Large Noise in Variational Regularization

Martin Burger[*]   Tapio Helin[†]   Hanne Kekkonen [‡]

February 9, 2018

Abstract

In this paper we consider variational regularization methods for inverse problems with large noise that is in general unbounded in the image space of the forward operator. We introduce a Banach space setting that allows to define a reasonable notion of solutions for more general noise in a larger space provided one has sufficient mapping properties of the forward operators.

A key observation, which guides us through the subsequent analysis, is that such a general noise model can be understood with the same setting as approximate source conditions (while a standard model of bounded noise is related directly to classical source conditions). Based on this insight we obtain a quite general existence result for regularized variational problems and derive error estimates in terms of Bregman distances. The latter are specialized for the particularly important cases of one- and p-homogeneous regularization functionals.

As a natural further step we study stochastic noise models and in particular white noise, for which we derive error estimates in terms of the expectation of the Bregman distance. The finiteness of certain expectations leads to a novel class of abstract smoothness conditions on the forward operator, which can be easily interpreted in the Hilbert space case. We finally exemplify the approach and in particular the conditions for popular examples of regularization functionals given by squared norm, Besov norm and total variation, respectively.

Keywords: Variational Regularization, Error Estimates, Large Noise, White Noise, Bregman Distances

[*]Institut für Numerische und Angewandte Mathematik, Westfälische Wilhelms-Universität (WWU) Münster. Einsteinstr. 62, D 48149 Münster, Germany. e-mail: martin.burger@wwu.de

[†]Department of Mathematics and Statistics, University of Helsinki (UH). P.O. Box 68 (Gustaf Hällströmin katu 2b) FI-00014 University of Helsinki, Finland. e-mail: tapio.helin@helsinki.fi

[‡]Cantab Capital institute, Faculty of Mathematics, University of Cambridge (UC), Wilberforce Road, Cambridge CB3 0WA, United Kingdom. e-mail: hnk22@cam.ac.uk

Contents

1   Introduction

Motivated by stochastic modelling of noise, in particular white noise, the treatment of
inverse problems with large noise has received strong attention recently [23, 24, 42, 43,
53]. In this case large noise means that the norm of the data perturbation introduced
by the noise is not small or might be even unbounded in the image space of the forward
operator. Recently several papers have tackled such problems in the setting of linear
regularization methods (corresponding to quadratic variational regularization), but also
in those approaches some points were restrictive. The work by Eggermont et al. [24]
assumes noise potentially large in the image space of the forward operator, but still
being an element of this space. This allows to gain some insight, but still excludes white
noise, where the latter condition is satisfied with probability zero. Also some difficulties
related to the appropriate formulation of the regularized problem with white noise are
not appearing in this way. Another line of research restricts to inverse problems with
special settings of function spaces, namely some Sobolev spaces [42, 43] or Hilbert scales

[49, 50, 51, 52]. In these works estimates are obtained in weaker norms however and the setting still partly shadows the general structure.

In this paper we directly tackle the issue of large noise variational regularization with convex regularization functionals in Banach spaces. We derive a rather general theory that can be adapted to special homogeneity properties of the regularization functional, in particular to quadratic (Tikhonov) and one-homogeneous regularizations as popularized via total variation methods [60, 9] and sparsity (see e.g. [6, 20, 56]). We consider the linear ill-posed problem

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

for a continuous linear operator $K : X \to Y$, where X and Y are separable Banach and Hilbert space, respectively. For our setting of the noise let (Z, Y, Z^*) be a Gelfand triple such that $Z \subset Y$ is a dense subspace with Banach structure and the dual pairing of Z and Z^* is compatible with the inner product of Y, i.e., by identifying $Y = Y^*$ we have

$$\langle u, v \rangle_{Z \times Z^*} = \langle u, v \rangle_Y$$

whenever $u \in Z \subset Y$ and $v \in Y = Y^* \subset Z^*$. The key assumption we make is that $K : X \to Z$ is continuous. It directly follows that K^* has a continuous extension $K^* : Z^* \to X^*$. The noisy data are given by

$$f^\delta = Ku^\dagger + \delta n, \tag{1.2}$$

where $n \in Z^*$ and $\delta > 0$ models the noise level. Notice carefully that $f^\delta \in Z^*$ can be unbounded in the norm of Y, which yields our setting of large noise. It is crucial that due to the continuous extension property $K^* n$ is bounded in X^*.

As usual in variational methods we obtain a regularized solution of (1.2) by computing a minimizer u_α^δ of a weighted sum of the square residual (in the norm of Y) and the regularization functional. However, since the (squared) norm of f^δ is not necessarily finite, it is more appropriate to consider an expansion of the square residual [42, 43] and compute u_α^δ as a minimizer of

$$J_\alpha^\delta(u) = \frac{1}{2} \|Ku\|_Y^2 - \langle Ku, f^\delta \rangle_{Z \times Z^*} + \alpha R(u) \tag{1.3}$$

with a convex regularization functional $R : X \to \mathbb{R} \cup \{\infty\}$.

Our main assumptions on R in addition to convexity are

(R1) the functional R is lower semicontinuous in some topology τ on X,

(R2) the sub-level sets $M_\rho = \{R \leq \rho\}$ are sequentially compact in the topology τ on X and

(R3) the convex conjugate $R^\star$ is finite on a ball in X^* centered at zero.

The first two are the standard conditions needed for existence proofs and as we shall see below together with (R3) they will also lead to a general existence result for minimizers

3

of J^δ_α in the case of positive α. Note that we assume that $K : X \to Z$ is continuous in τ topology. A standard example is R being a power of a norm in a Banach space possessing a predual space. In this case the Banach–Alaoglu theorem yields compactness in the weak-star topology, for which we have genuine lower semicontinuity of the norm. We mention that a major difference to the case of bounded noise is that there is no natural lower bound for J^δ_α (the lower bound in the case of bounded noise is $-\frac{1}{2}\|f^\delta\|^2_Y + \alpha R(u_0)$, with u_0 being a minimizer of R), which is the only complication in the analysis below and needs a suitable approximation of the noise together with (R3). To make some results below more accessible we will further employ the symmetry condition

(R4) $R(-u) = R(u)$ for all $u \in X$,

which is however not essential for the overall line of arguments.

Our key observation is related to error estimates between u^δ_α and a solution $u^\dagger$ minimizing R among all possible solutions of $Ku = f$. The usual way to obtain such is starting from the optimality condition for a minimizer

$$K^*(Ku^\delta_\alpha - f^\delta) + \alpha\mu^\delta_\alpha = 0, \qquad \mu^\delta_\alpha \in \partial R(u^\delta_\alpha), \tag{1.4}$$

where

$$\partial R(u) = \{\mu \in X^* \mid R(u) - R(v) \le \langle \mu, u - v \rangle_{X^* \times X} \text{ for all } v \in X\}$$

stands for the subdifferential. Next, the form (1.2) of f^δ is inserted and multiples of a subgradient $\mu^\dagger \in \partial R(u^\dagger)$ are added on both sides to arrive at

$$K^*K(u^\delta_\alpha - u^\dagger) + \alpha(\mu^\delta_\alpha - \mu^\dagger) = \delta\eta - \alpha\mu^\dagger, \tag{1.5}$$

where $\eta = K^*n \in X^*$. The following step is to take a duality product with $u^\delta_\alpha - u^\dagger$ and hence derive error estimates in the Bregman distance [5, 7]. In doing so one can strongly benefit if $\mu^\dagger$ satisfies a source condition, i.e., if $\mu^\dagger = K^*w^\dagger$ for some $w^\dagger \in Y$. Note that in the bounded noise model also η satisfies such a condition, which becomes violated in our setting. Since η and $\mu^\dagger$ appear in a similar fashion on the right-hand side we see that the unboundedness of the noise in Y leads to a similar technical issue as the violation of the source condition for $\mu^\dagger$. However, the latter is reasonably well understood and has been tackled by the concept of distance functions and approximate source conditions [32, 36, 38, 61], which are related to the growth rate of $\|w^\dagger\|_Y$ as $K^*w^\dagger$ approximates $\mu^\dagger$. Due to the analogous role of $\mu^\dagger$ and η it is natural to use the same paradigm for approximating the large noise and this is the basic foundation of the analysis in this paper.

Following this idea our key contribution is to derive Bregman distance based error estimates between u^δ_α and $u^\dagger$ for a general convex R. Given deterministic noise model, one can derive explicit converge rate results given (a variation of) an approximate source condition on $\mu^\dagger$ and η. In this paper, we prove convergence rates for the special cases of 1-homogeneous $R(u) = \|u\|_X$ as well as the p-homogeneous $R(u) = \frac{1}{p}\|u\|^p_X$ for $1 < p < \infty$.

For our main motivation, random noise, the approximate source condition needs to be reconsidered in a statistical framework. In this work our interest lies in the frequentist risk between estimator $U^\delta_\alpha = U^\delta_\alpha(\omega)$ and the true unknown $u^\dagger$. In such paradigm

we find that the expected decay rate of the approximate source condition of the noise term is sufficient to guarantee a convergence rate result. Here we study and derive the convergence rate of frequentist risk for three examples: quadratic Tikhonov regularization, Besov norm regularization and total variation regularization. As for the noise we assume the canonical Gaussian white noise model on the Gelfand triplet (Z, Y, Z^*) that has the well-known property that n is almost surely unbounded in Y.

Let us shortly discuss some earlier work. After introducing the idea in [8], Bregman distances have been frequently used as an error measure for studying convergence rates of regularized solutions in Banach spaces. Convergence rates for the Bregman distance were further developed in e.g. [3, 10, 31, 37, 45, 47, 58, 59]. Iterative regularization based on Bregman distances were analysed e.g. in [10, 55]. The literature on regularization theory in Banach spaces is quite extensive, but throughout the paper we often refer to an excellent textbook on the topic [61]. For a recent discussion of Bregman distances we refer to [7].

