Error Estimates for Variational Models with Non-Gaussian Noise

Martin Benning* 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 such as those arising from Bayesian models with non-Gaussian noise, 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 noise models in imaging (Poisson, Laplacian, Multiplicative) and the corresponding Bayesian MAP estimation.

Key words: Error Estimation, Bregman Distance, Imaging, Image Processing, Laplace Noise, Poisson Noise, Multiplicative Noise, Sparsity.

1 Introduction

Image processing and inversion with structural prior information (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 and compact 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

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

and allows to define stochastic noise models in a meaningful way. In the course of this work we want to call g the *exact data* and $\tilde{u}$ the *exact solution*.

Since most inverse problems are ill-posed (due to the compactness of the forward operator), K cannot be inverted continuously. Furthermore, in real-life applications the exact data g is usually not attainable. 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 in 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 attainable, 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. E.g., in the limiting case of Gaussian-distributed noise(to be detailed below), we have a noise bound in the L_2-measure, i.e.

$$\|g - f\|_{L_2(\Sigma)} \leq \delta\,. \tag{1.3}$$

In this case δ^2 is an upper estimate of the noise variance σ^2.

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

$$\hat{u} \in \operatorname*{arg\,min}_{u \in \mathcal{W}(\Omega)} \left\{ H_f(Ku) + \alpha J(u) \right\} , \tag{1.4}$$

with $H_f : \mathcal{V}(\Sigma) \to \mathbb{R} \cup \{\infty\}$ and $J : \mathcal{W}(\Omega) \to \mathbb{R} \cup \{\infty\}$, $\mathcal{W}(\Omega) \subseteq \mathcal{U}(\Omega)$. The scheme contains the 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 literature, schemes based on (1.4) are often referred to as *variational regularization schemes*. Throughout this paper we shall assume that J is chosen such that a minimizer of (1.4) exists, the proof of which is not an easy task for many important choices of H_f (cf. e.g. [AA08, BL09]). Notice that if $H_f(Ku) + \alpha J(u)$ in (1.4) 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. [EHN96]) 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 [ROF92]) 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 [EHN96]). A major step for error estimates in the case of regularization with singular energies has been the introduction of (generalized) Bregman distances (cf. [B67, K97]) as an error measure (cf. [BO04]). The Bregman distance for general convex, not necessarily differentiable functionals, is defined as follows.

Definition 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*

$$D_J(u, v) := \{ J(u) - J(v) - \langle p, u - v \rangle_{\mathcal{U}} \mid p \in \partial J(v) \} \ . \tag{1.5}$$

The Bregman distance for a specific subgradient ζ is defined as $D_J^\zeta : \mathcal{U} \times \mathcal{U} \to \mathbb{R}^+$ with

$$D_J^\zeta(u, v) := J(u) - J(v) - \langle \zeta, u - v \rangle_{\mathcal{U}} \ , \quad \zeta \in \partial J(v) \ . \tag{1.6}$$

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 $D_J(u, u) = 0$ and $D_J(u, v) \geq 0$ hold, the latter due to convexity of J. If J is strictly convex, we even obtain $D_J(u, v) > 0$ for $u \neq v$. 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 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 $D_J^{symm} : \mathcal{U} \times \mathcal{U} \to \mathbb{R}^+$ with*

$$D_J^{symm}(u_1, u_2) := D_J^{p_1}(u_2, u_1) + D_J^{p_2}(u_1, u_2) = \langle u_1 - u_2, p_1 - p_2 \rangle_{\mathcal{U}^*} \ , \tag{1.7}$$

with

$$p_i \in \partial J(u_i) \quad for \quad i \in \{1, 2\} \ . \tag{1.8}$$

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 yet 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. [BRH07, BSH08, HKP$^+$07, L08, LT08, RS06]). 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

$$\exists \, \xi \in \partial J(\tilde{u}), \exists \, q \in L^2(\Sigma) \ : \ \xi = K^* q \, . \tag{1.9}$$

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. [BGO$^+$06, OBG$^+$05, BSB09]).

Most works deal with the case of quadratic fitting related to Gaussian distributed noise so far, with only few exceptions (see e.g. [P08]). 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. For such cases different variational models (fidelities related to the log-likelihood of the noise distribution and its asymptotic) can be derived in the framework of MAP estimation, which need different analysis as in the case of Gaussian distributed noise.

In the following we want to derive some non-Gaussian noise models as recently used in various imaging applications. Then, we are going to present basic error estimates for general, convex variational regularization methods, which we will apply to the specific noise models. Subsequently we are going to support these estimates and test their sharpness by computational results. Concludingly, we will give a brief outlook and will formulate open questions. We would also like to mention the parallel development on error estimates for variational models with non-quadratic fidelity in [], which yields the same results as our paper in the case of Laplacian noise. Since the analysis in [] is rather 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-Gaussian Noise Models

In this section we want to discuss a statistical approach to formulate variational regularization methods based on Bayes' formula and MAP estimation. This approach will allow to take into account of how the data f are random variables arising from a specific probability law. With this approach we want to present three typical, non-Gaussian noise models: the Laplace model, the Poisson model and a multiplicative noise model based on the Gamma distribution.

Stochastic noise modelling is done reasonably via random variables in finite-dimensional spaces, where the dimension of the space corresponds to a number of detectors or pixels. In some of the noise models like the Gaussian one (usually arising as a limit distribution) the averaging on a pixel or detector of finite size is inherent. In order make the potential dependence of the noise model on the dimension m more concrete we introduduce linear and compact operator

$$K^m : \mathcal{U}(\Omega) \to \mathcal{V}^m \,, \tag{2.1}$$

with a finite m-dimensional range $\mathcal{V}^m$. K^m is usually the concatenation of K with a sampling or local averaging. In order to derive a continuum model, we will - if the situations require it - start with the semi-discrete setting to derive a maximum a-posteriori probability density based on Bayes' formula and will then pass over to the continuum case in order to obtain a variational model in this setting, too.

2.1 MAP Estimation

A very common approach to derive a variational regularization method is to maximize the a-posteriori probability density $p(u|f)$, which is known as the maximum a-posteriori probability (MAP) estimation. Due to Bayes' formula, we have

$$p(u|f) = \frac{p(f|u)p(u)}{p(f)} \,, \tag{2.2}$$

where $p(u|f)$ is the a-posteriori likelihood function. The probability density $p(u)$ is called the prior probability density, which provides a-priori information on u, e.g. the function space that u lies in. Most frequently the Gibbs model is used to model the prior probability, i.e.

$$p(u) = e^{-\alpha J(u)} \,, \tag{2.3}$$

for a positive constant $\alpha > 0$ and a convex regularization energy $J : \mathcal{W}(\Omega) \to \mathbb{R} \cup \{\infty\}$. The conditional probability $p(f|u)$ of the data f with given image u is the part of the Bayesian model that is influenced by the type of distribution of the noise (and hence the noisy data f). If, e.g. the data f are Gaussian distributed (with expectation g), more precisely each sample of $\frac{f-g}{\sigma}$ is i.i.d. standard normal, the conditional probability $p(f|u)$ can be described as

$$p(f|u) \sim e^{-\frac{\|K^m u - f\|_2^2}{2\sigma^2}} \,, \tag{2.4}$$

with σ^2 being the variance of the data f. Minimizing the negative log-likelihood of the Bayes model, i.e. computing $u \in \arg\min_{u \in \mathcal{U}(\Omega)} \left(-\ln(p(u|f))\right)$, yields

$$\hat{u} = \underset{u \in \mathcal{W}(\Omega)}{\arg\min} \left\{ \frac{1}{2} \|K^m u - f\|_2^2 + \alpha J(u) \right\} \,, \tag{2.5}$$

which is a standard model used for convex regularization. As $m \to \infty$, a natural infinite-dimensional asymptotic of the variational model is usually studied,

$$\hat{u} = \underset{u \in \mathcal{W}(\Omega)}{\arg\min} \left\{ \frac{1}{2} \|Ku - f\|_{\mathcal{V}(\Sigma)}^2 + \alpha J(u) \right\} \,, \tag{2.6}$$

However, many types of application provide non-Gaussian-distributed data, which result in different models for $p(f|u)$ and subsequently for $p(u|f)$, and we shall discuss few important ones in the following.

