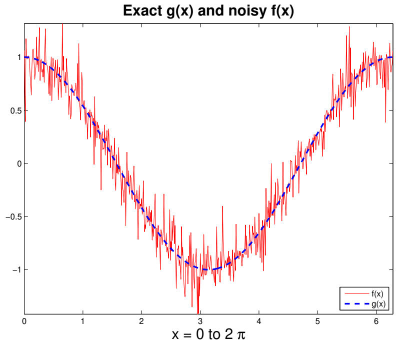


FIG. 6.2. *Exact $g(x) = \cos(x)$ and noisy $f(x)$, corrupted by Laplace noise with mean zero, $\sigma = 0.1$ and $\delta \approx 0.1037$.*

and the source condition becomes

$$(\mathrm{SC}H^1) \qquad \begin{aligned} q(x) &= -\Delta g(x) + g(x)\,, \quad \text{for } x \in \Omega \text{ and } q \in H^1(\Omega), \\ \frac{\partial q}{\partial n} &= 0\,, \quad \text{for } x \in \partial\Omega. \end{aligned}$$

In the following we present two examples and their related error estimates.

EXAMPLE 6.1. For our first data example we choose $g(x) = \cos(x)$, for $x \in [0, 2\pi]$. Since $g \in C^\infty([0, 2\pi])$ and $g'(0) = g'(2\pi) = 0$, the source condition (SCH^1) is fulfilled. Hence, the derived error estimates in Section 4 should work.

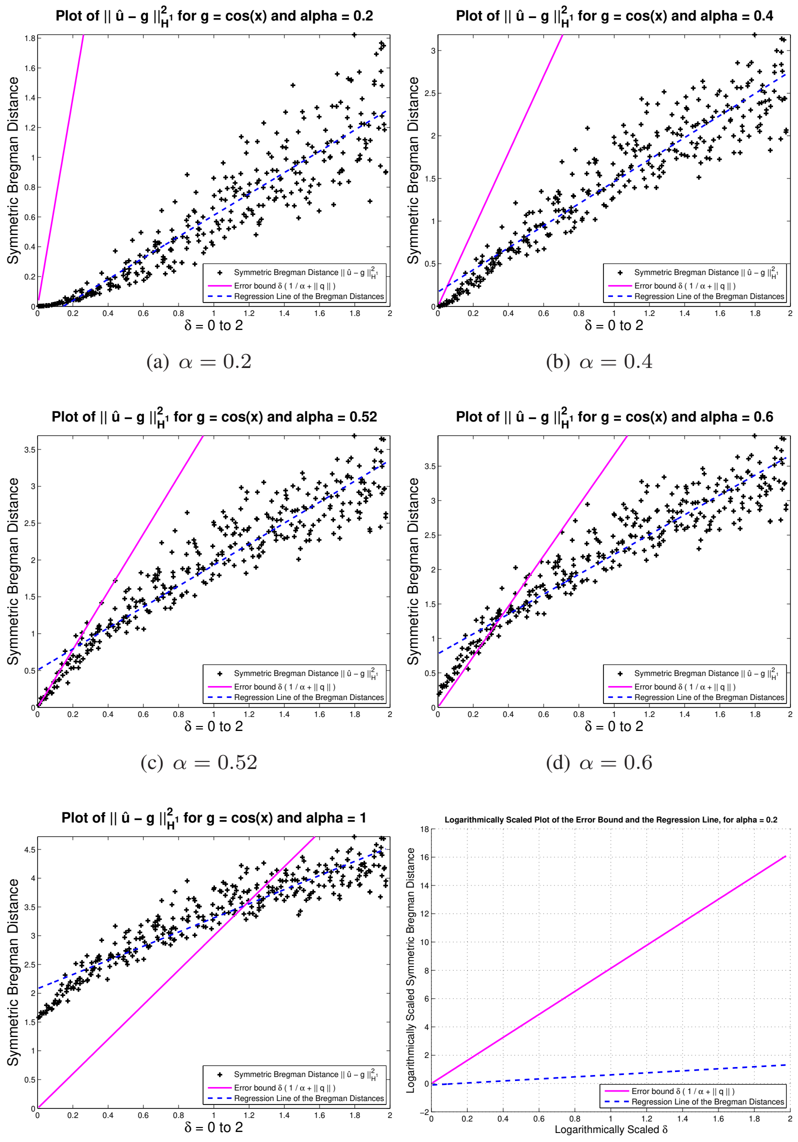
First of all we check (4.1) and (4.2) numerically for noise-free data, i.e., $f = g$ and $\delta = 0$. The estimates predict that as soon as $\alpha \leq \frac{1}{2}$ (note that $\|q\|_{L^\infty([0,2\pi])} = 2\,\|\cos(x)\|_{L^\infty([0,2\pi])} = 2$) holds, the regularized solution $\hat{u}$ should be identical to the exact solution g in the Bregman distance setting (6.2). This is also found in computational practice, as Figure 6.1 confirms.