The general approach in statistical literature for solving frequentist inverse problems with white noise is typically based on obtaining a singular value decomposition (SVD) of the forward operator K and then constructing a procedure based on spectral regularization, see, e.g. [1, 12, 15, 21, 30, 44, 46]. However, in general inverse problems settings the SVD can rarely be computed analytically. Hence our approach which does not require the identification of the SVD basis of K can be applied for wider range of inverse problems.

The remainder of the paper is organized as follows: in Section 2 we consider the theory for general convex functional R. Main results of this section include the proof of existence of u_α^δ as well as a related a-priori estimate in Section 2.1. The general error estimates are given in Section 2.4. In Section 3 we derive convergence rates for different homogeneous examples of R. Next, we turn our focus on random noise in Section 4 and consider examples of regularization by a quadratic Tikhonov functional (Section 4.2), Besov norm (Section 4.3) and total variation functional (Section 4.4). Finally, we give an outlook to applications of our work to Bayesian inference in Section 5.

2 General Estimates

In the following we discuss the general approach for variational regularization under the assumptions above. We start by establishing the existence of a minimizer of J_α^δ for $\alpha > 0$, which also yields some a-priori bounds for the solution.

2.1 Existence and a-priori Estimates

For general noise the existence of J_α^δ is not clear from standard arguments. While usual lower semicontinuity arguments remains unchanged, the key issue is compactness, which follows from an a-priori estimate on R due to the compactness of sublevel sets. In deriving such an estimate we need to bypass the missing lower bound of J_α^δ.

5

Proposition 2.1. *Let R satisfy the assumptions (R1)-(R4), then the functional J_α^δ has a minimizer. Moreover, any such minimizer u_α^δ satisfies*

$$R(u_\alpha^\delta) \leq \frac{1+\gamma}{1-\gamma} R(u^\dagger) + \frac{\delta^2}{2\alpha(1-\gamma)} \|w\|_Y^2 + \frac{2\gamma}{1-\gamma} R^\star \left(\frac{\delta}{\alpha\gamma} (K^* w - \eta) \right) \qquad (2.1)$$

for any $\gamma \in (0,1)$ and $w \in Y$. Above $\eta = K^* n$.

Proof. Consider the sublevel set $M = \{u \in X \mid J_\alpha^\delta(u) \leq J_\alpha^\delta(u^\dagger)\}$. Clearly, M is non-empty since $u^\dagger \in M$. Further, any $u \in M$ satisfies

$$
\begin{aligned}
\frac{1}{2} \|K(u - u^\dagger)\|_Y^2 + \alpha R(u) &\leq \delta \langle K(u - u^\dagger), n \rangle_{Z \times Z^*} + \alpha R(u^\dagger) \\
&= \delta \langle u - u^\dagger, \eta \rangle_{X \times X^*} + \alpha R(u^\dagger) \\
&= \delta \langle u - u^\dagger, \eta - K^* w \rangle_{X \times X^*} + \delta \langle K(u - u^\dagger), w \rangle_Y + \alpha R(u^\dagger) \\
&\leq \alpha\gamma R(u) + 2\alpha\gamma R^\star \left(\frac{\delta}{\alpha\gamma} (K^* w - \eta) \right) + \frac{1}{2} \|K(u - u^\dagger)\|_Y^2 \\
&\quad + \frac{\delta^2}{2} \|w\|_Y^2 + \alpha(1 + \gamma) R(u^\dagger),
\end{aligned}
$$

where $0 < \gamma < 1$ and $w \in Y$ is arbitrary. The last inequality follows from using generalized Young's inequality. For the definition of the convex conjugate $R^\star$ see Appendix A. Due to assumptions (R2), (R3) and Y being dense in Z^* we can now choose $w \in Y$ such that for a constant $C > 0$

$$R^\star \left(\frac{\delta}{\alpha\gamma} (K^* w - \eta) \right) \leq C,$$

and hence we obtain

$$R(u) \leq \frac{1+\gamma}{1-\gamma} R(u^\dagger) + \frac{\delta^2}{2\alpha(1-\gamma)} \|w\|_Y^2 + \frac{2C\gamma}{1-\gamma}$$

which implies M is compact due to assumption (R2).

Now the existence follows by standard arguments. Without loss of generality we can assume that $\{u_j\}_{j=1}^\infty \subset M$ is a minimizing sequence of J_α^δ. Since M is compact, there exists a converging subsequence $u_{j_k} \to \widetilde{u} \in X$. Finally, the lower semicontinuity of J_α^δ yields that $\widetilde{u}$ is a minimizer. Note that with existence of a minimizer $\widetilde{u}$ we directly obtain the a-priori estimate (2.1). $\qquad\square$

Remark 2.2. We can prove a similar a-prior estimate for R also without the symmetry assumption (R4). In that case we get for the minimizer u_α^δ

$$
\begin{aligned}
R(u_\alpha^\delta) \leq \frac{1+\gamma}{1-\gamma} R(u^\dagger) &+ \frac{\delta^2}{2\alpha(1-\gamma)} \|w\|_Y^2 \\
&+ \frac{\gamma}{1-\gamma} \left(R^\star \left(\frac{\delta}{\alpha\gamma} (\eta - K^* w) \right) + R^\star \left(\frac{\delta}{\alpha\gamma} (K^* w - \eta) \right) \right).
\end{aligned}
$$

6

2.2 Basic Ingredients of Error Estimates

In the following we discuss some basics needed for the derivation of error estimates and the use of the approximate source conditions. The starting point for error estimates is the optimality condition mentioned above. Since the first two terms are linear and quadratic it is straight-forward to verify that they are Frechet-differentiable in our setting. Then the subdifferential of the whole functional equals the sum of the Frechet derivative of the first part and the subdifferential of the regularization functional (cf. [25]), which immediately implies the following statement:

Proposition 2.3. *Under the assumptions above, a minimizer u_α^δ of J_α^δ satisfies the optimality condition (1.4).*

As mentioned above, error estimates are based on rewriting (1.4) and then taking a duality product with $u_\alpha^\delta - u^\dagger$. This naturally leads to estimates in the Bregman distance, whose definition we recall for completeness:

Definition 2.4 (Bregman distance)**.** Let $R : X \to \mathbb{R} \cup \{\infty\}$ be a convex functional. Then for each $\mu_v \in \partial R(v) \subset X^*$ we define generalised Bregman distance between u and v as

$$D_R^{\mu_v}(u, v) = R(u) - R(v) - \langle \mu_v, u - v \rangle_{X^* \times X}.$$

Moreover, for $\mu_u \in \partial R(u)$ we define symmetric Bregman distance between u and v as

$$D_R^{\mu_u, \mu_v}(u, v) = \langle \mu_u - \mu_v, u - v \rangle_{X^* \times X}. \tag{2.2}$$

Let us now sketch the basic steps in the derivation of error estimates and the standard route in the case of bounded noise. Taking a duality product with (1.4) and $u_\alpha^\delta - u^\dagger$ we get

$$\|K(u_\alpha^\delta - u^\dagger)\|_Y^2 + \alpha D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \leq \langle \delta\eta - \alpha\mu^\dagger, u_\alpha^\delta - u^\dagger \rangle_{X^* \times X}.$$

The nice case leading directly to estimates is $\eta = K^* n$ with $n \in Y$ and the additional source condition $\mu^\dagger = K^* w^\dagger \in X^*$ for $w^\dagger \in Y$. Then the right-hand side becomes

$$\langle \delta\eta - \alpha\mu^\dagger, u_\alpha^\delta - u^\dagger \rangle_{X^* \times X} = \langle \delta n - \alpha w^\dagger, K(u_\alpha^\delta - u^\dagger) \rangle_Y,$$

and Young's inequality implies

$$\frac{1}{2}\|K(u_\alpha^\delta - u^\dagger)\|_Y^2 + \alpha D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \leq \frac{1}{2}\|\delta n - \alpha w^\dagger\|_Y^2.$$

The problem now becomes more difficult if η or $\mu^\dagger$ are not in the range of K^* (if the range is defined as $K^* Y$ and not K^* on a larger space including the noise). Note that with the notation using η instead of $K^* n$ it becomes apparent that technically η not in the range of K^* is equally difficult as $\mu^\dagger$ not in the range of K^*. The latter case is however reasonably well understood, at least in the case of strictly convex functionals R. This is discussed in detail in [61]. The idea is to use a so-called approximate source condition,

quantifying how well $\mu^\dagger$ can be approximated by elements in the range of K^*. Since $\mu^\dagger$ needs to be in the closure of the range, there exists a sequence w_n with $K^* w_n \to \mu^\dagger$. On the other hand it is not in the range, hence $\|w_n\|$ necessarily diverges. Thus, one can measure how well $\mu^\dagger$, respectively in our case $\delta\eta - \alpha\mu^\dagger$ can be approximated by elements $K^* w$ with a given upper bound on $\|w\|$. The best estimates are then obtained by balancing errors containing the approximation of $\delta\eta - \alpha\mu^\dagger$ and $\|w\|$.

In the case of no strict source condition and unbounded noise we will approximate $\mu^\dagger$ and η with separate elements $K^* w_1$ and $K^* w_2$ respectively. Then we can write

$$
\begin{aligned}
\langle \delta\eta - \alpha\mu^\dagger, u_\alpha^\delta - u^\dagger \rangle_{X^* \times X} = \\
\langle \delta(\eta - K^* w_2) - \alpha(\mu^\dagger - K^* w_1), u_\alpha^\delta - u^\dagger \rangle_{X^* \times X} + \langle \delta w_2 - \alpha w_1, K(u_\alpha^\delta - u^\dagger) \rangle_Y,
\end{aligned}
$$

where $w_1, w_2 \in Y$. The second term on the right hand side can now be estimated using Young's inequality as above, while for the first term it is natural to apply the generalized Young's inequality as in the proof of Proposition 2.1. We shall estimate the terms multiplied by δ and α separately and overall study a problem of estimating a term of the form $\langle \eta, u_\alpha^\delta - u^\dagger \rangle_{X^* \times X}$. For this sake we could separately estimate the duality products with u_α^δ and $u^\dagger$ as in the proof of Proposition 2.1. However, as we are interested mainly in functionals with some homogeneity properties and in particular (R4) we shall see that it is beneficial to use the following direct estimate

$$
\left\langle \eta, u_\alpha^\delta - u^\dagger \right\rangle_{X^* \times X} = \zeta \left\langle \frac{\eta}{\zeta}, u_\alpha^\delta - u^\dagger \right\rangle_{X^* \times X} \leq \zeta R\left(u_\alpha^\delta - u^\dagger\right) + \zeta R^\star\left(\frac{\eta}{\zeta}\right), \qquad (2.3)
$$

which we shall employ further with appropriately chosen $\zeta > 0$. We observe that in proceeding as above we are left with two terms in dependence on w_1, namely $\frac{\alpha^2}{2}\|w_1\|^2$ and $\alpha\zeta R^\star(\frac{K^* w_1 - \mu^\dagger}{\zeta})$. Analogous reasoning holds for w_2, with α replaced by δ. This motivates our approach to the approximate source conditions to be detailed in the following.