2.2 Laplace Noise

A simple model for Laplace noise is an additive noise model in the discrete setting, i.e., for each pixel $i = 1, \dots, m$ the measured value is distorted by independent identically Laplace distributed random variables. This means if we simply identify $\mathcal{V}^m$ with $\mathbb{R}^m$, then

$$f_i = g_i + n_i \quad i = 1, \dots, m, \tag{2.7}$$

with n_i being a Laplace distributed random variable with mean zero and variance $2\sigma^2$. Hence, the probability density $p(f|u)$ is given by

$$p(f|u) = (2\sigma)^{-m} \exp\left(- \sum_{i=1}^{m} \frac{|f_i - (K^m u)_i|}{\sigma} \right). \tag{2.8}$$

After appropriate rescaling the log-likelihood is given by

$$LL_m(u) = \frac{1}{m} \sum_{i=1}^{m} |f_i - (K^m u)_i|, \tag{2.9}$$

and in the limit $m \to \infty$ this sum converges to

$$LL(u) = \int_\Sigma |(Ku)(y) - f(y)| \; d\mu(y), \tag{2.10}$$

In order to understand the noise level in the continuum limit we consider $LL_m(g_i)$ as a random variable depending on n_i. Due to the above assumptions, the variables $|n_i| = |f_i - g_i|$ are i.i.d. with mean σ and variance σ^2. It is straight-forward to see that there holds a law of large numbers for the log-likelihood, $LL_m(g)$ becomes a random variable with mean σ and variance $\frac{\sigma^2}{m}$. Thus, the limit $LL(g)$ becomes deterministic and equals σ. If an upper estimate δ of σ is available, it is thus reasonable to assume the noise bound

$$LL(g) \leq \delta \tag{2.11}$$

as we shall do in the following.

If we use the additional prior probability according to (2.2) together with the negative loglikelihood (2.9) respectively (2.10), we obtain the asymptotic variational problem

$$\hat{u} = \arg\min_{u \in \mathcal{W}(\Omega)} \left\{ \int_\Sigma |(Ku)(y) - f(y)| \; d\mu(y) + \alpha J(u) \right\}. \tag{2.12}$$

The optimality condition of (2.12) can easily be computed as

$$K^* \hat{s} + \alpha \hat{p} = 0, \qquad \hat{s} \in \text{sign}(K\hat{u} - f), \quad \hat{p} \in \partial J(\hat{u})\,. \tag{2.13}$$

2.3 Poisson Noise

In the case of Poisson distributed data the conditional probability $p(f|u)$ is modelled in discrete terms as

$$p(f|u) = \prod_{i=1}^{m} \frac{(K^m u)_i^{f_i}}{f_i!} e^{-(K^m u)_i}\,, \tag{2.14}$$

where $K^m : \mathcal{U}(\Omega) \to \mathcal{V}^m$ is the semi-discrete, linear and compact operator introduced in (2.1) given at m samples (which represent random variables in the modelling). If we consider the negative log-likelihood $-\ln(p(f|u))$, we obtain

$$-\ln(p(f|u)) = -\left(\sum_{i=1}^{m} f_i \ln(K^m u)_i - \ln(f_i!) - (K^m u)_i\right). \tag{2.15}$$

If we furthermore consider the limit $m \to \infty$ and neglect the $f_i!$ terms (which are independent of u and hence will not matter for the solution of the variational model) we obtain the asymptotic log-likelihood

$$LL(u) = \int_\Sigma [(Ku)(y) - f(y)\ln((Ku)(y))] \; d\mu(y). \tag{2.16}$$

If we add terms in f, independent of u and therefore not affecting the optimality condition, to force the functional to be zero if $Ku = f$, this yields

$$H_f(Ku) = \int_\Sigma \left[f(y)\ln\left(\frac{f(y)}{(Ku)(y)}\right) - f(y) + (Ku)(y)\right] \; d\mu(y), \tag{2.17}$$

which is actually known as the Kullback-Leibler functional. Hence, the negative loglikelihood minimization of $p(u|f)$ yields the optimization problem

$$\hat{u} = \arg\min_{u \in \mathcal{W}(\Sigma)} \left\{ \int_\Sigma \left[f(y)\ln\left(\frac{f(y)}{(Ku)(y)}\right) - f(y) + (Ku)(y)\right] \; d\mu(y) + \alpha J(u) \right\}. \tag{2.18}$$

With the natural scaling assumption

$$K^* \mathbf{1} = \mathbf{1}, \tag{2.19}$$

we obtain the complementarity condition

$$\begin{aligned} u \geq 0, \qquad K^* \frac{f}{Ku} - \alpha p \leq 1, \\ \hat{u}\left(1 - K^* \frac{f}{K\hat{u}} + \alpha \hat{p}\right) = 0, \qquad \hat{p} \in \partial J(\hat{u}). \end{aligned} \tag{2.20}$$

2.4 Multiplicative Noise

In this section we want to consider a multiplicative noise model $f = (Ku)\,v$, where v is the noise and $Ku \geq 0$ is assumed. Again, we start with a discrete setting and work with the semi-discrete operator K^m introduced in (2.1) instead of K. Furthermore, we suppose that each f_i, $i \in \{1, \dots, m\}$, is the mean over n measurements. In [AA08], Aubert and Aujol assumed v to follow a gamma law with mean one and derived the conditional probability $p(f|u)$ with

$$p(f|u) = \prod_i^m \frac{n^n}{(K^m u)_i^n \, \Gamma(n)} f_i^{n-1} e^{-n \frac{f_i}{(K^m u)_i}}, \tag{2.21}$$

for $(K^m u)$ being positive. Hence, due to (2.2) we obtain

$$
\begin{aligned}
-\ln\left(p(u|f)\right) = &\sum_{i=1}^{m} n\left(\ln\left((K^m u)_i\right) + \frac{f_i}{(K^m u)_i} - \ln(n)\right) \\
&+ \ln\left(\Gamma(n)\right) - (n-1)\ln\left(f_i\right) \, .
\end{aligned}
\tag{2.22}
$$

If we neglect terms independent of u as well as the constant n and consider the continuum case $m \to \infty$ we obtain

$$
LL(u) = \int_{\Sigma} \ln\left(\left(Ku\right)(y)\right) + \frac{f(y)}{\left(Ku\right)(y)} d\mu(y) \, , \tag{2.23}
$$

Adding up terms independent of u to guarantee that the integral in (2.23) is zero if $Ku = f$ yields

$$
H_f(Ku) = \int_{\Sigma} \left[\ln\left(\frac{(Ku)(y)}{f(y)}\right) + \frac{f(y)}{(Ku)(y)} - 1\right] d\mu(y) \, . \tag{2.24}
$$

Minimizing the negative loglikelihood $p(u|f)$ due to (2.2) respectively its asymptotic yields the variational problem

$$
\hat{u} \in \arg\min_{u \in \mathcal{W}(\Omega)} \left\{ \int_{\Sigma} \left[\ln\left(\frac{(Ku)(y)}{f(y)}\right) + \frac{f(y)}{(Ku)(y)} - 1\right] d\mu(y) + \alpha J(u) \right\} \, , \tag{2.25}
$$

with the optimality condition

$$
0 = K^* \left(\frac{(K\hat{u})(y) - f(y)}{((K\hat{u})(y))^2}\right) + \alpha \hat{p} \qquad \hat{p} \in \partial J(\hat{u}) \, . \tag{2.26}
$$

One main drawback of (2.25) is that the fidelity term is not globally convex and therefore will not allow unconditional use of the general error estimates we will derive in Section 3. In order to convexify this speckle noise removal model, in [HNW09] Huang et al. suggested the substitution $z(y) := \ln((Ku)(y))$ to obtain the entirely convex optimization problem

$$
\hat{z} = \arg\min_{z \in \mathcal{W}(\Sigma)} \left\{ \int_{\Sigma} \left[z(y) + f(y)e^{-z(y)} - 1 - \ln(f(y))\right] d\mu(y) + \alpha J(z) \right\} \, , \tag{2.27}
$$

with optimality condition

$$
1 - f(y)e^{-\hat{z}(y)} + \alpha \hat{p} = 0 \, , \tag{2.28}
$$

for $\hat{p} \in \partial J(\hat{z})$. This model is a special case of the general multiplicative noise model presented in [SO08]. The disadvantage of this approach however is that the regularization is applied to a logarithmically scaled version of Ku instead of u, which might not be desirable. Moreover, $\hat{u}$ might not be computable since K^{-1} does not need to exist. At least in the denoising case with $K = Id$ this substitution seems to be natural and furthermore will allow the derivation of an error estimate for the Bregman distance between $\hat{z}$ and $\tilde{z} = \ln(g)$.