In the following we want to illustrate the sharpness of (4.2) in the case of non-zero δ. For this reason, we have generated Laplace-distributed random variables and have added them to g to obtain f. We have generated random variables with different values for the variance of the Laplace distribution to obtain different noise levels δ in the L^1-measure. Figure 6.2 shows g and an exemplary noisy version of g with $\delta \approx 0.1037$. In the following, we computed δ as the L^1-norm over $[0, 2\pi]$, to adjust the dimension of δ to the H^1-norm (in the above example δ then approximately becomes $\delta \approx 0.6$).

In order to validate (4.2) we produced many noisy functions f with different noise levels δ in the range of 0 to 2. For five fixed α values ($\alpha = 0.2$, $\alpha = 0.4$, $\alpha = 0.52$, $\alpha = 0.6$, and $\alpha = 1$) we have plotted the symmetric Bregman distances between the regularized solutions $\hat{u}$ and g, the regression line of these distances and the error bound given via (4.2); the results can be seen in Figure 6.3. It can be observed that for $\alpha = 0.2$ and $\alpha = 0.4$ the computed Bregman distances lie beneath that bound, while for $\alpha = 0.52$, $\alpha = 0.6$ and $\alpha = 1$ the error bound is violated, which seems to be a good indicator of the sharpness of (4.2).

EXAMPLE 6.2. In order to validate the need for the source condition (SCH^1) we want to consider two more examples; $g_1(x) = \sin(x)$ and $g_2(x) = |x - \pi|$, $x \in [0, 2\pi]$. Both functions violate (SCH^1); g_1 does not fulfill the Neumann boundary conditions, while the second derivative of g_2 is a δ-distribution centered at $\pi/2$ and therefore not integrable. In the case of g_2 there does not exist a q such that there could exist an α to guarantee (4.2). Nevertheless, in order to visualize that there exists no such error bound, we want to introduce

ETNA
Kent State University
http://etna.math.kent.edu



(a) $\alpha = 0.2$ (b) $\alpha = 0.4$

(c) $\alpha = 0.52$ (d) $\alpha = 0.6$

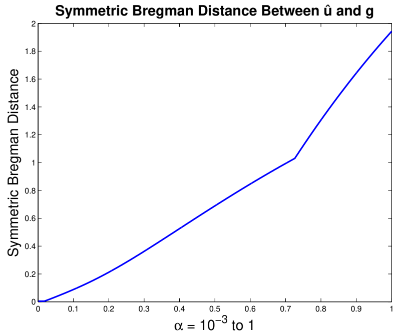
(e) $\alpha = 1$ (f) $\alpha = 0.2$

FIG. 6.3. *The plots of computed symmetric Bregman distances for* $\alpha = 0.2,\ 0.4,\ 0.52,\ 0.6$ *and* $\alpha = 1$*, against* $\delta = 0$ *to* $\delta = 2$*. It can be seen that in (a) and (b) the computed Bregman distances lie beneath the error bound derived in* (4.2)*, while the distances in (c), (d), and (e) partly violate this bound. Figure (f) shows the logarithmically scaled error bound in comparison to the logarithmically scaled regression line of the Bregman distances for* $\alpha = 0.2$*. It can be observed, that the slope of the regression line is smaller than the slope of the error bound. Hence, for the particular choice of* $g(x) = \cos(x)$ *there might exist an even stronger error bound than* (4.2)*.*

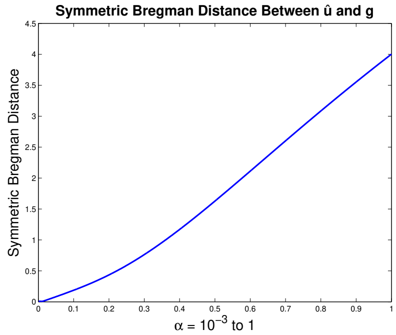
a reference bound $\delta\left(1/\alpha + \|w\|_{L^\infty([0,2\pi])}\right)$ with $w(x) := -\Delta g_2(x) + g_2(x)$, $x \in ([0,\pi[) \cup (]\pi, 2\pi])$, which yields $\|w\|_{L^\infty([0,2\pi])} = \pi$.