2.3 A Variation on Approximate Source Condition

The standard concept of approximate source condition is to consider the case $R(u) = \|u\|_X^r$ for some power $r > 1$ (cf. [61]). The key concept is the so-called distance function

$$
d_\rho(\vartheta) := \inf_{w \in Y} \{ \|K^* w - \vartheta\|_{X^*} \mid \|w\|_Y \leq \rho \}, \qquad (2.4)
$$

and its asymptotics as $\rho \to \infty$. Note that in the case of a fulfilled source condition $d_\rho(\vartheta) = 0$ for ρ sufficiently large, while in the really approximate case $d_\rho(\vartheta)$ decays to zero at a finite rate. Hence, the speed of decay of $d_\rho(\vartheta)$ is a natural measure to quantify the approximateness of the source condition. Unfortunately the existing theory employing the approximate source conditions or the even more implicit variational inequalities only works for the special norm-type functionals above (cf. [61]) and in addition uses some moduli of strict convexity of the norms. This of course excludes the most interesting cases of one-homogeneous regularizations such as sparsity and total variation. Hence we propose to consider a more general formulation based on convex duality.

As we have seen above it is crucial to approximate some elements $\vartheta \in X^*$ by $K^* w$ with $w \in Y$ in some kind of Fenchel dual problem defined by K and R. More precisely, we are interested in minimal values of the functional

$$E_{\alpha,\zeta}(w;\vartheta) = \zeta R^\star \left(\frac{K^* w - \vartheta}{\zeta} \right) + \frac{\alpha}{2} \|w\|_Y^2,$$

which we shall denote as

$$e_{\alpha,\zeta}(\vartheta) = \inf_{w \in Y} E_{\alpha,\zeta}(w;\vartheta). \tag{2.5}$$

In this paper approximated source conditions correspond to determining decay rates for (2.5).

Remark 2.5. Indeed it can be inferred from the Fenchel duality theorem (cf. [25]) that $E_{\alpha,\zeta}(w;\vartheta)$ is dual (as a functional of w) to

$$F_{\alpha,\zeta}(v;\vartheta) = \frac{1}{2\alpha} \|Kv\|_Y^2 - \langle \vartheta, v \rangle_{X^* \times X} + \zeta R(v)$$

and it holds that

$$e_{\alpha,\zeta}(\vartheta) = - \inf_{v \in X} F_{\alpha,\zeta}(v;\vartheta). \tag{2.6}$$

Thus, the measure $e_{\alpha,\zeta}$ measures how fast a regularization method approximating ϑ (related to the noise or source element) diverges and is hence a natural quantity. For $R^\star(\vartheta)$ being finite, this immediately implies a bound on $e_{\alpha,\zeta}(\vartheta)$ via the generalized Young inequality

$$\langle \vartheta, v \rangle_{X^* \times X} \leq \frac{1}{\zeta} R^\star(\vartheta) + \zeta R(v).$$

This results into

$$e_{\alpha,\zeta}(\vartheta) \leq \frac{1}{\zeta} R^\star(\vartheta).$$

Obviously, this estimate is not optimal under most conditions since it does not involve the first term in $F_{\alpha,\zeta}$. As we shall see below the bound can be improved under certain conditions, depending also on the homogeneity properties of R.

In the case of a Hilbert space regularization, $R(u) = \frac{1}{2}\|u\|_X^2$, we have

$$E_{\alpha,\zeta}(w;\vartheta) = \frac{1}{2\zeta}\|\vartheta - K^* w\|_X^2 + \frac{\alpha}{2}\|w\|_Y^2 = \alpha E_{1,\zeta\alpha}(w;\vartheta)$$

and the problem of computing the minimizer is a classical Tikhonov regularization problem. In particular in this example but also in the more general case the minimization of $E_{\alpha,\zeta}$ is closely related to the minimization in the definition of distance functions, roughly it can be understood as some kind of Lagrange multiplier formulation of the constrained problem for computing d_ρ. Moreover, it can be related to classical source conditions $\vartheta = (K^* K)^\nu w^*$, which are routinely used in the linear theory (cf. [26]). We will provide other examples of approximate source conditions and their implications for functionals with a certain degree of homogeneity in Section 3.

9

We finally mention that we can also rewrite the a-priori estimate from Proposition 2.1 in terms of the approximate source condition (2.5)

$$R(u_\alpha^\delta) \leq \frac{1+\zeta}{1-\zeta} R(u^\dagger) + \frac{2\delta}{\alpha(1-\zeta)} e_{\frac{\delta}{2}, \frac{\alpha\zeta}{\delta}}(\eta) \tag{2.7}$$

with any $\zeta \in (0, 1)$. Moreover, the approximate source conditions match well the use of Bregman distances as an error measure. Indeed, using the definition (2.2) of Bregman distance and completely analogous techniques as in the following one can show a conditional well-posedness result in the (symmetric) Bregman distances for all elements u_1, u_2 and their subgradients satisfying an approximate source condition and a bound $R(u_1 - u_2) \leq \gamma$, i.e.

$$D_R^{\mu_1, \mu_2}(u_1, u_2) \leq \varphi(\|Ku_1 - Ku_2\|), \tag{2.8}$$

with

$$\varphi(t) = \inf_{\alpha, \zeta} \left(\frac{t^2}{2\alpha} + \gamma\zeta + e_{\alpha, \zeta}(\mu_1) + e_{\alpha, \zeta}(\mu_2) \right).$$

2.4 Error Estimates

In order to obtain error estimates we start from the rewritten version of the optimality condition (1.5) and take a duality product with $u_\alpha^\delta - u^\dagger$ in the same way as sketched above. Then the right-hand side is estimated as

$$\langle \delta\eta - \alpha\mu^\dagger, u_\alpha^\delta - u^\dagger \rangle_{X^* \times X} \leq$$
$$(\alpha\zeta_1 + \delta\zeta_2) R(u_\alpha^\delta - u^\dagger) + \frac{1}{2}\|K(u_\alpha^\delta - u^\dagger)\|_Y^2 + \alpha e_{\alpha, \zeta_1}(\mu^\dagger) + \delta e_{\delta, \zeta_2}(\eta) \tag{2.9}$$

This immediately leads to the following error estimates:

Proposition 2.6. *Let R satisfy (R1)-(R4). Then with the assumptions above we obtain for any positive real numbers ζ_1, ζ_2:*

$$\|K(u_\alpha^\delta - u^\dagger)\|_Y^2 + 2\alpha D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \leq 2(\alpha\zeta_1 + \delta\zeta_2) R(u_\alpha^\delta - u^\dagger) + 2\alpha e_{\alpha, \zeta_1}(\mu^\dagger) + 2\delta e_{\delta, \zeta_2}(\eta) \tag{2.10}$$

and furthermore

$$D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \leq (\zeta_1 + \frac{\delta}{\alpha}\zeta_2) R(u_\alpha^\delta - u^\dagger) + e_{\alpha, \zeta_1}(\mu^\dagger) + \frac{\delta}{\alpha} e_{\delta, \zeta_2}(\eta). \tag{2.11}$$

In order to obtain meaningful estimates we need to further estimate $R(u_\alpha^\delta - u^\dagger)$, ideally in terms of the Bregman distance, which however strongly depends on the specific scaling properties of the underlying functional R. Inspired by p-convex functionals (cf. [4]), we shall consider the following assumption: There exists $\theta \in [0, 1]$ such that

$$R(u - v) \leq C_\theta(u, v) \left(D_R^{\mu_u, \mu_v}(u, v) \right)^\theta \tag{2.12}$$

for all $u, v \in X$, $\mu_u \in \partial R(u)$ and $\mu_v \in \partial R(v)$. Above the constant C_θ is bounded on sets where $R(u)$ and $R(v)$ are bounded. The canonical examples to be considered are square norms (leading to $\theta = 1$) and one-homogeneous functionals (leading to $\theta = 0$).

Example 2.7. Let X be a Hilbert space, L a bounded linear operator, and $R(u) = \frac{1}{2}\|Lu\|_X^2$. In consequence, $D_R^{\mu_u,\mu_v}(u,v) = \|L(u-v)\|_X^2 = 2R(u-v)$ and inequality (2.12) holds with $\theta = 1$ and $C_\theta(u,v) \equiv \frac{1}{2}$.

Example 2.8. Let R be one-homogeneous, symmetric around zero, and convex. We immediately obtain a triangle inequality

$$R(u-v) \le R(u) + R(v),$$

and hence (2.12) holds with $\theta = 0$ and $C_0(u,v) = R(u) + R(v)$. It is easy to see that for R of the above form no estimate with $\theta > 0$ can hold. As an example consider $R : \mathbb{R} \to \mathbb{R}$, $R(u) = |u|$. If u and v differ, but have equal sign, we obtain $|u - v| \ne 0$, but $D_R^{p,q}(u,v) = 0$.