3 Results for General Models

After introducing some frequently used variational schemes dealing with non-Gaussian noise, we want to present general error estimates for convex variational schemes. These basic estimates will allow us to derive specific error estimates for the models presented in Section 2. Furthermore we want to explore duality and will 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 will ensure that a solution $\tilde{u}$ contains features that are enhanced by the regularization term J.

3.1 Basic Estimates

In this section we are going to derive basic error estimates in the Bregman distance measure for general, convex variational regularization methods. First of all we want to define the following class of functionals we further want to investigate:

Definition 3. *The class of functionals $\mathcal{C}(\Psi)$ is defined to fulfill the following equivalence:*

- $H \in \mathcal{C}(\Psi)$

- $H : \Psi \to \mathbb{R} \cup \{\infty\}$ *is proper, convex and semi-lower continuous .*

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 (1.4). If we denote a solution of (1.4), which fulfills the optimality condition due to the Karush-Kuhn-Tucker conditions (KKT), by $\hat{u}$, we can obtain the following error estimates for $D_J^{\text{symm}}(\hat{u}, \tilde{u})$, which are the main results of this work.

Theorem 1 (Basic Estimate I)**.** *Let $K : \mathcal{U}(\Omega) \to \mathcal{V}(\Sigma)$ be a linear and compact operator between Banach spaces $\mathcal{U}(\Omega)$ and $\mathcal{V}(\Sigma)$, on compact and bounded sets Ω and Σ. Furthermore, let $H_f \in \mathcal{C}(\mathcal{V}(\Sigma))$ and $J \in \mathcal{C}(\mathcal{W}(\Omega))$, $\mathcal{W}(\Omega) \subseteq \mathcal{U}(\Omega)$ and let H_f be locally bounded. Then, if the source condition* (SC) *is fulfilled, the error estimate*

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

holds.

Proof. Since H_f and J are convex, the optimality condition of (1.4) 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 H_f is locally bounded, 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 [ET99]. Hence, the following equality holds:

$$\hat{\eta} + \alpha \hat{p} = 0 \,,$$

for $\hat{\eta} \in \partial H_f(K\hat{u})$ and $\hat{p} \in \partial J(\hat{u})$. Since K is compact and linear, for every element $\eta \in \partial H_f(Ku)$ there exists an element $s = K^* \eta \in \mathcal{U}(\Omega)^*$ such that $\langle s, v - u \rangle_{\mathcal{U}(\Omega)} = \langle \eta, Kv - Ku \rangle_{\mathcal{V}(\Sigma)}$ holds, for all $v \in \mathcal{U}(\Omega)$. Hence, there exists an element $\hat{s}$ with

$$\langle \hat{\eta}, K\hat{u} - K\tilde{u} \rangle_{\mathcal{V}(\Sigma)} = \langle \hat{s}, \hat{u} - \tilde{u} \rangle_{\mathcal{U}(\Omega)} \,.$$

Therefore we are able to consider $\hat{s} + \alpha \hat{p} = 0$ instead of $\hat{\eta} + \alpha \hat{p} = 0$. If we subtract $\alpha \xi$, with ξ fulfilling (SC), and take the duality product with $\hat{u} - \tilde{u}$, we obtain

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

which equals

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

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

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

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 want to notice that we could have also claimed locally boundedness of J instead of H_f or even both to prove Theorem 1. Usually the chosen fidelities are locally bounded. It can easily be proved that all fidelity terms we will investigate in the following section will satisfy the conditions of Definition 3 and will be locally bounded. Therefore we decided to claim locally boundedness only for H_f.

We can further generalize the estimate of Theorem 1 to obtain the second important general estimate in this work.

Theorem 2 (Basic Estimate II)**.** *Let $K : \mathcal{U}(\Omega) \to \mathcal{V}(\Sigma)$ be a compact and linear operator between Banach spaces $\mathcal{U}(\Omega)$ and $\mathcal{V}(\Sigma)$, on bounded and compact sets Ω and Σ. Furthermore, let $H_f \in \mathcal{C}(\mathcal{V}(\Sigma))$ and $J \in \mathcal{C}(\mathcal{W}(\Omega))$, $\mathcal{W}(\Omega) \subseteq \mathcal{U}(\Omega)$ and let H_f be locally bounded. Then, if the source condition* (SC) *is fulfilled, the error estimate*

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

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

Proof. Due to Theorem 1 we have

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

The left hand side is equivalent to

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

while the right hand side can be rewritten to

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

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 (cf. [ET99]), 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

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

Under appropriate conditions, the Fenchel duality theorem (cf. [ET99]) implies the primal-dual relation

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

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

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

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

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

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, which will also be 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. Therefore we want to specify the source condition (SC) to

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

Although L^1 is only a subspace of $(L^\infty)^*$ we want to identify L^1 as the dual space of L^∞ (and vice versa) in the remaining course of this work. Hence, if (SCL^1) is fulfilled we have $q \in L^\infty(\Sigma)$. Furthermore, in the following it will be assumed that the operator K will satisfy the conditions of Theorem 1 and Theorem 2.

4.1 Laplace Noise

With the use of Theorem 1 we can - in analogy to the error estimates for the exact penalization model in [BO04] - obtain the following estimate for $H_f(Ku) := \int_\Sigma |Ku - f| d\mu(y)$, with $\hat{u}$ satisfying the optimality condition (2.13) and $\tilde{u}$ being the exact solution of (1.1).

Theorem 3. *Let $\hat{u}$ satisfy the optimality condition* (2.13) *and let $\tilde{u}$ denote the exact solution of* (1.1). *Furthermore, the difference between exact data g and noisy data f is bounded in the L^1-norm, i.e. $\int_\Sigma |f - g| d\mu(y) \leq \delta$ and* (SCL^1) *holds. Then, for the symmetric Bregman distance $D_J^{symm}(\hat{u}, \tilde{u})$ for a specific regularization functional $J \in \mathcal{C}(\mathcal{W}(\Omega))$, $\mathcal{W}(\Omega) \subseteq L^1(\Omega)$, the estimate*

$$\left(1 - \alpha \|q\|_{L^\infty(\Sigma)}\right) H_f(K\hat{u}) + \alpha D_J^{symm}(\hat{u}, \tilde{u}) \leq \left(1 + \alpha \|q\|_{L^\infty(\Sigma)}\right) \delta \tag{4.1}$$

holds. Furthermore, for $\alpha < \frac{1}{\|q\|_{L^\infty(\Sigma)}}$*, we obtain*

$$D_J^{symm}(\hat{u}, \tilde{u}) \leq \delta \left(\frac{1}{\alpha} + \|q\|_{L^\infty(\Sigma)} \right). \qquad (4.2)$$