As in Example 6.1 we want to begin with the case of exact data, i.e., $f = g$. If we plot the symmetric Bregman distance against α, then we obtain the graphs displayed in Figure 6.4. It can be seen that for g_1 as well as for g_2 the error tends to be zero only if α gets very small. To

ETNA
Kent State University
http://etna.math.kent.edu



(a) $D^{\text{symm}}_{H^1}(\hat{u}, g_1)$ vs α



(b) $D^{\text{symm}}_{H^1}(\hat{u}, g_2)$ vs α

FIG. 6.4. *The symmetric Bregman distances $D^{symm}_{H^1}(\hat{u}, g_1)$ (a) and $D^{symm}_{H^1}(\hat{u}, g_2)$ (b), for $\alpha \in [10^{-3}, 1]$.*

illustrate the importance of the source condition in the noisy case with non-zero δ, we have proceeded as in Example 6.1. We generated Laplace-type noise and added it to g_1 and g_2 to obtain f_1 and f_2 for different error values δ. Figure 6.5 shows the Bregman distance error in comparison to the error bound given via (4.2) and in comparison to the reference bound as described above, respectively. It can be seen that in comparison to Example 6.1 the error and reference bounds are completely violated, even for small α. Furthermore, in the worst case of g_2 for $\alpha = 1$ the slope of the logarithmically scaled regression line is equal to the slope of the reference bound, which indicates that the error assumingly will in general never get beyond this reference bounds. The results support the need for the source condition to find quantitave error estimates.

6.2. Compressive sensing with Poisson noise. For the validation of the error estimates in the Poisson case we consider the following discrete inverse problem. Given a two dimensional function u at discrete sampling points, i.e., $u = (u_{i,j})_{i=1,\dots,n,\ j=1,\dots,m}$, we investigate the operator $\overline{K} : \ell^1(\Omega) \to \ell^1(\Sigma)$ and the operator equation

$$\overline{K}u = g$$

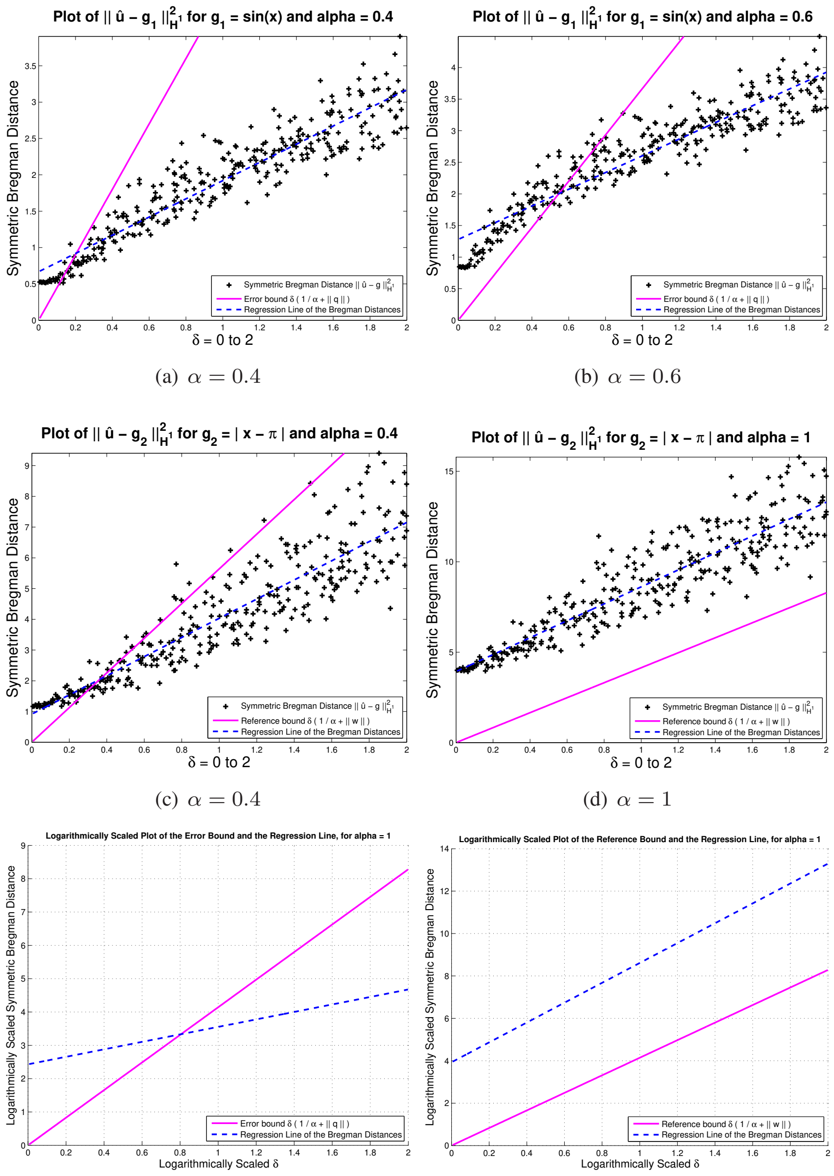
with

$$g_{i,j} = \sum_{k=1}^{n} \phi_{i,k} u_{k,j}\ , \qquad \text{for } i = 1, \dots, l,\ j = 1, \dots, m\,,$$

where $\phi_{i,j} \in [0, 1]$ are uniformly distributed random numbers, $\Omega = \{1, \dots, n\} \times \{1, \dots, m\}$, $\Sigma = \{1, \dots, l\} \times \{1, \dots, m\}$ and $l >> n$, such that $\overline{K}$ has a large nullspace. Furthermore, we consider f instead of g, with f being corrupted by Poisson noise.