2.5 Convergence theorems

With assumption (2.12) we can further estimate the right-hand side in the above estimates as

$$\left(\zeta_1 + \frac{\delta}{\alpha}\zeta_2\right) R(u_\alpha^\delta - u^\dagger) \;\le\; \left(\zeta_1 + \frac{\delta}{\alpha}\zeta_2\right) C_\theta(u_\alpha^\delta, u^\dagger) D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger)^\theta$$

$$\le\; \theta D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) + (1-\theta)\left(\zeta_1 + \frac{\delta}{\alpha}\zeta_2\right)^{1/(1-\theta)} C_\theta(u_\alpha^\delta, u^\dagger)^{1/(1-\theta)},$$

if $\theta < 1$. In the case $\theta = 1$ the first estimate is the only relevant one. This leads to the following result

Theorem 2.9. *Let R satisfy the assumptions of Proposition 2.6 and (2.12). Then for, $\theta < 1$ we obtain*

$$D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \le \inf_{(\zeta_1,\zeta_2)\in\mathbb{R}_+^2} \left\{ \left(\zeta_1 + \frac{\delta}{\alpha}\zeta_2\right)^{1/(1-\theta)} C_\theta(u_\alpha^\delta, u^\dagger)^{1/(1-\theta)} \right.$$

$$\left. + \frac{1}{1-\theta} e_{\alpha,\zeta_1}(u^\dagger) + \frac{\delta}{\alpha(1-\theta)} e_{\delta,\zeta_2}(\eta) \right\}. \quad (2.13)$$

For $\theta = 1$ the estimate

$$D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \le \inf_{\zeta_1,\zeta_2\in\Sigma} \frac{e_{\alpha,\zeta_1}(u^\dagger) + \frac{\delta}{\alpha} e_{\delta,\zeta_2}(\eta)}{1 - (\zeta_1 + \frac{\delta}{\alpha}\zeta_2) C_1(u_\alpha^\delta, u^\dagger)} \quad (2.14)$$

holds with

$$\Sigma = \left\{ (\zeta_1,\zeta_2) \in \mathbb{R}_+^2 \;\middle|\; \left(\zeta_1 + \frac{\delta}{\alpha}\zeta_2\right) C_1(u_\alpha^\delta, u^\dagger) < 1 \right\}.$$

We finally mention an alternative statement of Theorem 2.9, which also takes into account an estimate of the residual. In the subsequent parts of the paper we will not discuss estimates for the residual, but obviously those can be obtained in the same way using the following result:

Theorem 2.10. *Let R satisfy the assumptions of Proposition 2.6 and (2.12). Then for, $\theta < 1$ we obtain*

$$\|K(u_\alpha^\delta - u^\dagger)\|_Y^2 + (2\alpha - \theta)D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger)$$

$$\leq \inf_{(\zeta_1, \zeta_2) \in (\mathbb{R}^+)^2} \left\{ (1-\theta)(2(\alpha\zeta_1 + \delta\zeta_2))^{1/(1-\theta)} C_\theta(u_\alpha^\delta, u^\dagger)^{1/(1-\theta)} + 2\alpha e_{\alpha, \zeta_1}(\mu^\dagger) + 2\delta e_{\delta, \zeta_2}(\eta) \right\}$$

Note that the constant $C_\theta(u_\alpha^\delta, u^\dagger)$ above depends on $R(u_\alpha^\delta)$ and hence also on the corresponding a-priori estimate.

3 Convergence Rates for Homogeneous Regularizations

Let us shortly introduce some notation. Throughout the following sections we denote $f \lesssim g$ for two functions if there exists a universal constant $C > 0$ such that $f \leq Cg$ as functions. Moreover, if functions f and g are equivalent we write $f \simeq g$. Notice that if a random variable X has a probability distribution π, we write $X \sim \pi$.

3.1 Regularization by one-homogeneous functionals

Let us directly proceed to the case of a one-homogeneous functional R such as Besov-one norms or total variation. We assume that X is a suitable space such that R has a trivial nullspace (note that the nullspace of a one-homogeneous convex functional is always a linear space, and if it is finite-dimensional this component can be eliminated via similar arguments as in the total variation case detailed in [9]).

In this case we can define a dual "norm" S on X^* via

$$S(q) = \sup_{R(u) \leq 1} \langle q, u \rangle_{X^* \times X}. \tag{3.1}$$

Note that S is again one-homogeneous. The one-homogeneity of R implies

$$\langle q, u \rangle_{X^* \times X} \leq R(u) \, S(q) \tag{3.2}$$

for all $u \in X$ and $q \in X^*$. In the case of one-homogeneous R we can relate $R^\star$ and S as follows:

Lemma 3.1. *Let $R : X \to \mathbb{R} \cup \{\infty\}$ be convex, non-negative and one-homogeneous and let $S : X^* \to \mathbb{R} \cup \{\infty\}$ be defined by (3.1). Then for any $c \in \mathbb{R}^+$, we have*

$$R^\star(cq) = \begin{cases} 0 & \text{if } S(q) \leq \frac{1}{c} \\ +\infty & \text{else.} \end{cases} \tag{3.3}$$

Note that under the convexity condition and the homogeneity $R(cu) = |c|R(u)$, that is, the regularisation functional R is sublinear. Hence, the proof follows from general results on sublinear functionals in [35, Section V]. Next we formulate an alternative approximate source condition for the unknown and the noise term in one-homogeneous case.

Assumption 3.2. We assume to have an approximate source condition of order $r_1 \geq 0$ for the unknown, that is we require

$$\inf_{w \in Y} \left\{ \|w\|_Y^2 \ \middle| \ S(\mu^\dagger - K^* w) \leq \beta \right\} = C_1 \beta^{-r_1} \tag{3.4}$$

when $\beta > 0$ small enough. We also require similar condition of order $r_2 \geq 0$ for the noise term and assume

$$\inf_{w \in Y} \left\{ \|w\|_Y^2 \ \middle| \ S(\eta - K^* w) \leq \beta \right\} = C_2 \beta^{-r_2}. \tag{3.5}$$

Notice carefully that in the case when we do not have strict source condition the corresponding parameter r_j must be strictly positive. Before proceeding, let us record the following technical lemma:

Lemma 3.3. *The minimum of a problem*

$$M = \inf_{\zeta \in \mathbb{R}_+} \left(a \zeta^s + b \zeta^{-t} \right)$$

for $a, b, s, t > 0$ is achieved at

$$\zeta = \left(\frac{bt}{as} \right)^{\frac{1}{s+t}} \tag{3.6}$$

yielding a minimum

$$M \simeq a^{\frac{t}{s+t}} b^{\frac{s}{s+t}}.$$

Proof. Variational calculus yields

$$as \zeta^{s-1} - bt \zeta^{-t-1} = 0$$

at the minimum and hence (3.6) holds. Moreover, we obtain

$$M = a \left(\frac{bt}{as} \right)^{\frac{s}{s+t}} \left(1 + \frac{s}{t} \right) \simeq a^{\frac{t}{s+t}} b^{\frac{s}{s+t}}.$$

$\square$

Theorem 3.4. *Let X be a Banach space and $R(u) = \|u\|_X$. Suppose that Assumption 3.2 is satisfied with some orders $r_1, r_2 \geq 0$. For the choice $\alpha \simeq \delta^\kappa$ where*

$$\kappa = \begin{cases} \frac{(1+r_1)(2+r_2)}{(2+r_1)(1+r_2)} & \text{for } r_1 \leq r_2 \text{ and} \\ 1 & \text{for } r_2 < r_1, \end{cases}$$

we have that

$$D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \lesssim \begin{cases} \delta^{\frac{2+r_2}{(2+r_1)(1+r_2)}} & \text{for } r_1 \leq r_2 \text{ and} \\ \delta^{\frac{1}{1+r_1}} & \text{for } r_2 < r_1. \end{cases}$$

13

Proof. Using Lemma 3.1 we can write

$$e_{\delta,\varsigma}(\eta) = \frac{\delta}{2} \inf_{w \in Y} \left\{ \|w\|_Y^2 \;\middle|\; S(\eta - K^*w) \leq \varsigma \right\}. \tag{3.7}$$

Recall from Example 2.8 that the one-homogeneous case corresponds to parameter $\theta = 0$ in condition (2.12) and $C_0(u,v) = R(u) + R(v)$. The a priori estimate in Proposition 2.1 gives us

$$R(u_\alpha^\delta) \leq \frac{1+\gamma}{1-\gamma} R(u^\dagger) + \frac{\delta^2}{2\alpha(1-\gamma)} \inf_{w \in Y} \{ \|w\|_Y^2 \mid S(K^*w - \eta) \leq \frac{\alpha\gamma}{\delta} \}$$

for any $\gamma \in [0,1)$ and $w \in Y$. Now it follows from Theorem 2.9 that

$$
\begin{aligned}
D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) &\leq \inf_{\zeta_1, \zeta_2 \in \mathbb{R}_+^2} \left(\zeta_1 C_0(u_\alpha^\delta, u^\dagger) + e_{\alpha,\zeta_1}(\mu^\dagger) + \frac{\delta}{\alpha} \zeta_2 C_0(u_\alpha^\delta, u^\dagger) + \frac{\delta}{\alpha} e_{\delta,\zeta_2}(\eta) \right) \\
&\lesssim M_1 + M_2.
\end{aligned} \tag{3.8}
$$

where

$$M_1 = \inf_{\zeta_1 \in \mathbb{R}_+} \left\{ \zeta_1 \left(1 + \frac{\delta}{\alpha} e_{\delta, \frac{\alpha\gamma}{\delta}}(\eta) \right) + e_{\alpha,\zeta_1}(\mu^\dagger) \right\}$$

and

$$M_2 = \inf_{\zeta_2 \in \mathbb{R}_+} \left\{ \frac{\zeta_2 \delta}{\alpha} \left(1 + \frac{\delta}{\alpha} e_{\delta, \frac{\alpha\gamma}{\delta}}(\eta) \right) + \frac{\delta}{\alpha} e_{\delta,\zeta_2}(\eta) \right\}.$$