Proof. It can be proved that $H_f \in \mathcal{C}(L^1(\Sigma))$ holds and that H_f is locally bounded. Hence, we obtain (due to Theorem 1)

$$\begin{aligned}
H_f(K\hat{u}) + \alpha D_J^{\mathrm{symm}}(\hat{u}, \tilde{u}) &\leq \underbrace{H_f(g)}_{\leq \delta} - \alpha \langle q, K\hat{u} - g \rangle_{L^1(\Sigma)} \\
&\leq \delta - \alpha \langle q, K\hat{u} - f + f - g \rangle_{L^1(\Sigma)} \\
&= \delta - \alpha \big(\langle q, K\hat{u} - f \rangle_{L_1(\Sigma)} \\
&\quad + \langle q, f - g \rangle_{L_1(\Sigma)} \big) \\
&\overset{\text{Hölder inequality}}{\leq} \delta + \alpha \|q\|_{L^\infty(\Sigma)} \Big(\|K\hat{u} - f\|_{L^1(\Sigma)} \\
&\quad + \|f - g\|_{L^1(\Sigma)} \Big) \\
&\leq \delta + \alpha \|q\|_{L^\infty(\Sigma)} \Big(\|K\hat{u} - f\|_{L^1(\Sigma)} + \delta \Big) ,
\end{aligned}$$

which leads us to (4.1). If we insert $H_f(K\hat{u}) = \|K\hat{u} - f\|_{L^1(\Sigma)}$ and set $\alpha < \frac{1}{\|q\|_{L^\infty(\Sigma)}}$ we can subtract $\|K\hat{u} - f\|_{L^1(\Sigma)}$ on both sides. If we multiply with α we obtain (4.2). □

As expected from the dual perspective above (note that the dual of the L^1-norm is a characteristic function), we obtain in the case of exact data ($\delta = 0$) for α sufficiently small

$$D_J^{\mathrm{symm}}(\hat{u}, \tilde{u}) = 0, \qquad H_g(K\hat{u}) = 0. \qquad (4.3)$$

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

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

We finally remark on the necessity of the source condition (SCL^1). 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 (cf. [N97]). Such results so far exist only for quadratic regularization and cannot be expected for general Bregman distance estimates - even less with non-Gaussian noise 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 Laplace noise 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

$$\hat{p} = K^* q, \qquad \hat{p} \in \partial J(\hat{u}), q \in L^1(\Sigma)^* = L^\infty(\Sigma),$$

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 (SCL^1).

4.2 Poisson Noise

In the case of Poisson noise we have

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

Theorem 2 will allow us to derive an error estimate that lies in $\mathcal{O}(\delta)$ for $\alpha \to 0$. Before that, we have to prove the following lemma.

Lemma 1. *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*

$$h_n(x) := (-1)^n \alpha\gamma(\varphi - x) - c\left(\varphi \ln\left(\frac{\varphi}{x}\right) - \varphi + x\right) , \tag{4.4}$$

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

$$\overline{x}_n^h = \frac{\varphi}{1 + (-1)^n \frac{\alpha}{c}\gamma} \tag{4.5}$$

and are therefore bounded by

$$h_n(x) < h_n(\overline{x}_n^h) = (-1)^n \alpha\gamma\varphi - c\varphi \ln\left(1 + (-1)^n \frac{\alpha}{c}\gamma\right) , \tag{4.6}$$

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 minima $\overline{x}_n^h$ 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 minimum. □

With the help of Lemma 1 we are able to prove the following error estimate.

Theorem 4. *Let $\hat{u}$ satisfy the optimality condition (2.20) with K 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. $\int_\Sigma f \, d\mu(y) = 1$. Furthermore, 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 (SCL^1) holds. Then, for $c \in]0,1[$ and $\alpha < \frac{c}{\|q\|_{L^\infty(\Sigma)}}$, the symmetric Bregman distance $D_J^{symm}(\hat{u}, \tilde{u})$ for a specific regularization functional $J \in \mathcal{C}(\mathcal{W}(\Omega))$, $\mathcal{W}(\Omega) \subseteq L^1(\Omega)$, is bounded via*

$$(1-c)H_f(K\hat{u}) + \alpha D_J^{symm}(\hat{u}, \tilde{u}) \leq (1+c)\delta - c\ln\left(1 - \frac{\alpha^2}{c^2} \|q\|^2_{L^\infty(\Sigma)}\right) . \tag{4.7}$$

Proof. It can easily be seen $H_f \in \mathcal{C}(L^1(\Sigma))$ holds and that H_f is locally bounded. Therefore, Theorem 2 states that (3.2) holds. 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 $\alpha^2 < \left(\frac{c}{q}\right)^2$, we can use Lemma 1 to estimate

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

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

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

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

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

and hence, (4.7) holds. □

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 1 can still be applied to find an upper estimate, the only difference is that there can be more than just one maximum.

One observes from a Taylor approximation of the second term on the right-hand side of (4.7) around $\alpha = 0$ that

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

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

4.3 Multiplicative Noise

In the case of multiplicative noise we are going to examine model (2.27) instead of (2.25), since (2.27) is convex for all z and therefore allows the application of Theorem 2. The source condition slightly differs, since there is no operator in that type of model. Therefore we get

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

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 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*

$$k_n(x) := (-1)^n \alpha\gamma(\varphi - x) - c(x + \varphi e^{-x} - 1 - \ln(\varphi)) \,, \tag{4.8}$$

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

$$\overline{x}_n^k = -\ln\left(\frac{c + (-1)^n \alpha\gamma}{c\varphi}\right) \,, \tag{4.9}$$

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

$$\begin{aligned} k_n(x) < k_n(\overline{x}_n^k) = \alpha\gamma\left((-1)^n\left(\varphi + \ln\left(\frac{c + (-1)^n\alpha\gamma}{c\varphi}\right)\right) - 1\right) \\ + c\ln\left(\frac{c + (-1)^n\alpha\gamma}{c}\right) \,, \end{aligned} \tag{4.10}$$

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

Proof. Similarly to Lemma 1, 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 minimum for all $n \in \mathbb{N}$. □

With the help of Lemma 2 we are able to prove the following error estimate.

Theorem 5. *Let* $\hat{z}$ *satisfy the optimality condition* (2.28) *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). *Furthermore, the difference between noisy data* f *and exact data* g *is bounded in the measure of* (2.25), *i.e.* $\int_\Sigma \ln\left(\frac{g}{f}\right) + \frac{f}{g} - 1 \, d\mu(y) \leq \delta$ *and* (zSCL^1) *holds. Then, for* $c \in]0,1[$ *and* $\alpha < \frac{c}{\|q\|_{L^\infty(\Sigma)}}$, *the symmetric Bregman distance* $D_J^{symm}(\hat{z}, \tilde{z})$ *for a specific regularization functional* $J \in \mathcal{C}(\mathcal{W}(\Sigma))$, $\mathcal{W}(\Sigma) \subseteq L^1(\Sigma)$, *is bounded via*

$$\begin{aligned} (1-c)H_f(\hat{z}) + \alpha D_J^{symm}(\hat{z}, \tilde{z}) \leq & (1+c)\delta \\ & + \alpha\,|\Sigma|\,\|q\|_{L^\infty(\Sigma)} \ln\left(\frac{c + \alpha\,\|q\|_{L^\infty(\Sigma)}}{c - \alpha\,\|q\|_{L^\infty(\Sigma)}}\right) . \end{aligned} \tag{4.11}$$

Proof. Again, it can easily be seen that $H_f \in \mathcal{C}(L^1(\Sigma))$ holds and that H_f is locally bounded. Hence, we can apply Theorem 2 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 2 we have

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

$$= \alpha \int_\Sigma q \underbrace{\left(\ln\left(\frac{c+\alpha q}{cf}\right) - \ln\left(\frac{c-\alpha q}{cf}\right)\right)}_{=\ln\left(\frac{c+\alpha q}{c-\alpha q}\right)} d\mu(y)$$

$$+ c \int_\Sigma \underbrace{\left(\ln\left(\frac{c+\alpha q}{c}\right) + \ln\left(\frac{c-\alpha q}{c}\right)\right)}_{=\ln\left(1-\frac{\alpha^2}{c^2}q^2\right)} d\mu(y)\,,$$

for $\alpha < \frac{c}{q}$. 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 can also easily be verified that the function $l(x) := \ln\left(1 - \frac{\alpha^2}{c^2}x^2\right)$ is strictly concave and has it's 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 (4.11). □

Again a Taylor approximation of the second term on the right-hand side of (4.11) around $\alpha = 0$ yields the asymptotic behaviour

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

5 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 Laplace noise (due to the interesting asymptotic exactness) and Poisson noise (due to the practical importance).