6.2.1. Sparsity regularization. In the case of sparsity regularization, we assume u to have a sparse representation with respect to a certain basis. Therefore, we consider an operator $B : \ell^1(\Theta) \to \ell^1(\Omega)$ such that $u = Bc$ holds for coefficients $c \in \ell^1(\Theta)$. If we want to apply a regularization that minimizes the ℓ^1-norm of the coefficients, we obtain the minimization problem

$$\int_\Sigma f \ln\left(\frac{f}{\overline{K}Bc}\right) + \overline{K}Bc - f\, d\mu(y) + \alpha \sum_{i,j} |c^{i,j}|_{\ell^1(\Theta)} \to \min_{c \in \ell^1(\Theta)}\,. \tag{6.3}$$



(a) $\alpha = 0.4$ (b) $\alpha = 0.6$

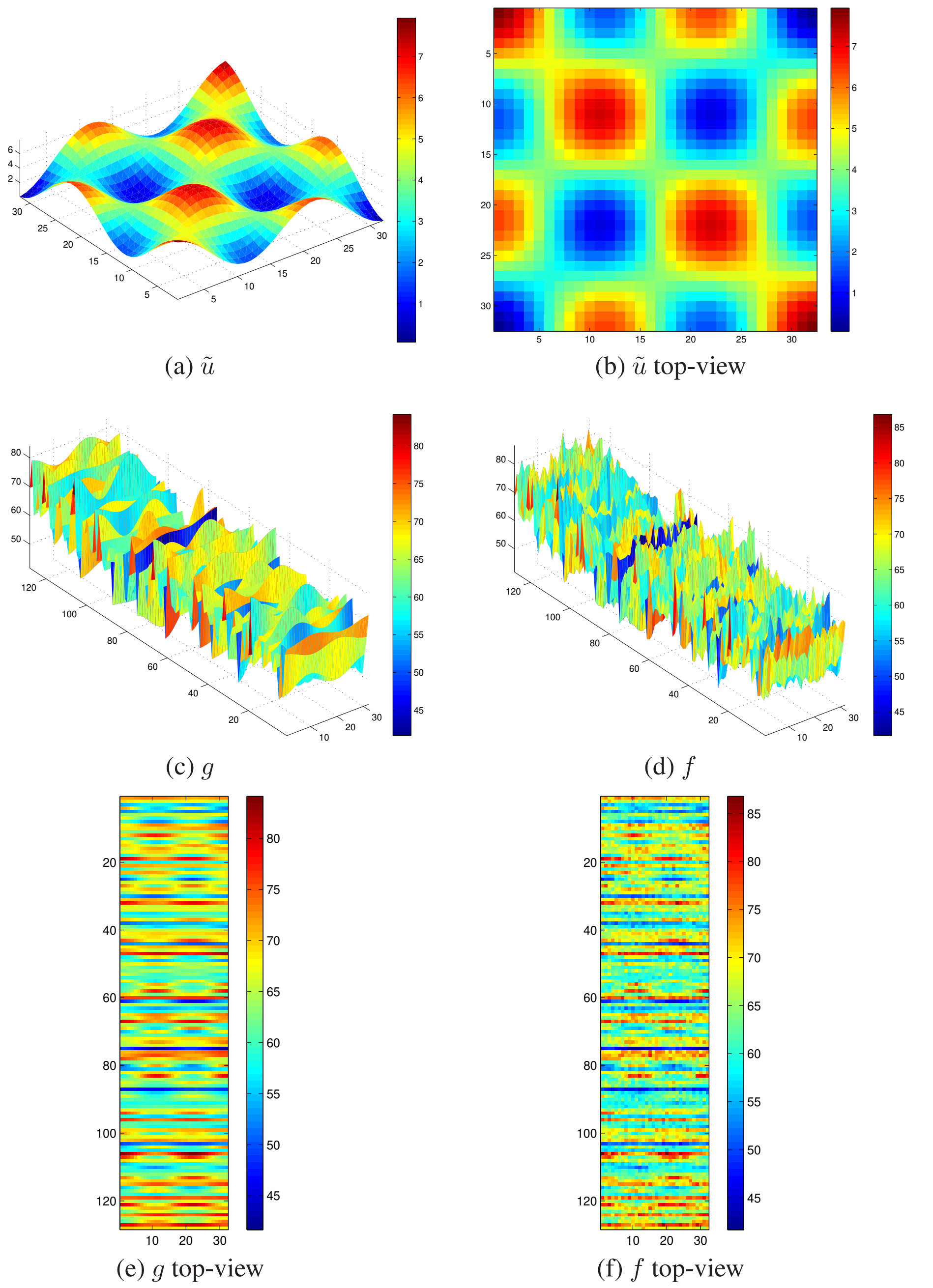
(c) $\alpha = 0.4$ (d) $\alpha = 1$

(e) $\alpha = 1$ (f) $\alpha = 1$

FIG. 6.5. *The plots of the computed Bregman distances with violated* (SCH^1). *Figures (a) and (b) show the Bregman distances* $D^{symm}_{H^1}(\hat{u}, g_1)$ *for* $\alpha = 0.4$ *and* $\alpha = 0.6$*, respectively. Figures (c) and (d) represent the Bregman distances* $D^{symm}_{H^1}(\hat{u}, g_2)$ *for* $\alpha = 0.4$ *and* $\alpha = 1$*. Furthermore, Figures (e) and (f) show the logarithmically scaled versions of the error/reference bound in comparison to a line regression of the Bregman distances for* $\alpha = 1$.