From assumption 3.2 we get the following estimates:

$$e_{\delta, \frac{\alpha\gamma}{\delta}}(\eta) \lesssim \delta^{1+r_2} \alpha^{-r_2} \gamma^{-r_2},$$
$$e_{\alpha,\zeta_1}(\mu^\dagger) \lesssim \alpha \zeta_1^{-r_1},$$
$$e_{\delta,\zeta_2}(\eta) \lesssim \delta \zeta_2^{-r_2}.$$

By assuming that $\alpha \simeq \delta^\kappa$ with some $\kappa > 0$ we can write

$$D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \leq \inf_{\zeta_1 \in \mathbb{R}_+} \left\{ (1 + \delta^{r_3})\zeta_1 + \delta^\kappa \zeta_1^{-r_1} \right\} + \inf_{\zeta_2 \in \mathbb{R}_+} \left\{ \delta^{1-\kappa}(1 + \delta^{r_3})\zeta_2 + \delta^{2-\kappa} \zeta_2^{-r_2} \right\}.$$

Above $r_3 = 2 - \kappa + r_2(1 - \kappa) > 0$ when $\kappa \leq 1$. Now by Lemma 3.3 we get estimate

$$M_1 + M_2 \simeq \delta^{\frac{\kappa}{1+r_1}} + \delta^{\frac{(1-\kappa)r_2 + 2 - \kappa}{1+r_2}}.$$

Optimizing the above we get $\kappa = \frac{(1+r_1)(2+r_2)}{(2+r_1)(1+r_2)}$ when $r_1 \leq r_2$ which gives us the convergence rate

$$D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \lesssim \delta^{\frac{2+r_2}{(1+r_2)(2+r_1)}}.$$

In the case $r_1 \geq r_2$ we choose $\kappa = 1$ to get

$$D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \lesssim \delta^{\frac{1}{1+r_1}}.$$

$\square$

14

Corollary 3.5. *If in addition to the assumptions of Theorem 3.4 we assume the exact source condition for the unknown $u^\dagger$, i.e., $r_1 = 0$, we get a convergence rate*

$$D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \simeq \delta^{\frac{2+r_2}{2(1+r_2)}}.$$

We finally provide some examples of one-homogeneous functionals and the meaning of the approximate source conditions in such cases:

Example 3.6. We start with a slightly artificial example, which however provides consistency with the linear theory. Assume X is a Hilbert space and let $R(u) = \|u\|_X$. Then $S(v) = \|v\|_X$ and for $u \neq 0$ we have $\partial R(u) = \left\{ \frac{u}{\|u\|_X} \right\}$. In (3.4) we thus look for the norm of w when

$$\left\| \frac{u^\dagger}{\|u^\dagger\|_X} - K^* w \right\|_X \leq \beta.$$

Setting $\widetilde{w} = w\|u^\dagger\|_X$, $\widetilde{\beta} = \beta\|u^\dagger\|_X$, and $\widetilde{C}_1 = C_1\|u^\dagger\|_X^{r_1}$ we can reformulate the approximate source condition in terms of $\widetilde{\beta}$ tending to zero as

$$\inf_{\widetilde{w} \in Y} \left\{ \|\widetilde{w}\|_Y^2 \;\middle|\; \|u^\dagger - K^*\widetilde{w}\|_X \leq \widetilde{\beta} \right\} = \widetilde{C}_1 \widetilde{\beta}^{-r_1},$$

which is related to the approximate source condition for the regularization with the quadratic norm $\frac{1}{2}\|u\|_X^2$, whose subdifferential is $\{u\}$, see Assumption 3.10.

We can further relate the approximate source condition to standard source conditions in the linear case. Assume that K is a compact operator and let $u^\dagger = (K^*K)^\nu v$ for $v \in X$ and $0 < \nu < \frac{1}{2}$. Then a simple calculation based on the singular value expansion shows that for each $\beta > 0$ there exists w with $\|u^\dagger - K^*w\|_X \leq \beta$ and $\|w\|_Y \sim \beta^{1-1/2\nu}$. Hence, an approximate source condition is satisfied with $r_1 \to 0$ as $\nu \to \frac{1}{2}$ and $r_1 \to \infty$ as $\nu \to 0$.

Example 3.7. We proceed to one of the most canonical examples of a one-homogeneous functional, namely $X = \ell^1(\mathbb{N})$, i.e. $u = (u_i)_{i=1}^\infty$, $u_i \in \mathbb{R}$, and

$$R(u) = \|u\|_X = \sum_{i=1}^\infty |u_i|.$$

In order to obtain a first insight we also consider a simple diagonal operator $K : \ell^1(\mathbb{N}) \to \ell^2(\mathbb{N})$, $(u_i) \mapsto (k_i u_i)$, with a decreasing sequence k_i of nonzero real values converging to zero. In addition we require that $K : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$ is bounded. Then the dual norm is given by $S(v) = \|v\|_\infty$, i.e., in the definition of the source condition (3.4) for β arbitrarily small we need to find w such that

$$\sup_{i \in \mathbb{N}} |\mu_i^\dagger - k_i w_i| = \|\mu^\dagger - K^*w\|_\infty = S(\mu^\dagger - K^*w) \leq \beta.$$

Now assume that $u^\dagger$ has an infinite support, i.e., there exists a nontrivial sequence i_j with $u_{i_j}^\dagger \neq 0$ and hence $|\mu_{i_j}^\dagger| = 1$. Then we conclude for i_j sufficiently large (note that $k_{i_j} w_{i_j}$ converges to zero)

$$1 - |k_{i_j}|\,|w_{i_j}| \leq |\mu_{i_j}^\dagger - k_{i_j} w_{i_j}| \leq \beta.$$

This implies $|w_{i_j}| \geq \frac{1-\beta}{|k_{i_j}|}$ for i_j sufficiently large, hence the corresponding sequence w cannot be an element of $\ell^2(\mathbb{N})$. Thus, the source condition can only be satisfied for $u^\dagger$ having a finite support.

On the other hand, if $u^\dagger$ has finite support contained in $\{1, \ldots, M\}$ we can choose a subgradient $\mu^\dagger$ with $\mu_i^\dagger = 0$ for $i > M$. Then for element w with $w_i = \frac{\mu_i^\dagger}{k_i}$ for $i \leq M$ and $w_i = 0$ else, we have $\mu^\dagger = K^* w$ and w has finite ℓ^2-norm, i.e., a standard source condition is satisfied. We thus see that in this case the asymptotic source condition does not seem useful, it is as strong as the original source condition due to the special structure of the subgradients. This behaviour is related to the degenerate behaviour of the ℓ^1-regularization, which has some phase transition from a well-posed finite dimensional to an ill-posed infinite dimensional problem depending on the support (cf. [31, 28]). We mention however that it is easy to see that the set of subgradients $\mu^\dagger$ for which the approximate source condition holds is larger for $r_1 > 0$ than for the standard source condition $r_1 = 0$, indeed the set is strictly increasing with r_1. The implication of this fact for the error estimation is not clear at this moment however.

We finally mention that approximate source conditions are useful in any case to quantify large noise as in (3.5), since the elements η are then arbitrary and not characterized by the structure of subgradients. The condition simply measures how well the noise can be approximated in the ℓ^∞-norm by elements $K^* w$.

Example 3.8. A synthesis of the last two examples is group sparsity in Hilbert spaces. For simplicity let H be a single Hilbert space and $X = \ell^1(\mathbb{N}; H)$, $u = (u_i)_{i=1}^\infty$, $u_i \in H$, with

$$R(u) = \sum_{i=1}^\infty \|u_i\|_H.$$

A subgradient $\mu \in \partial R(u)$ is given by $\mu = (\mu_1, \mu_2, \ldots)$ with $\mu_i \in H$ such that $\|\mu_i\|_H \leq 1$ and $\mu_i = \frac{u_i}{\|u_i\|_H}$ if $u_i \neq 0$. As in the previous example of ℓ^1-regularization one can verify that an approximate source condition can only hold if only a finite number of the $u_i^\dagger$ are different from zero. On the other hand a source condition is not automatically satisfied in this case, we also need $\mu_i^\dagger = (K^* w)_i$, which requires analogous properties of the $u_i^\dagger$ as for $u^\dagger$ in Example 3.6.

Example 3.9. We finally provide a standard example as already used in [8], namely total variation denoising by the Rudin–Osher–Fatemi (ROF) functional (cf. [60]). This means we assume $D \subset \mathbb{R}^2$, $X = BV(D)$, $Y = L^2(D)$ and K is the embedding operator

between these spaces. The regularization functional is given by

$$R(u) = \sup_{\varphi \in C_0^\infty(D), \|\varphi\|_\infty \le 1} \int_D \nabla \cdot \varphi u \, dx. \tag{3.9}$$

It is well-known that source conditions for ROF denoising are related to square integrability of the curvature of level sets (cf. [8, 14]). On the other hand the approximation properties of ROF are particularly bad if the exact solution is the characteristic function of a square, whose curvature is just a Radon measure on the jump set (cf. [11]). Hence, a natural conjecture is that approximate source conditions with $0 < r_1 < \infty$ are related to q-integrability of the curvature of level sets for $1 < q < 2$, which we make more explicit in the following. Assume for this sake that u is the indicator function of a simply connected compact subset $D_0 \subset D$, such that $\Gamma = \partial D_0$ is of class C^1 and the curvature κ is an element of $L^q(\Gamma)$. An elementary computation then yields that the normal and tangent fields are Hölder continuous along Γ with exponent $\gamma = 1 - \frac{1}{q}$. Using this kind of regularity one obtains that the signed distance function b_Γ is of class $C^{1,\gamma}$ in a neighbourhood of Γ and the curvature of level sets Δb_Γ is q-integrable in this neighbourhood. Now, similar to [8, 9], we can construct a subgradient μ of the form

$$\mu = \nabla \cdot g, \quad g = \psi(b_\Gamma) \nabla b_\Gamma$$

with ψ be a continuously differentiable function with local support around zero, $\psi(0) = 1$ and $0 \le \psi \le 1$ else. For this subgradient we easily verify $\|g\|_{L^\infty} \le 1$ and $\|\mu\|_{L^p} < \infty$.