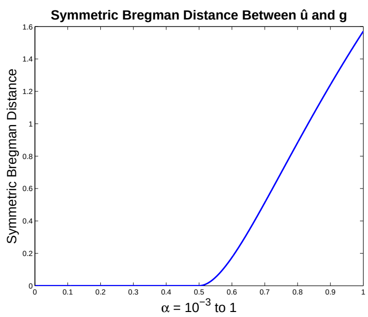


Figure 1: The Bregman distance error between $\hat{u}$ and g for $\alpha \in [10^{-3}, 1]$. As soon as $\alpha \leq \frac{1}{2}$, the error tends to be zero.

5.1 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

$$\int_\Omega |u - f| \; dx + \frac{\alpha}{2} \int_\Omega (|\nabla u|^2 + u^2) \; dx \to \min_{u \in H^1(\Omega)}, \tag{5.1}$$

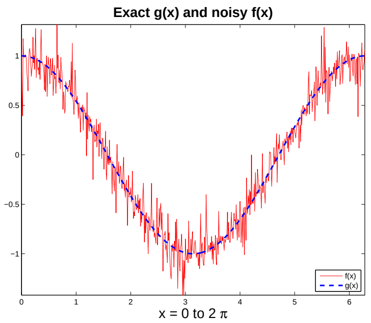


Figure 2: Exact $g(x) = \cos(x)$ and noisy $f(x)$, corrupted by Laplace noise with mean zero, $\sigma = 0.1$ and $\delta \approx 0.1037$.

whose optimality condition is

$$-\alpha\Delta u + \alpha u + s = 0, \qquad s \in \text{sign } (u - f). \tag{5.2}$$

A common approach to the numerical minimization of functionals like (5.1) 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:

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

Exchanging infimum and supremum in the last formula can easily be justified with standard methods in convex analysis (cf. [ET99]). 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 $A := (-\Delta \cdot + \cdot)$)

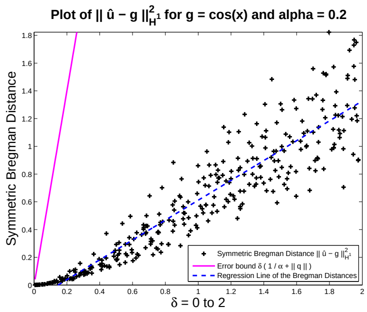
$$\frac{1}{2} \int_\Omega s(A^{-1}s) \; dx + \alpha \int_\Omega fs \; dx \to \min_{s \in L^\infty(\Omega), \|s\|_\infty \leq 1}. \tag{5.3}$$

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

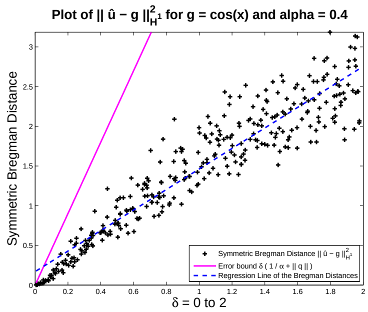
$$s^{k+1} = \mathbf{P}_1 \left(s^k - \tau \left(A^{-1} s^k + \alpha f \right) \right), \tag{5.4}$$

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

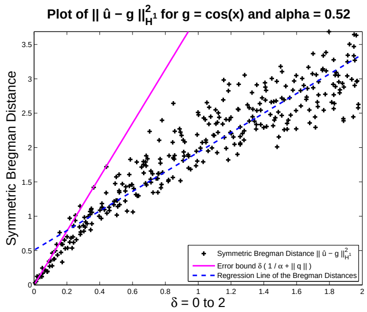
$$\mathbf{P}_1(v)(x) = \begin{cases} v(x) & \text{if } |v(x)| \leq 1 \\ \frac{v(x)}{|v(x)|} & \text{else} \end{cases}.$$



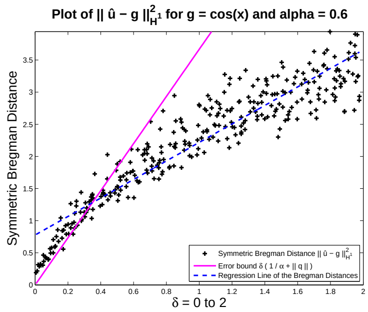
(a) $\alpha = 0.2$



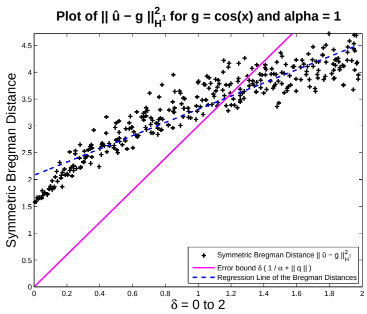
(b) $\alpha = 0.4$



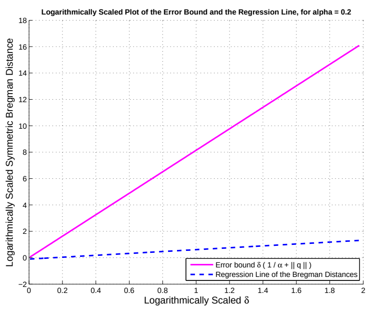
(c) $\alpha = 0.52$



(d) $\alpha = 0.6$



(e) $\alpha = 1$



(f) $\alpha = 0.2$

Figure 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 3(a) and 3(b) the computed Bregman distances lie beneath the error bound derived in (4.2), while the distances in 3(c), 3(d) and 3(e) partly violate this bound. Figure 3(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).

Due to the quadratic H^1 regularization we obtain

$$D^{\text{symm}}_{H^1}(\hat{u}, g) = 2D_{H^1}(\hat{u}, g) = \|\hat{u} - g\|^2_{H^1(\Omega)} \ , \tag{5.5}$$

and the source condition becomes

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

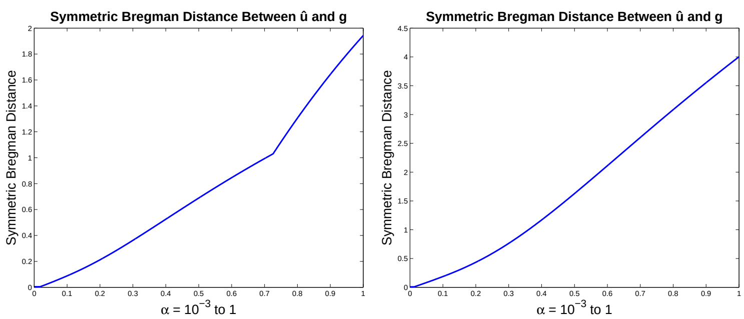
In the following we want to present two examples and their related error estimates.

Example 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 $\|q\|_{L^\infty([0,2\pi])} = \|\sin(x) + \cos(x)\|_{L^\infty([0,2\pi])} = 2$) holds, the regularized solution $\hat{u}$ should be identical to the exact solution g in the Bregman distance setting (5.5). This is also found in computational practice, as Figure 1 confirms.

In the following we want to illustrate the sharpness of (4.2) in the case of non-zero δ. For that 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 2 shows g and an exemplarily 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 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).