Notice that the measure μ is a point measure and the integral in (6.3) is indeed a sum over all discrete samples, which we write as an integral in order to keep the notation as introduced in the previous sections. To perform a numerical computation, we use a forward-backward splitting algorithm based on a gradient descent of the optimality condition. For an initial set

ETNA
Kent State University
http://etna.math.kent.edu



(a) $\tilde{u}$ (b) $\tilde{u}$ top-view

(c) g (d) f

(e) g top-view (f) f top-view

FIG. 6.6. *In (a) we can see $\tilde{u}$ as defined via $\tilde{u} = B\tilde{c}$. A top-view of $\tilde{u}$ can be seen in (b). In (c) and (d) the plots of $g = \overline{K}\tilde{u}$ and its noisy version f are described, while (e) and (f) show top-view illustrations of (c) and (d).*

of $|\Theta|$ coefficients $c_0^{i,j}$, we compute the iterates

$$c_{k+\frac{1}{2}}^{i,j} = c_k^{i,j} - \frac{1}{\tau}\left(B^T\overline{K}^T\left(1 - \frac{f}{\overline{K}Bc_k}\right)\right)_{i,j},$$

$$c_{k+1}^{i,j} = \operatorname{sign}\left(c_{k+\frac{1}{2}}^{i,j}\right)\left(\left|c_{k+\frac{1}{2}}^{i,j}\right| - \frac{\alpha}{\tau}\right)_+,$$

with $\tau > 0$ being a damping parameter and $a_+ = \max\{0, a\}$. The disadvantage of this approach is that positivity of u is not guaranteed. However, it is easy to implement and produces satisfactory results as we will see in the following.

As a computational example, we choose u to be sparse with respect to the cosine basis. We therefore define $B^{-1} : \ell^1(\Theta) \to \ell^1(\Omega)$ as the two-dimensional cosine-transform,

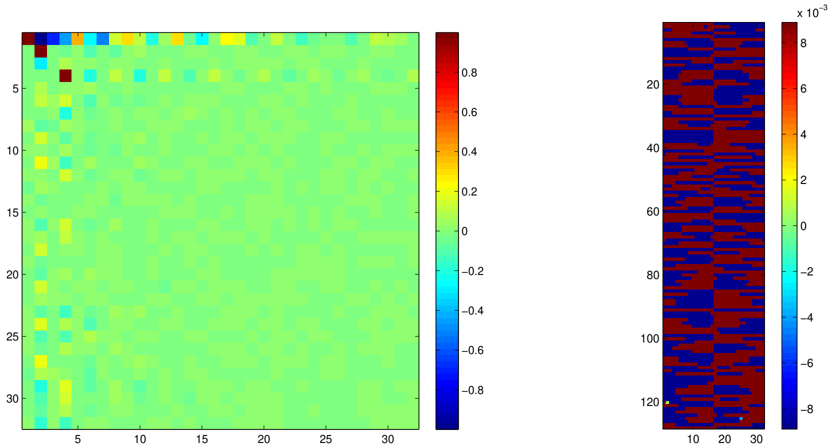
$$c^{a,b} = \gamma(a)\gamma(b) \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} u_{i,j} \cos\left(\frac{\pi(2i+1)a}{2n}\right) \cos\left(\frac{\pi(2j+1)b}{2m}\right), \tag{6.4}$$

for $a \in \{1, \dots, n\}$, $b \in \{1, \dots, m\}$ and

$$\gamma(x) = \begin{cases} \sqrt{1/n} & \text{for } x = 0 \\ \sqrt{2/n} & \text{for } x \neq 0 \end{cases}.$$

Remember that since the cosine-transform defined as in (6.4) is orthonormal we have $B = (B^{-1})^T$.

We set $\tilde{u} = B\tilde{c}$ with $\tilde{c}$ being zero except for $\tilde{c}^{1,1} = 4\sqrt{nm}$, $\tilde{c}^{2,2} = 1/2\sqrt{nm}$, and $\tilde{c}^{4,4} = 3/2\sqrt{nm}$. With this choice of $\tilde{c}$ we guarantee $\tilde{u} > 0$. Furthermore, we obtain $g = \overline{K}\tilde{u}$ with $g > 0$. Finally, we generate a noisy version f of g by replacing every sample $g_{i,j}$ with a Poisson random number $f_{i,j}$ with expected value $g_{i,j}$. As an example we chose $n = m = 32$.