The dual norm of BV given by

$$S(v) = \inf\{\|h\|_{L^\infty} \mid \nabla \cdot h = v\},$$

thus in the approximate source condition we know that $S(\mu - K^* w) \le \beta$ as soon as we find any h with $\|h - g\|_{L^\infty} \le \beta$. A sufficient condition for the approximate source condition is thus

$$\inf\{\|\nabla \cdot h\|_{L^2} \mid \|h - g\|_{L^\infty} \le \beta\} \le C\beta^{-r_1}.$$

To verify such a condition let G be a standard kernel with unit integral such as the Gaussian, $G_\epsilon = \epsilon^{-2} G(\frac{\cdot}{\epsilon})$ and $h = G_\epsilon * g$. Then it is a standard computation for convolutions to show that

$$\|G_\epsilon * g - g\|_{L^\infty} \le C_1 \epsilon^\gamma, \qquad \|\nabla \cdot G_\epsilon * g\|_{L^2} \le C_2 \epsilon^{2(1 - 2/q)} = C_2 \epsilon^{2(2\gamma - 1)}.$$

For $q > 1$ we obtain $\gamma > 0$ and hence we can choose $\beta \sim \epsilon^\gamma$, which implies an approximate source condition with $r_1 = \frac{2}{\gamma} - 4 = \frac{4 - 2q}{q - 1}$. With $q = 2$ we recover the standard source condition for square integrable curvature, with $q \to 1$ we obtain $r_1 \to \infty$.

3.2 Regularization by p-homogeneous functional for $1 < p < \infty$

In this section we consider regularization with functionals of type $R(u) = \frac{1}{p} \|u\|_X^p$ for $1 < p < \infty$. Below $p, q \in (1, \infty)$ are Hölder conjugates, i.e.,

$$\frac{1}{p} + \frac{1}{q} = 1.$$

17

Here we utilize additional assumptions regarding the Banach space X. Let $J_p : X \to X^*$ denote the set-valued duality mapping

$$J_p(u) = \{\mu \in X^* \mid \langle \mu, u \rangle_{X^* \times X} = \|u\|_X \|\mu\|_{X^*} \quad \text{and} \quad \|\mu\|_{X^*} = \|u\|_X^{p-1}\}.$$

A Banach space X is said to be *p-convex* if there exists a constant $c_p > 0$ such that

$$\frac{1}{p} \|u - v\|_X^p \geq \frac{1}{p} \|u\|_X^p - \langle j_p^X(u), v \rangle_{X^* \times X} + \frac{c_p}{p} \|v\|_X^p$$

for all $u, v \in X$ and all $j_p \in J_p$. Moreover, X is called *p-smooth* if there exists a constant $G_p > 0$ such that

$$\frac{1}{p} \|u - v\|_X^p \leq \frac{1}{p} \|u\|_X^p - \langle j_p^X(u), v \rangle_{X^* \times X} + \frac{G_p}{p} \|v\|_X^p$$

for all $u, v \in X$ and all $j_p \in J_p$. The basic consequences and properties of these geometrical assumptions are listed in [61]. For what follows, an important connection between the convexity and smoothness assumptions is given in [61, Thm 2.52]: X is p-smooth if and only if X^* is q-convex. Moreover, X is p-convex if and only if X^* is q-smooth. Some examples of $\max\{2, p\}$-convex and $\min\{2, p\}$-smooth spaces are sequence spaces ℓ^p, Lebesgue spaces L^p, and Sobolev spaces $W^{m,p}$. Notice also that in this Section we consider a p-smooth Banach space X for some $p > 1$. In that case it is well known (see [61, Remark 2.38]) that the duality mapping J_p is single-valued.

Next we define an alternative approximate source condition for the unknown and noise in case $R(u) = \frac{1}{p} \|u\|_X^p$ for $1 < p < \infty$.

Assumption 3.10. We assume to have an approximate source conditions of order $r_1 \geq 0$ for the unknown, i.e., we require that

$$\inf_{w \in Y} \left\{ \frac{1}{\beta} \|K^* w - \mu^\dagger\|_{X^*}^q + \frac{1}{2} \|w\|_Y^2 \right\} \leq C \beta^{-r_1}$$

when $\beta > 0$ is small enough. We also require a similar condition of order $r_2 \geq 0$ for the noise term and assume

$$\inf_{w \in Y} \left\{ \frac{1}{\beta} \|K^* w - \eta\|_{X^*}^q + \frac{1}{2} \|w\|_Y^2 \right\} \leq C \beta^{-r_2}.$$

For comparison of the above approximate source condition to distance function see Remark 3.17.

Case $1 < p < 2$

Theorem 3.11. *Suppose that the Banach space X is p-smooth and 2-convex and $R(u) = \frac{1}{p} \|u\|_X^p$ for some $1 < p < 2$. Moreover, suppose that Assumption 3.10 is satisfied with some orders $r_1, r_2 \geq 0$ and $r_1 < 1$. Then for the choice $\alpha \simeq \delta^\kappa$ where*

$$\kappa = \begin{cases} \frac{\nu_1 \nu_2}{\nu_1 \nu_2 + q(r_2 - r_1)} & \text{for } r_1 \leq r_2 \text{ and} \\ 1 & \text{for } r_2 < r_1 < 1 \end{cases}$$

we have convergence

$$D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \leq \begin{cases} C_p \delta^{\frac{2\nu_2(1-r_1)}{\nu_1\nu_2+q(r_2-r_1)}} & \text{for } r_1 \leq r_2 \text{ and} \\ C_p \delta^{\frac{2(1-r_1)}{2+r_1(q-2)}} & \text{for } r_2 < r_1 < 1. \end{cases}$$

Above we have denoted $\nu_i = 2 + r_i(q-2)$ and $q = \frac{p}{p-1}$. For the constant C_p we have $C_p \to \infty$ when $p \to 2$.

Proof. We can apply the Xu–Roach inequality II [61, Thm. 2.40 (b)] in X to obtain

$$D_R^{\mu_u, \mu_v}(u, v) = \langle j_p(u) - j_p(v), u - v \rangle_{X^* \times X} \geq C \max\{\|u\|_X, \|v\|_X\}^{p-2} \|u-v\|_X^2.$$

This gives us an estimate

$$R(u - v) \leq C_{\frac{p}{2}}(u, v) D_R^{\mu_u, \mu_v}(u, v)^{\frac{p}{2}}$$

with

$$C_{\frac{p}{2}}(u, v) = \frac{C}{p} \max\{\|u\|_X, \|v\|_X\}^{\frac{p(2-p)}{2}}.$$

By applying the trivial upper bound $\max\{\|u\|_X, \|v\|_X\} \leq \|u\|_X + \|v\|_X$ and the a priori bound given in (2.7) for any $\gamma \in (0, 1)$ we have

$$C_{\frac{p}{2}}(u_\alpha^\delta, u^\dagger) \leq C\left(\|u^\dagger\|_X^p + \frac{\delta}{\alpha} e_{\frac{\delta}{2}, \frac{\alpha\gamma}{\delta}}(\eta))\right)^{\frac{2-p}{2}}.$$

Considering Theorem 2.9 we now obtain

$$D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \lesssim \inf_{(\zeta_1, \zeta_2) \in \mathbb{R}_+^2} \left\{ \left(\zeta_1 + \frac{\delta}{\alpha}\zeta_2\right)^{\frac{2}{2-p}} C_{\frac{p}{2}}(u_\alpha^\delta, u^\dagger)^{\frac{2}{2-p}} + \frac{2}{2-p}\left(e_{\alpha, \zeta_1}(\mu^\dagger) + \frac{\delta}{\alpha} e_{\delta, \zeta_2}(\eta)\right) \right\}$$

$$\lesssim M_1 + M_2. \tag{3.10}$$

Above we have for $s = \frac{2}{2-p}$ that

$$M_1 = \inf_{\zeta_1 \in \mathbb{R}_+} \left(\zeta_1^s \left(1 + \frac{\delta}{\alpha} e_{\frac{\delta}{2}, \frac{\alpha\gamma}{\delta}}(\eta)\right) + e_{\alpha, \zeta_1}(\mu^\dagger) \right)$$

and

$$M_2 = \inf_{\zeta_2 \in \mathbb{R}_+} \left(\zeta_2^s \left(\frac{\delta}{\alpha}\right)^s \left(1 + \frac{\delta}{\alpha} e_{\frac{\delta}{2}, \frac{\alpha\gamma}{\delta}}(\eta)\right) + \frac{\delta}{\alpha} e_{\delta, \zeta_2}(\eta) \right).$$

Since $R(u) = \frac{1}{p}\|u\|_X^p$ we can write

$$e_{\alpha, \zeta}(\eta) = \inf_{w \in Y} \left\{ \zeta R^\star\left(\frac{1}{\zeta}(K^*w - \eta)\right) + \frac{\alpha}{2}\|w\|_Y^2 \right\}$$

$$= \alpha \inf_{w \in Y} \left\{ \frac{1}{q\alpha\zeta^{q-1}}\|K^*w - \eta\|_{X^*}^q + \frac{1}{2}\|w\|_Y^2 \right\}.$$

19

From Assumption 3.10 we directly obtain following estimates:

$$
\begin{aligned}
e_{\frac{\delta}{2},\frac{\alpha\gamma}{\delta}}(\eta) &\lesssim \gamma^{-t_2}\delta^{1-r_2+t_2}\alpha^{-t_2}, \\
e_{\alpha,\zeta_1}(\mu^{\dagger}) &\lesssim \alpha^{1-r_1}\zeta_1^{-t_1}, \\
e_{\delta,\zeta_2}(\eta) &\lesssim \delta^{1-r_2}\zeta_2^{-t_2},
\end{aligned}
\tag{3.11}
$$

where we have set $t_i = (q-1)r_i \geq 0$. By assuming further that $\alpha \simeq \delta^{\kappa}$ for some $\kappa > 0$, we can reduce the two upper-most estimates to

$$
e_{\frac{\delta}{2},\frac{\alpha\gamma}{\delta}}(\eta) \lesssim \delta^{1-r_2+(1-\kappa)t_2} \quad \text{and} \quad e_{\alpha,\zeta_1}(\mu^{\dagger}) \lesssim \delta^{\kappa(1-r_1)}\zeta_1^{-t_1}.
$$