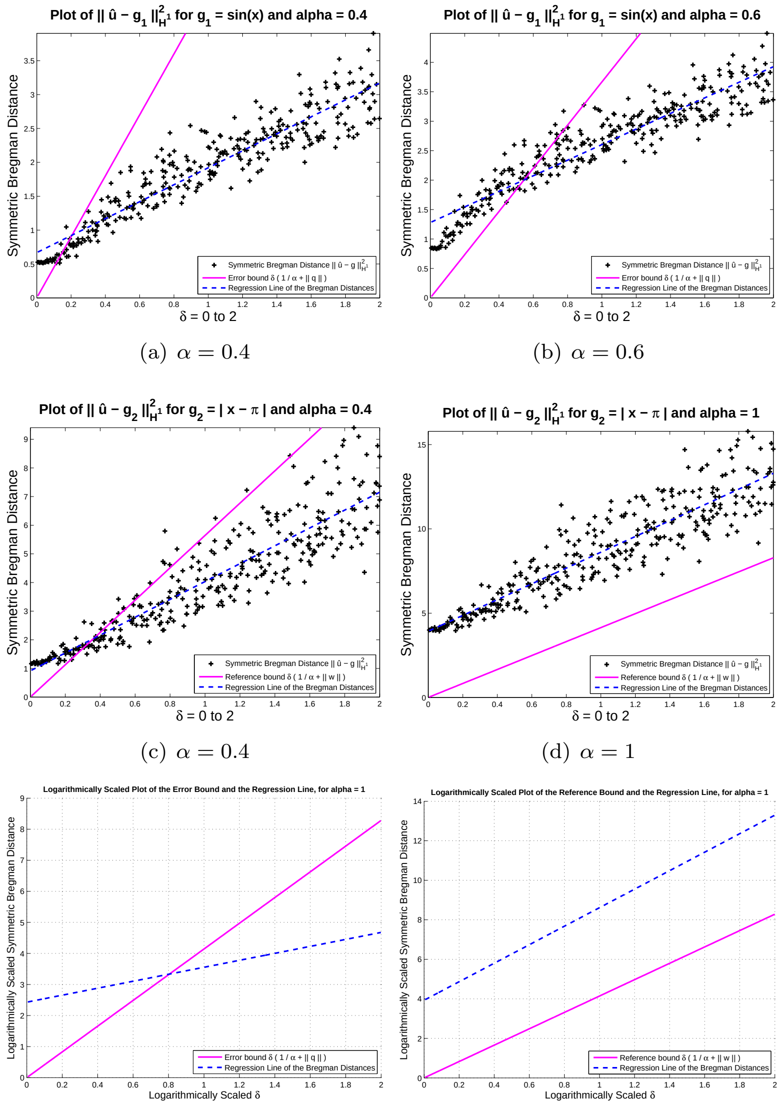


(a) $D^{\text{symm}}_{H^1}(\hat{u}, g_1)$ vs α (b) $D^{\text{symm}}_{H^1}(\hat{u}, g_2)$ vs α

Figure 4: The symmetric Bregman distances $D^{\text{symm}}_{H^1}(\hat{u}, g_1)$ 4(a) and $D^{\text{symm}}_{H^1}(\hat{u}, g_2)$ 4(b), for $\alpha \in \left[10^{-3}, 1\right]$.

Example 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 do 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,



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

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

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

Figure 5: The plots of the computed Bregman distances with violated (SCH^1). Figure 5(a) and Figure 5(b) show the Bregman distances $D_{H^1}^{\text{symm}}(\hat{u}, g_1)$ for $\alpha = 0.4$ and $\alpha = 0.6$, respectively. Figure 5(c) and Figure 5(d) represent the Bregman distances $D_{H^1}^{\text{symm}}(\hat{u}, g_2)$ for $\alpha = 0.4$ and $\alpha = 1$. Furthermore, Figure 5(e) and Figure 5(f) show the logarithmically scaled versions of the error/reference bound in comparison to a line regression of the Bregman distances for $\alpha = 1$.

in order to visualize that there exists no such error bound, we want to introduce 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 1 we want to begin with the case of exact data, i.e. $f = g$. If we plot the symmetric Bregman distance against α we obtain the graphs displayed in Figure 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 illustrate the importance of the source condition in the noisy case with non-zero δ we have proceeded as in Example 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 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 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.

5.2 Compressive Sensing with Poisson Noise

For the validation of the error estimates in the Poisson case we want to discover 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

$$\overline{K}u = g \tag{5.6}$$

with

$$g_{i,j} = \sum_{k=1}^{n} \phi_{i,k} u_{k,j} \ , \qquad \text{for } i = 1,\dots,l,\ j = 1,\dots,m\,, \tag{5.7}$$

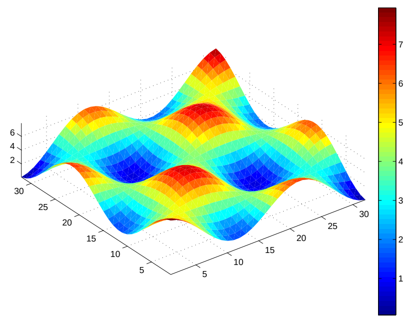
where $\phi_{i,j} \in [0,1]$ are uniformly distributed random numbers, $\Omega = \{1,\dots,n\} \times \{1,\dots,m\}$, $\Sigma = \{1,\dots,l\} \times \{1,\dots,m\}$ and $l >> n$, such that $\overline{K}$ has a large nullspace. Furthermore we consider f instead of g, with f being corrupted by Poisson noise.

5.2.1 Sparsity Regularization

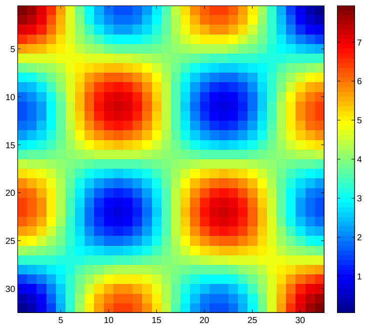
In the case of sparsity regularization we assume u to have a sparse representation due to a certain basis. Therefore, we consider an operator $B : \ell^1(\Theta) \to \ell^1(\Omega)$ such that $u = Bc$ holds, for coefficients $c \in \ell^1(\Theta)$. If we want to apply a regularization that minimizes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

$$\int_\Sigma f \ln\left(\frac{f}{\overline{K}Bc}\right) + \overline{K}Bc - f \ d\mu(y) + \alpha \sum_{i,j} |c^{i,j}|_{\ell^1(\Theta)} \to \min_{c \in \ell^1(\Theta)} \ . \tag{5.8}$$

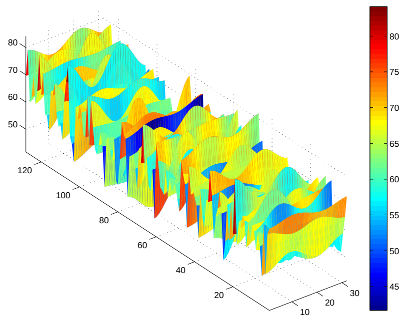
Notice that the measure μ is a point measure and the integral in (5.8) is indeed a sum over all discrete samples, which we write as an integral in order to keep



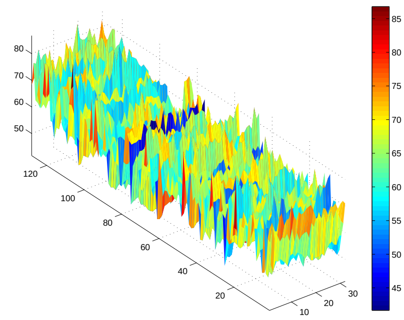
(a) $\tilde{u}$



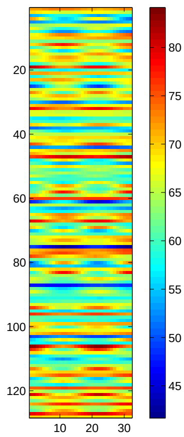
(b) $\tilde{u}$ top-view



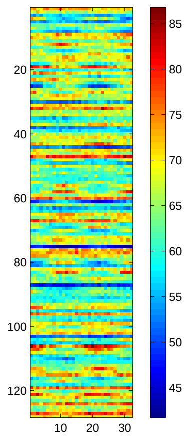
(c) g



(d) f



(e) g top-view



(f) f top-view

Figure 6: In 6 (a) we can see $\tilde{u}$ as defined via $\tilde{u} = B\tilde{c}$. A top-view of $\tilde{u}$ can be seen in 6 (b). In 6 (c) and 6 (d) the plots of $g = \overline{K}\tilde{u}$ and its noisy version f are described, while 6 (e) and 6 (f) show top-view illustrations of 6 (c) and 6 (d).