(a) ξ (b) q

FIG. 6.7. *The computed subgradient ξ (a) and the function q (b), which are related to each other by the source condition (SCℓ^1). The computed q has minimum ℓ^∞-norm among all q's satisfying (SCℓ^1); $\|q\|_{\ell^\infty} \approx 8.9 \times 10^{-3}$.*

Hence we obtain $\tilde{c}^{1,1} = 128$, $\tilde{c}^{2,2} = 16$, and $\tilde{c}^{4,4} = 48$, while the other coefficients remain zero. Furthermore, we obtain $\tilde{u}$, g and f as described in Figure 6.6. The operator dimension is chosen to be $l = 128$. The damping parameter is set to the constant value $\tau = 0.1225$.

The source condition for this discrete data example becomes

$$\exists\, \xi \in \partial|\tilde{c}|_{\ell^1(\Omega)},\ \exists\, q \in (\ell^1(\Sigma))^* : \quad \xi = B^T K^T q. \tag{SCℓ^1}$$

It can be shown that the subgradient of $|\tilde{c}|_{\ell^1(\Omega)}$ is simply

$$\partial|\tilde{c}^{i,j}|_{\ell^1(\Omega)} = \operatorname{sign}(\tilde{c}^{i,j}) = \begin{cases} 1 & \text{for } \tilde{c}^{i,j} > 0 \\ \in [-1, 1] & \text{for } \tilde{c}^{i,j} = 0 \\ -1 & \text{for } \tilde{c}^{i,j} < 0 \end{cases}, \tag{6.5}$$

for all $(i, j) \in \Omega$. Hence, to validate the error estimates and their sharpness, we have computed ξ and q with minimal ℓ^∞-norm, in order to satisfy (SCℓ^1) and (6.5). The computational results can be seen in Figure 6.7, where $\|q\|_{\ell^\infty} \approx 8.9 \times 10^{-3}$ holds. With ξ computed, the symmetric Bregman distance easily can be calculated via

$$D^{\text{symm}}_{|\cdot|_{\ell^1}}(\hat{c}, \tilde{c}) = \langle \hat{p} - \xi, \hat{c} - \tilde{c} \rangle_{\ell^1(\Omega)}$$

for $\hat{p} \in \partial |\hat{c}|_{\ell^1}$. We did several computations with different values for α and the constant $c \in]0, 1[$ (not to be confused with the cosine transform coefficients) to support the error estimate (4.3). The results can be seen in Table 6.1. We note that Theorem 4.3 can only be applied if $\int_\Sigma f \; d\mu(y) = 1$, which is obviously not the case for our numerical example. But, due to the proof of Theorem 4.3, the only modification that has to be made is to multiply $|f| = \int_\Sigma f \; d\mu(y)$ by the logarithmic term in (4.3) to obtain a valid estimate.

7. Outlook and open questions. We have seen that under rather natural source conditions error estimates in Bregman distances can be extended from the well-known quadratic fitting (Gaussian noise) case to general convex fidelities. We have seen that the appropriate definition of noise level in the convex case is not directly related to the norm difference, but rather to the data fidelity. With this definition the estimates indeed yield the same asymptotic order with respect to regularization parameter and noise level as in the quadratic case. The constants are again related to the smoothness of the solution (norm of the source element), with the technique used in the general case one obtains slightly larger constants than in the original estimates. The latter is caused by the fact that the general approach to error estimation cannot exploit linearity present in the case of quadratic fidelity

Error estimation is also important for other than variational approaches, in particular iterative or flow methods such as scale space methods, inverse scale space methods or Bregman iterations. The derivation of such estimates will need a further understanding of dual iterations or flows, which are simple gradient flows in the case of quadratic fidelity but have a much more complicated structure in general.

Another interesting question for future research, which is also of practical importance, is an understanding of error estimation in a stochastic framework. It will be an interesting task to further quantify uncertainty or provide a comprehensive stochastic framework. In the case of Gaussian noise such steps have been made, e.g., by lifting pointwise estimates to estimates for the distributions in different metrics (cf. [14, 15, 16]) or direct statistical approaches; cf. [3]. In the case of Poisson noise a direct estimation of mean deviations has been developed in parallel in [18], which uses similar techniques as our estimation and also includes a novel statistical characterization of noise level in terms of measurement times. We think that our approach of estimating pointwise errors via fidelities, i.e., data likelihoods (more precisely negative log likelihoods), will have an enormous potential to generalize to a statistical setting. In particular, we expect the derivation of confidence regions by further studies of distributions of the log-likelihood for different noise models.

Acknowledgments. This work has been supported by the German Research Foundation DFG through the project *Regularization with singular energies* and *SFB 656: Molecular cardiovascular imaging*. The second author acknowledges further support by the German ministry for Science and Education (BMBF) through the project INVERS: Deconvolution with sparsity constraints in nanoscopy.