Applying all the estimates above to Bregman distance in (3.10) we get

$$
\begin{aligned}
D_R^{\mu_{\alpha}^{\delta},\mu^{\dagger}}(u_{\alpha}^{\delta}, u^{\dagger}) &\lesssim \inf_{\zeta_1 \in \mathbb{R}_+} \left\{ (1+\delta^{r_3})\zeta_1^s + \delta^{\kappa(1-r_1)}\zeta_1^{-t_1} \right\} + \\
&\quad \inf_{\zeta_2 \in \mathbb{R}_+} \left\{ \delta^{(1-\kappa)s}(1+\delta^{r_3})\zeta_2^s + \delta^{2-\kappa-r_2}\zeta_2^{-t_2} \right\}
\end{aligned}
$$

where we have assumed that $r_3 = 1 - r_2 + (1-\kappa)(t_2+1) \geq 0$. Further, applying Lemma 3.3 yields us

$$
M_1 \simeq \delta^{\kappa(1-r_1)\frac{s}{s+t_1}}
$$

and

$$
M_2 \simeq \delta^{(1-\kappa)\frac{st_2}{s+t_2}}\delta^{(2-\kappa-r_2)\frac{s}{s+t_2}}.
$$

Consequently, we can reduce the estimate to

$$
\begin{aligned}
M_1 + M_2 &\simeq \delta^{\frac{2\kappa(1-r_1)}{2+(q-2)r_1}} + \delta^{\frac{2(2+r_2(q-2)-\kappa(r_2(q-1)+1))}{2+(q-2)r_2}} \\
&= \delta^{\frac{2\kappa(1-r_1)}{\nu_1}} + \delta^{\frac{2(\nu_2-\kappa(r_2(q-1)+1))}{\nu_2}}
\end{aligned}
$$

where $\nu_i = 2 + r_i(q-2)$.

When $r_1 \leq r_2$ the above expression is minimized at

$$
\kappa = \frac{\nu_1\nu_2}{\nu_1\nu_2 + q(r_2 - r_1)}
$$

yielding convergence rate

$$
D_R^{\mu_{\alpha}^{\delta},\mu^{\dagger}}(u_{\alpha}^{\delta}, u^{\dagger}) \leq C_p \delta^{\frac{2\nu_2(1-r_1)}{\nu_1\nu_2+q(r_2-r_1)}}.
$$

In order to attain convergence we have to assume $r_1 < 1$. If $r_2 < r_1 < 1$ the optimal convergence rate is achieved when $\kappa = 1$ and we obtain

$$
D_R^{\mu_{\alpha}^{\delta},\mu^{\dagger}}(u_{\alpha}^{\delta}, u^{\dagger}) \leq C_p \delta^{\min\left\{\frac{2(1-r_1)}{2+r_1(q-2)}, \frac{2(1-r_2)}{2+r_2(q-2)}\right\}} = C_p \delta^{\frac{2(1-r_1)}{2+r_1(q-2)}}.
$$

Above the constant $C_p \to \infty$ when $p \to 2$. Note that with the chosen κ the assumption $r_3 \geq 0$ is always true when $r_1 < 1$. $\quad\square$

Corollary 3.12. *If in addition to the assumptions of Theorem 3.11 the exact source condition $r_1 = 0$ is satisfied the estimate*

$$D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \le C_p \delta^{\frac{2\nu_2}{2\nu_2 + qr_2}}$$

holds. Furthermore, notice that assuming an exact source condition on the noise leads to the standard convergence rate of $\mathcal{O}(\delta)$ in the classical setting [61].

Remark 3.13. Let us illustrate another bound for $R(u-v)$ obtained via the Xu–Roach inequalities. Since X^* is q-convex and 2-smooth [61, Thm 2.52 (b)] we can apply [61, Lemma 2.63] and the Xu–Roach inequality II [61, Thm. 2.40 (b)] in X^* to obtain

$$D_R^{\mu_u, \mu_v}(u, v) = D_{R^*}^{u,v}(\mu_u, \mu_v) \ge C \max\{\|\mu_u\|_{X^*}, \|\mu_v\|_{X^*}\}^{q-q} \|\mu_u - \mu_v\|_{X^*}^q = C \|\mu_u - \mu_v\|_{X^*}^q.$$

Next by Xu–Roach inequality IV [61, Thm. 2.42] we obtain

$$\|\mu_u - \mu_v\|_{X^*} \ge C \max\{\|\mu_u\|_{X^*}, \|\mu_v\|_{X^*}\}^{2-q} \|u - v\|_X,$$

where we have considered the inequality in X^* which is 2-smooth by assumption.

Combining the two inequalities above yields

$$R(u - v) \le \frac{C}{p} \max\{\|u\|_X, \|v\|_X\}^{p(2-p)} D_R^{\mu_u, \mu_v}(u, v)^{\frac{p}{q}}$$

since $p - q + (2 - p)q = 0$.

Case $p = 2$

Finally we simplify the estimates in the quadratic case:

Theorem 3.14. *Suppose that X is a Banach space and $R(u) = \frac{1}{2}\|u\|_X^2$. Moreover, suppose that Assumption 3.10 is satisfied with some orders $r_1, r_2 \ge 0$ and $r_1 < 1$. For the choice $\alpha \simeq \delta^\kappa$, where*

$$\kappa = \begin{cases} \frac{2}{2 + r_2 - r_1} & \text{for } r_1 \le r_2 \text{ and} \\ 1 & \text{for } r_2 < r_1 < 1, \end{cases}$$

we get convergence

$$D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \lesssim \begin{cases} \delta^{\frac{2(1-r_1)}{2+r_2-r_1}} & \text{for } r_1 \le r_2 \text{ and} \\ \delta^{1-r_1} & \text{for } r_2 < r_1 < 1. \end{cases}$$

Proof. Recall from the Example 2.7 that case $R(u) = \frac{1}{2}\|u\|_X^2$ corresponds to parameter $\theta = 1$ and $C_2(u_\alpha^\delta, u^\dagger) = \frac{1}{2}$ in condition (2.12). Hence the second part of the Theorem 2.9 gives us

$$D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \lesssim \inf_{\zeta_1, \zeta_2 \in \Sigma} \frac{e_{\alpha, \zeta_1}(\mu^\dagger) + \frac{\delta}{\alpha} e_{\delta, \zeta_2}(\eta)}{2 - \zeta_1 - \frac{\delta}{\alpha}\zeta_2}$$

$$\lesssim \inf_{\zeta_1, \zeta_2 \in \Sigma} \frac{\zeta_1^{-r_1} \delta^{\kappa(1-r_1)} + \zeta^{-r_2} \delta^{2 - \kappa - r_2}}{2 - \zeta_1 - \frac{\delta}{\alpha}\zeta_2}$$

21

where $\zeta_1 + \frac{\delta}{\alpha}\zeta_2 < 2$ in Σ. If we choose $\zeta_1 = c < 1$ and $\zeta_2 = \frac{\alpha}{\delta}$ we can write

$$D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \lesssim \delta^{\kappa(1-r_1)} + \delta^{2-\kappa(1-r_2)}$$

where we need to assume $r_1 < 1$. The above convergence is optimized by $\kappa = \frac{2}{2+r_2-r_1}$ when $r_1 \le r_2$ in which case

$$D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \lesssim \delta^{\frac{2(1-r_1)}{2+r_2-r_1}}.$$

If $r_1 > r_2$ then we choose $\kappa = 1$ which gives us convergence

$$D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \lesssim \delta^{1-r_1}.$$

$\square$

Corollary 3.15. *If we assume that $u^\dagger$ fulfills the source condition, that is, $r_1 = 0$ we get convergence*

$$D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \lesssim \delta^{\frac{2}{2+r_2}}.$$

Case $p > 2$

Theorem 3.16. *Suppose that X is a p-convex Banach space with some $p > 2$ and $R(u) = \frac{1}{p}\|u\|_X^p$. Moreover, suppose that Assumption 3.10 is satisfied with some orders $r_1, r_2 \ge 0$ and $r_1 < 1$. For the choice $\alpha \simeq \delta^\kappa$, where*

$$\kappa = \begin{cases} \frac{2}{2+r_2-r_1} & \text{for } r_1 \le r_2 \text{ and} \\ 1 & \text{for } r_2 < r_1 < 1, \end{cases}$$

we have convergence

$$D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) \le \begin{cases} C_p \delta^{\frac{2(1-r_1)}{2+r_2-r_1}} & \text{for } r_1 \le r_2 \text{ and} \\ C_p \delta^{1-r_1} & \text{for } r_2 < r_1. \end{cases}$$

Proof. We can give an alternative definition for the general Bregman distance by

$$\begin{aligned} D_R^{\mu_u}(u, v) &= \frac{1}{q}\|\mu_u\|_{X^*}^q - \langle \mu_u, v \rangle_{X^* \times X} + \frac{1}{p}\|v\|_X^p \\ &= (p-1)R(u) - \langle \mu_u, v \rangle_{X^* \times X} + R(v) \end{aligned} \tag{3.12}$$

where $\mu_u \in \partial R(u)$. We get same kind of estimate for the Bregman distance as in [4]

$$\begin{aligned} D_R^{\mu_u}(u, v) &= \left(1 - \frac{1}{p}\right)\|u\|_X^p - \langle \mu_u, v \rangle_{X^* \times X} + \frac{1}{p}\|v\|_X^p \\ &= \frac{1}{p}\|u - (u - v)\|_X^p - \frac{1}{p}\|u\|_X^p + \langle \mu_u, u - v \rangle_{X^* \times X} \\ &\ge \frac{C_p}{p}\|u - v\|_X^p. \end{aligned}$$