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

$$
\begin{aligned}
c^{i,j}_{k+\frac{1}{2}} &= c^{i,j}_k - \frac{1}{\tau}\left(B^T\overline{K}^T\left(1-\frac{f}{\overline{K}Bc_k}\right)\right)_{i,j}\,,\\
c^{i,j}_{k+1} &= \operatorname{sign}\left(c^{i,j}_{k+\frac{1}{2}}\right)\left(\left|c^{i,j}_{k+\frac{1}{2}}\right| - \frac{\alpha}{\tau}\right)_+\,,
\end{aligned}
\tag{5.9}
$$

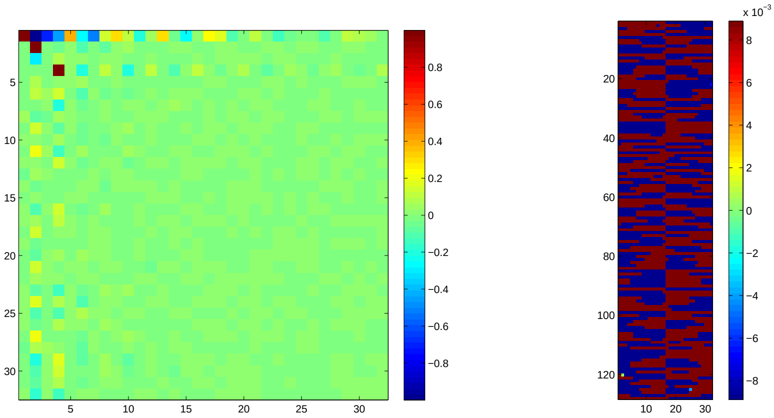
with $\tau > 0$ being a damping parameter and $a_+ = \max\{0, a\}$. The disadvantage of this approach is that positivity of u is not guaranteed. However, it is easy to implement and produces satisfactory results as we will see in the following.

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

$$
c^{a,b} = \gamma(a)\gamma(b)\sum_{i=0}^{n-1}\sum_{j=0}^{m-1} u_{i,j}\cos\left(\frac{\pi(2i+1)a}{2n}\right)\cos\left(\frac{\pi(2j+1)b}{2m}\right)\,, \tag{5.10}
$$

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 (5.10) 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 $g > 0$. Concludingly 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 $n = m = 32$. Hence we obtain $\tilde{c}^{1,1} = 128$,



(a) ξ (b) q

Figure 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); $\|q\|_{\ell^\infty} \approx 8.9 \times 10^{-3}$.

$\tilde{c}^{2,2} = 16$ 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. The operator dimension l is chosen to be $l = 128$. The damping parameter τ is set to the constant value of $\tau = 0.1225$.

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

$$\partial|\tilde{c}^{i,j}|_{\ell^1(\Omega)} = \operatorname{sign}(\tilde{c}^{i,j}) = \begin{cases} 1 & \text{for } \tilde{c}^{i,j} > 0 \\ \in [-1,1] & \text{for } \tilde{c}^{i,j} = 0 \\ -1 & \text{for } \tilde{c}^{i,j} < 0 \end{cases}\,, \tag{5.11}$$

for all $(i,j) \in \Omega$. Hence, to validate the error estimates and their sharpness we have computed ξ and q in order to satisfy (SCℓ^1), (5.11) and $q \in \arg\min_q \|q\|_{\ell^\infty}$. The computational results can be seen in Figure 7, where $\|q\|_{\ell^\infty} \approx 8.9 \times 10^{-3}$ holds. With ξ computed, the symmetric Bregman distance can easily be calculated via

$$D^{\text{symm}}_{|\cdot|_{\ell^1}}(\hat{c}, \tilde{c}) = \langle \hat{p} - \xi, \hat{c} - \tilde{c} \rangle_{\ell^1(\Omega)}\,, \tag{5.12}$$

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 (4.7). The results can be seen in Table 1. We want to notice that Theorem 4 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, the only modification that has to be made is to multiply $|f| = \int_\Sigma f \; d\mu(y)$ with the logarithmic term in (4.7) to obtain a valid estimate.

6 Outlook & 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 non-Gaussian models. We have seen that the appropriate definition of noise level in the non-Gaussian case is not directly related to the noise variance (which could be very large e.g. in the Poisson-distributed case), but rather to the data likelihood - more precisely the negative log-likelihood. With this definition the estimates indeed yield the same asymptotic order with respect to regularization parameter and noise level as in the Gaussian 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 Gaussian case.

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 a more detailed understanding of error estimation in a stochastic framework. Here we have only scratched the surface of statistics by using

c	α	(a)	(b)	(c)	(d)	(e)
0.5	0.05	85.25	16	178.7	0.1049	93.56
0.5	0.009	60.32	0.5133	178.7	0.0034	118.4
0.5	0.005	59.48	1.076	178.7	0.001049	119.2
0.5	0.003	58.39	8.756	178.7	0.0003777	120.3
0.5	0.0009	49.34	183.4	178.7	3.4e-005	129.4
0.5	0.0005	46.69	300.2	178.7	1.049e-005	132
0.5	0.0003	45.68	380.6	178.7	3.777e-006	133
0.5	5e-005	50.72	517.3	178.7	1.049e-007	128
0.1	0.009	108.6	0.5133	131	0.017	22.48
0.1	0.005	107.1	1.076	131	0.005246	23.99
0.1	0.003	105.1	8.756	131	0.001889	25.95
0.1	0.0009	88.8	183.4	131	0.00017	42.24
0.1	0.0005	84.04	300.2	131	5.246e-005	47.01
0.1	0.0003	82.23	380.6	131	1.889e-005	48.82
0.1	5e-005	91.29	517.3	131	5.246e-007	39.76
0.01	0.009	119.4	0.5133	120.3	0.17	1.054
0.01	0.005	117.8	1.076	120.3	0.05246	2.611
0.01	0.003	115.6	8.756	120.3	0.01889	4.732
0.01	0.0009	97.68	183.4	120.3	0.0017	22.64
0.01	0.0005	92.45	300.2	120.3	0.0005246	27.88
0.01	0.0003	90.45	380.6	120.3	0.0001889	29.87
0.01	5e-005	100.4	517.3	120.3	5.246e-006	19.91
0.001	0.05	170.3	16	119.3	58.45	7.382
0.001	0.009	120.5	0.5133	119.3	1.705	0.4317
0.001	0.005	118.8	1.076	119.3	0.5252	0.9405
0.001	0.003	116.7	8.756	119.3	0.1889	2.779
0.001	0.0009	98.57	183.4	119.3	0.017	20.7
0.001	0.0005	93.29	300.2	119.3	0.005246	25.97
0.001	0.0003	91.27	380.6	119.3	0.001889	27.98
0.001	5e-005	101.3	517.3	119.3	5.246e-005	17.92
0.0001	0.009	120.6	0.5133	119.1	27.14	25.65
0.0001	0.005	118.9	1.076	119.1	5.845	6.046
0.0001	0.003	116.8	8.756	119.1	1.959	4.337
0.0001	0.0009	98.66	183.4	119.1	0.1705	20.66
0.0001	0.0005	93.37	300.2	119.1	0.05252	25.83
0.0001	0.0003	91.36	380.6	119.1	0.01889	27.81
0.0001	5e-005	101.4	517.3	119.1	0.0005246	17.72
1e-005	0.0009	98.67	183.4	119.1	2.714	23.18
1e-005	0.0005	93.38	300.2	119.1	0.5845	26.34
1e-005	0.0003	91.36	380.6	119.1	0.1959	27.97
1e-005	5e-005	101.4	517.3	119.1	0.005252	17.71
1e-006	5e-005	101.4	517.3	119.1	0.05845	17.76

Table 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) $D^{\text{symm}}_{|\cdot|_{\ell^1}}(\hat{c},\tilde{c})$, (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 $\alpha D^{\text{symm}}_{|\cdot|_{\ell^1}}(\hat{c},\tilde{c})$ and hence, the error estimate (4.7) is always fulfilled.

Bayesian models and asymptotic formulas for the log-likelihood. 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. [EHK05, HP06, HP09]) or direct statistical approaches (cf.

[BHM+07]). In the case of Poisson noise a direct estimation of mean deviations has been developed in parallel in [H09], 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 data 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 levels.

Acknowledgements

This 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.