The authors thank Dirk Lorenz (University Bremen), Nico Bissantz (University Bochum), Thorsten Hohage (University Göttingen), and Christoph Brune (University Münster) for useful and stimulating discussions. Furthermore, the authors thank the referees for valuable suggestions.

TABLE 6.1
Computational Results. The first two columns represent different values for α and c. The remaining numerated columns denote: (a) $(1-c)H_f(\hat{c})$, (b) $\alpha\, \mathrm{D}^{\mathrm{symm}}_{|\cdot|_{\ell^1}}(\hat{c},\tilde{c})$, (c) $(1+c)\delta$, (d) $-c\ln\left(1-\left(\frac{\alpha}{c}\,\|q\|_{\ell^\infty}\right)^2\right)$, (e) $(1+c)\delta-(1-c)H_f(\hat{c})-c\ln\left(1-\left(\frac{\alpha}{c}\,\|q\|_{\ell^\infty}\right)^2\right)$. It can be seen that column (e) is always larger than (b) and hence, the error estimate (4.3) is always fulfilled.

c	α	(a)	(b)	(c)	(d)	(e)
0.5	0.05	85.25	0.8	178.7	0.1049	93.56
0.5	0.009	60.32	0.004619	178.7	0.0034	118.4
0.5	0.005	59.48	0.005379	178.7	0.001049	119.2
0.5	0.003	58.39	0.02627	178.7	0.0003777	120.3
0.5	0.0009	49.34	0.1651	178.7	3.4e-005	129.4
0.5	0.0005	46.69	0.1501	178.7	1.049e-005	132
0.5	0.0003	45.68	0.1142	178.7	3.777e-006	133
0.5	5e-005	50.72	0.02586	178.7	1.049e-007	128
0.1	0.009	108.6	0.004619	131	0.017	22.48
0.1	0.005	107.1	0.005379	131	0.005246	23.99
0.1	0.003	105.1	0.02627	131	0.001889	25.95
0.1	0.0009	88.8	0.1651	131	0.00017	42.24
0.1	0.0005	84.04	0.1501	131	5.246e-005	47.01
0.1	0.0003	82.23	0.1142	131	1.889e-005	48.82
0.1	5e-005	91.29	0.02586	131	5.246e-007	39.76
0.01	0.009	119.4	0.004619	120.3	0.17	1.054
0.01	0.005	117.8	0.005379	120.3	0.05246	2.611
0.01	0.003	115.6	0.02627	120.3	0.01889	4.732
0.01	0.0009	97.68	0.1651	120.3	0.0017	22.64
0.01	0.0005	92.45	0.1501	120.3	0.0005246	27.88
0.01	0.0003	90.45	0.1142	120.3	0.0001889	29.87
0.01	5e-005	100.4	0.02586	120.3	5.246e-006	19.91
0.001	0.05	170.3	0.8	119.3	58.45	7.382
0.001	0.009	120.5	0.004619	119.3	1.705	0.4317
0.001	0.005	118.8	0.005379	119.3	0.5252	0.9405
0.001	0.003	116.7	0.02627	119.3	0.1889	2.779
0.001	0.0009	98.57	0.1651	119.3	0.017	20.7
0.001	0.0005	93.29	0.1501	119.3	0.005246	25.97
0.001	0.0003	91.27	0.1142	119.3	0.001889	27.98
0.001	5e-005	101.3	0.02586	119.3	5.246e-005	17.92
0.0001	0.009	120.6	0.004619	119.1	27.14	25.65
0.0001	0.005	118.9	0.005379	119.1	5.845	6.046
0.0001	0.003	116.8	0.02627	119.1	1.959	4.337
0.0001	0.0009	98.66	0.1651	119.1	0.1705	20.66
0.0001	0.0005	93.37	0.1501	119.1	0.05252	25.83
0.0001	0.0003	91.36	0.1142	119.1	0.01889	27.81
0.0001	5e-005	101.4	0.02586	119.1	0.0005246	17.72
1e-005	0.0009	98.67	0.1651	119.1	2.714	23.18
1e-005	0.0005	93.38	0.1501	119.1	0.5845	26.34
1e-005	0.0003	91.36	0.1142	119.1	0.1959	27.97
1e-005	5e-005	101.4	0.02586	119.1	0.005252	17.71
1e-006	5e-005	101.4	0.02586	119.1	0.05845	17.76

REFERENCES

[1] G. AUBERT AND J. F. AUJOL, *A variational approach to remove multiplicative noise*, SIAM J. Appl. Math., 68 (2008), pp. 925–946.
[2] J. BARDSLEY AND A. LUTTMAN, *Total variation-penalized Poisson likelihood estimation for ill-posed prob-*

lems , Adv. Comput. Math., 31 (2009), pp. 35–59.

[3] N. BISSANTZ, T. HOHAGE, A. MUNK, AND F. RUYMGAART, *Convergence rates of general regularization methods for statistical inverse problems and applications*, SIAM J. Numer. Anal., 45 (2007), pp. 2610–2636.