References

[AA08] G.Aubert, J.F.Aujol, *A Variational approach to remove multiplicative noise*, SIAM J. Appl. Math. 68 (2008), 925-946.

[BL09] J.Bardsley, A.Luttman, *Total variation-penalized Poisson likelihood estimation for ill-posed problems* , Adv. Comp. Math. (2009), to appear.

[BHM+07] N.Bissantz, T.Hohage, A.Munk, F.Ruymgaart, *Convergence rates of general regularization methods for statistical inverse problems and applications*, SIAM J. Numer. Anal. **45** (2007), 2610-2636.

[B67] L.M.Bregman, *The relaxation method for finding the common point of convex sets and its application to the solution of problems in convex programming*, USSR Comp. Math. Math. Phys. 7 (1967), 200-217.

[BKM+08] T.Bonesky, K.S.Kazimierski, P.Maass, F.Schöpfer, T.Schuster, *Minimization of Tikhonov functionals in Banach spaces*, Abstr. Appl. Anal. 19 (2008), 192679

[BO04] M.Burger, S.Osher, *Convergence rates of convex variational regularization*, Inverse Problems 20 (2004), 1411-1421.

[BGO+06] M.Burger, G.Gilboa, S.Osher, J.Xu, *Nonlinear inverse scale space methods*, Comm. Math. Sci. 4 (2006), 179-212.

[BRH07] M.Burger, E.Resmerita, L.He, *Error estimation for Bregman iterations and inverse scale space methods*, *Computing*, 81 (2007), 109-135.

[BSH08] M.Burger, C.Schönlieb, L.He, *Cahn-Hilliard inpainting and a generalization to gray-value images*, Preprint (2008), and submitted.

[BSB09] C.Brune, A.Sawatzky, 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, vol. 5567 (2009), 235-246.

[CDL+98] A.Chambolle, R.DeVore, N.Y.Lee, B.Lucier, *Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Trans. Image Proc., 7 (1998), 319–335.

[CE05] T.F.Chan, S.Esedoglu, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math. 65 (2005), 1817-1837.

[CS05] T.Chan, J.Shen, *Image Processing and Analysis* (SIAM, Philadelphia, 2005).

[C91] I.Csiszar, *Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems*, Ann. Stat.19 (1991), 2032-2066.

[DBZ+06] N.Dey, L.Blanc-Feraud, C.Zimmer, Z.Kam, P.Roux, J.C.Olivo-Marin, J. Zerubia, *Richardson-Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution*, Microscopy Research Technique 69 (2006), 260-266.

[EHN96] H.W.Engl, M.Hanke, A.Neubauer, Regularization of Inverse Problems, Kluwer Academic Publishers, Dordrecht (1996) (Paperback edition 2000).

[EHK05] H.W.Engl, A.Hofinger, S.Kindermann, *Convergence rates in the Prokhorov metric for assessing uncertainty in ill-posed problems*, Inverse Problems **21** (2005), 399-412.

[EKN89] H.W.Engl, K.Kunisch, A.Neubauer, *Convergence rates for Tikhonov regularisation of non-linear ill-posed problems*, Inverse Problems 5 (1989) 523-540.

[ET99] I.Ekeland, R.Temam, *Convex analysis and variational problems*, Corrected Reprint Edition, SIAM, Philadelphia, 1999.

[GG84] S.Geman, D.Geman, *Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images*, IEEE Trans. Pattern Anal. Mach. Intell. 6 (1984), 721741.

[GM85] S.Geman, D.E.McClure, *Bayesian image analysis: an application to single photon emission tomography*, Proc. Statistical Computation Section (American Statistical Association, Washington, 1985) 1218.

[HP06] A.Hofinger, H.K.Pikkarainen, *Convergence rates for the Bayesian approach to linear inverse problems*, Inverse Problems **23** (2006), 2469-2484.

[HP09] A.Hofinger, H.K.Pikkarainen, *Convergence rates for linear inverse problems in the presence of an additive normal noise*, Stoch. Anal. Appl. **27** (2009), 240-257.

[HKP+07] B.Hofmann, B.Kaltenbacher, C.Pöschl, O.Scherzer, *A Convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators*, Inverse Problems 23 (2007), 987–1010.

[H09] T.Hohage, *Variational regularization of inverse problems with Poisson data*, Preprint (2009).

[HNW09] Y.M.Huang, M.K.Ng, Y.W.Wen, *A new total variation method for multiplicative noise Removal*, SIAM J. Imaging Sciences 2 (2009), 20-40.

[JHC98] E.Jonsson, S.C.Huang, T.Chan, *Total variation regularization in positron emission tomography*, CAM Report 98-48 (UCLA, 1998).

[K97] K.C.Kiwiel, *Proximal minimization methods with generalized Bregman functions*, SIAM J. Cont. Optim. 35 (1997), 1142 - 1168

[LCA07] T.Le, R.Chartrand, T.J.Asaki, *A variational approach to reconstructing images corrupted by Poisson noise*, J. Math. Imaging Vision 27 (2007), 257–263.

[L08] D.A.Lorenz, *Convergence rates and source conditions for Tikhonov regularization with sparsity constraints*, J. Inverse Ill-Posed Problems 16(2008), 463-478.

[LT08] D.A.Lorenz, D.Trede, *Optimal Convergence rates for Tikhonov regularization in Besov Scales*, Inverse Problems 24 (2008), 055010.

[NW01] F.Natterer, F.Wübbeling, *Mathematical Methods in Image Reconstruction*, (SIAM, Philadelphia, 2001).

[N97] A.Neubauer, *On converse and saturation results for Tikhonov regularization of linear ill-posed problems*, SIAM J. Numer. Anal. 34 (1997), 517527.

[OBG+05] S.Osher, M.Burger, D.Goldfarb, J.Xu, W.Yin, An iterative regularization method for total variation-based image restoration, SIAM Multiscale Model. Simul. 4 (2005), 460-489.

[PZG99] V.Y.Panin, G.L.Zeng, G.T.Gullberg, *Total variation regulated EM Algorithm*, IEEE Trans. Nucl. Sci. NS-46 (1999) 2202-2010.

[P08] Christiane Pöschl, *Tikhonov Regularization with General Residual Term*, Leopold-Franzens-Universität Innsbruck, 2008, October

[RS] S.Remmele, M.Seeland, J.Hesser, *Fluorescence microscopy deconvolution based on Bregman iteration and Richardson-Lucy algorithm with TV regularization*, Preprint (University Heidelberg, 2008).

[R05] E.Resmerita,*Regularization of ill-posed problems in Banach spaces: convergence rates*, Inverse Problems 21 (2005), 13031314.

[RS06] E.Resmerita, O.Scherzer, *Error estimates for non-quadratic regularization and the relation to enhancing*, Inverse Problems 22 (2006) 801-814.

[RLO03] L.Rudin, P.L.Lions, S.Osher, *Multiplicative denoising and debluring: Theory and algorithms*, in: S.Osher, N.Paragios, eds., *Geometric Level Sets in Imaging, Vision, and Graphics* (Springer, New York, 2003), 103119.

[ROF92] L.Rudin, S.Osher, E.Fatemi, *Nonlinear total variation based noise removal algorithms* Physica D, 60 (1992), 259-268.

[SO08] J.Shi, S.Osher, *A nonlinear inverse scale space method for a convex multiplicative noise model*, SIAM Journal on Imaging Sciences (2008) 1(3):294321

[SHW93] D.L.Snyder, A.M.Hammoud, R.L.White, *Image recovery from data acquired with a charge- coupled-device camera*, J. Opt. Soc. Amer. A 10 (1993), 1014-1023.

[SHL+95] D.L.Snyder, C.W.Helstrom, A.D.Lanterman, M.Faisal, R.L.White, *Compensation for readout noise in CCD images*, J. Opt. Soc. Amer. A 12 (1995), 272-283.

[V02] C.Vogel, *Computational Methods for Inverse Problems* (SIAM, Philadelphia, 2002).

[WA04] M.N.Wernick, J.N.Aarsvold, eds., *Emission Tomography: The Fundamentals of PET and SPECT*, (Academic Press, San Diego, 2004)