[4] 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*, Comput. Math. Math. Phys., 7 (1967), pp. 200–217.

[5] C. BRUNE, A. SAWATZKY, AND M. BURGER, *Bregman-EM-TV methods with application to optical nanoscopy*, Proceedings of the 2nd International Conference on Scale Space and Variational Methods in Computer Vision, X.-C. Tai, K. Mørken, M. Lysaker, and K.-A. Lie, eds., Lecture Notes in Comput. Sci., Vol. 5567, Springer, Berlin, 2009, pp. 235–246.

[6] M. BURGER, G. GILBOA, S. OSHER, AND J. XU, *Nonlinear inverse scale space methods*, Commun. Math. Sci., 4 (2006), pp. 179–212.

[7] M. BURGER, L. HE, AND C. SCHÖNLIEB, *Cahn-Hilliard inpainting and a generalization to grayvalue images*, SIAM J. Imaging Sci., 2 (2009), pp. 1129–1167.

[8] M. BURGER AND S. OSHER, *Convergence rates of convex variational regularization*, Inverse Problems, 20 (2004), pp. 1411–1421.

[9] M. BURGER, E. RESMERITA, AND L. HE, *Error estimation for Bregman iterations and inverse scale space methods*, Computing, 81 (2007), pp. 109–135.

[10] R. CHAN, C. HO, AND M. NIKOLOVA, *Salt-and-pepper noise removal by median-type noise detector and detail-preserving regularization*, IEEE Trans. on Image Process., 14 (2005), pp. 1479–1485.

[11] T. F. CHAN AND S. ESEDOGLU, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math., 65 (2005), pp. 1817–1837.

[12] I. EKELAND AND R. TEMAM, *Convex Analysis and Variational Problems*, Corrected Reprint Edition, SIAM, Philadelphia, 1999.

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

[14] H. W. ENGL, A. HOFINGER, AND S. KINDERMANN, *Convergence rates in the Prokhorov metric for assessing uncertainty in ill-posed problems*, Inverse Problems, 21 (2005), pp. 399–412.

[15] A. HOFINGER AND H. K. PIKKARAINEN, *Convergence rates for the Bayesian approach to linear inverse problems*, Inverse Problems, 23 (2006), pp. 2469–2484.

[16] ———, *Convergence rates for linear inverse problems in the presence of an additive normal noise*, Stoch. Anal. Appl., 27 (2009), pp. 240–257.

[17] B. HOFMANN, B. KALTENBACHER, C. PÖSCHL, AND O. SCHERZER, *A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators*, Inverse Problems, 23 (2007), pp. 987–1010.

[18] T. HOHAGE, *Variational regularization of inverse problems with Poisson data*, Preprint, 2009.

[19] Y. M. HUANG, M. K. NG, AND Y. W. WEN, *A new total variation method for multiplicative noise removal*, SIAM J. Imaging Sci., 2 (2009), pp. 20–40.

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

[21] D. A. LORENZ, *Convergence rates and source conditions for Tikhonov regularization with sparsity constraints*, J. Inverse Ill-Posed Probl., 16 (2008), pp. 463–478.

[22] D. A. LORENZ AND D. TREDE, *Optimal convergence rates for Tikhonov regularization in Besov scales*, Inverse Problems, 24 (2008), 055010.

[23] Y. MEYER, *Oscillating patterns in image processing an in some nonlinear evolution equations*, The Fifteenth Dean Jacquelines B. Lewis Memorial Lectures, American Mathematical Society, Providence, 2001.

[24] F. NATTERER AND F. WÜBBELING, *Mathematical Methods in Image Reconstruction*, SIAM, Philadelphia, 2001.

[25] A. NEUBAUER, *On converse and saturation results for Tikhonov regularization of linear ill-posed problems*, SIAM J. Numer. Anal., 34 (1997), pp. 517–527.

[26] S. OSHER, M. BURGER, D. GOLDFARB, J. XU, AND W. YIN, *An iterative regularization method for total variation-based image restoration*, Multiscale Model. Simul., 4 (2005), pp. 460–489.

[27] C. PÖSCHL, *Tikhonov regularization with general residual term*, PhD thesis, Department of Mathematics, Computer Science, and Physics, Leopold-Franzens-Universität Innsbruck, 2008.

[28] E. RESMERITA AND O. SCHERZER, *Error estimates for non-quadratic regularization and the relation to enhancing*, Inverse Problems, 22 (2006), pp. 801–814.

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

[30] J. SHI AND S. OSHER, *A nonlinear inverse scale space method for a convex multiplicative noise model*, SIAM J. Imaging Sci., 1 (2008), pp. 294–321.

[31] C. WU, J. ZHANG, AND X. TAI, *Augmented Lagrangian method for total variation restoration with non-quadratic fidelity*, Inverse Problems and Imaging, 5 (2011), pp. 237–261.