22

The last estimate above is given by the Xu-Roach inequalities [66]. The Bregman distance given by (3.12) coincides with our previous definition (2.2)

$$
\begin{aligned}
D_R^{\mu_u,\mu_v}(u,v) &= D_R^{\mu_u}(u,v) + D_R^{\mu_v}(v,u) \\
&= p(R(u) + R(v)) + \langle \mu_u - \mu_v, u - v \rangle_{X^* \times X} - \|u\|_X^p - \|v\|_X^p \\
&= \langle \mu_u - \mu_v, u - v \rangle_{X^* \times X},
\end{aligned}
$$

for any $\mu_u \in \partial R(u)$ and $\mu_v \in \partial R(v)$. Hence we get an estimate

$$
R(u - v) \leq C_p D_R^{\mu_u,\mu_v}(u,v).
$$

That is, (2.12) holds with $\theta = 1$ and $C_\theta(u,v) = C_p$. Hence when $p > 2$ we get the same convergence rate as in case $p = 2$. $\qquad\square$

Remark 3.17. It is straightforward to see that polynomial decay of the distance function [61] implies an approximate source condition in Assumption 3.10. Suppose we have

$$
d_\rho(\mu^\dagger) = \inf_{w \in Y} \{ \|K^*w - \mu^\dagger\|_{X^*} \mid \|w\|_Y \leq \rho \} \leq \rho^{-k},
$$

where $k > 0$. This yields an estimate

$$
\begin{aligned}
e_{\alpha,\zeta}(\mu^\dagger) &= \inf_{w \in Y} \left\{ \zeta R^\star \left(\frac{1}{\zeta}(K^*w - \mu^\dagger) \right) + \frac{\alpha}{2}\|w\|_Y^2 \right\} \\
&\simeq \inf_{\rho > 0} \left\{ \zeta^{1-q}\rho^{-kq} + \delta^\kappa \rho^2 \right\} \\
&\simeq \delta^{\frac{kq\kappa}{kq+2}} \zeta^{-\frac{2(q-1)}{kq+2}}.
\end{aligned}
$$

Choosing $k = \frac{2(1-r_1)}{r_1 q}$, where $r_1 \in [0,1)$, we see that the last estimate above can be written

$$
e_{\alpha,\zeta}(\mu^\dagger) \simeq \delta^{\kappa(1-r_1)} \zeta^{-(q-1)r_1}
$$

which corresponds to the estimate given by Assumption 3.10 and (3.11).

3.3 Hilbert Space Embedding

Since many estimates are crucially simplified by using Hilbert space structures, we discuss in the following an approach to obtain (possibly suboptimal) rates deduced from the results above using embedding. We consider the case where R is the p-th power of a norm in a Banach space, with $p \geq 1$, and there exists a continuous embedding into a Hilbert space X_0. Indeed, we can assume the slightly weaker condition

$$
R(u) \geq C\|u\|_{X_0}^p \tag{3.13}
$$

for all $u \in X$. Note that by extending R as infinite outside X we can also state the same condition for arbitrary $u \in X_0$. Obviously the case $p = 1$ is of particular interest

here to cover e.g. total variation regularization (with the obvious embedding into L^2 for dimension less or equal two) and sparsity regularization (with the obvious embedding of ℓ^1 into ℓ^2).

In order to reduce to a Hilbert space framework, we assume that K can be extended to X_0 and maps this space continuously to Z. Thus, $L = K^*K$ is a bounded self-adjoint operator on X_0 and thus has a spectral decomposition. In particular, we can formulate smoothness of a vector $\vartheta \in X_0$ with the condition

$$\vartheta = L^\mu \omega \tag{3.14}$$

for $\omega \in X_0$ and some $\mu \in (0, \frac{1}{2})$. We then use the relation $e_{\alpha,\zeta}(\vartheta) = -\inf_{v \in X} F_{\alpha,\zeta}(v; \vartheta)$ and estimate $F_{\alpha,\zeta}(v; \vartheta)$ from below. For this sake we use (3.13) and (3.14) to get

$$
\begin{aligned}
F_{\alpha,\zeta}(v; \vartheta) &= \frac{1}{2\alpha}\|Kv\|_Y^2 - \langle \vartheta, v \rangle_{X^* \times X} + \zeta R(v) \\
&\geq \frac{1}{2\alpha}\|L^{\frac{1}{2}}v\|_{X_0}^2 - \|L^\mu v\|_{X_0}\|\omega\|_{X_0} + \zeta C\|v\|_{X_0}^p.
\end{aligned}
$$

Using the interpolation inequality

$$\|L^\mu v\|_{X_0} \leq \|L^{\frac{1}{2}}v\|_{X_0}^{2\mu}\|v\|_{X_0}^{1-2\mu}$$

and Young's inequality we get estimate

$$F_{\alpha,\zeta}(v; \vartheta) \geq -C\|\omega\|_{X_0}^{\frac{p}{p-1+2\mu-p\mu}}\zeta^{-\frac{1-2\mu}{p-1+2\mu-p\mu}}\alpha^{\frac{p\mu}{p-1+2\mu-p\mu}}$$

for some constant C independent of v, ζ, and α, which directly yields an upper bound for $e_{\alpha,\zeta}(\vartheta)$. We mention that in the case $p = 1$ we obtain

$$e_{\alpha,\zeta}(\vartheta) \leq C\|\omega\|_{X_0}^{\frac{1}{\mu}}\zeta^{-\frac{1-2\mu}{\mu}}\alpha. \tag{3.15}$$

4 Examples with random noise

4.1 Frequentist framework

Let us recall that our work above towards unbounded noise was mostly motivated by random noise models, especially, the statistics of white noise. It is hence natural to reinterpret the results of Theorem 2.9 as pointwise estimates for a random variable U_α^δ, which arises due to the randomness of the noise N. In the frequentist settings one is interested in the model

$$F^\delta = Ku^\dagger + \delta N, \tag{4.1}$$

where the data F^δ are generated by a deterministic true solution $u^\dagger$. In (4.1) the measurement $F^\delta = F^\delta(\omega)$ and the noise $N = N(\omega)$ are thought to be random variables. Here $\omega \in \Omega$ is an element of a complete probability space $(\Omega, \Sigma, \mathbb{P})$.

Following the idea in the earlier sections we consider a general frequentist risk denoted by E_B between the estimator $U_\alpha^\delta = U_\alpha^\delta(\omega)$ and $u^\dagger$. Here, our error measure is given by the Bregman distance

$$E_B(U_\alpha^\delta, u^\dagger) = \mathbb{E}\big(D_R^{\mu_\alpha^\delta, \mu^\dagger}(U_\alpha^\delta, u^\dagger)\big). \tag{4.2}$$

From the previous section we directly obtain a bound

$$E_B(U_\alpha^\delta, u^\dagger) = \mathbb{E}\bigg\{ \inf_{(\zeta_1, \zeta_2) \in (\mathbb{R}^+)^2} \bigg(\Big(\zeta_1 + \frac{\delta}{\alpha}\zeta_2\Big)^{1/(1-\theta)} C_\theta(U_\alpha^\delta, u^\dagger)^{1/(1-\theta)} +$$

$$\frac{1}{1-\theta} e_{\alpha,\zeta_1}(\mu^\dagger) + \frac{\delta}{\alpha(1-\theta)} e_{\delta,\zeta_2}(K^*N) \bigg) \bigg\}.$$

A canonical example of frequentist risk (4.2) is the mean integrated squared error (MISE)

$$E_B(U_\alpha^\delta, u^\dagger) = \mathbb{E}\|U_\alpha^\delta - u^\dagger\|_X^2,$$

where a quadratic regularization term $R(u) = \|u\|_X^2$ is assumed. Convergence rates of MISE have been widely studied in the literature, see [12, 13].

We observe that a finite estimate can only be obtained if $\mathbb{E}(e_{\delta,\zeta}(K^*N)) < \infty$ at least for some $\zeta > 0$. Under the typical choices of R the finiteness for any δ and ζ is obtained if

$$\mathbb{E}(e_{1,1}(K^*N)) < \infty.$$

This condition can be interpreted as an abstract smoothing condition for the operator K, as we shall see it can be identified with K being a trace-class operator.

In order to choose optimal parameters we first have to clarify which of them are random. Since ζ_2 is an auxiliary parameter appearing in the estimates only, not affecting any computation, it can be optimized in dependence of K^*N and hence it also becomes a random variable. The situation is less obvious with respect to α. Indeed it turns out that the question is exactly related to the issue of a-priori vs. a-posteriori parameter choice in the deterministic setup (cf. [26]). The *a-priori parameter choice* $\alpha = \alpha(\delta)$ leads to a parameter independent of the realization of the noise N, while the *a-posteriori parameter choice* $\alpha = \alpha(\delta, F)$ makes the parameter a random variable of N. Since the specific choices of α rely on the form of the regularization functional, we shall further investigate the general risk (4.2) in three very prominent cases, the classical one of Tikhonov regularisation (two-homogeneous R), the more general regularisation with Besov penalty and the popular total variation regularisation.

4.2 Gaussian case

Let us review the implications of our results in the canonical special case of a squared norm based regularization penalty $R(u) = \frac{1}{2}\|u\|_X^2$ for $X = Y = L^2(\mathbb{T}^d)$. We assume that N is generalized white noise statistics in $\mathcal{D}'(\mathbb{T}^d)$, that is, we have $\mathbb{E}N = 0$ and

$$\mathbb{E}\langle N, \phi\rangle_{\mathcal{D}' \times \mathcal{D}}\langle N, \psi\rangle_{\mathcal{D}' \times \mathcal{D}} = \langle \phi, \psi\rangle_{\mathcal{D}' \times \mathcal{D}}$$

for any test functions $\phi, \psi \in C^\infty(\mathbb{T}^d)$, where $\langle \cdot, \cdot \rangle_{\mathcal{D}' \times \mathcal{D}}$ denotes the duality pairing. It is well-known that the realizations of N belong to $Z^* = H^{-d/2-\epsilon}(\mathbb{T}^d)$ almost surely for any $\epsilon > 0$. For a sharp result, see [64]. We want to concentrate on the phenomena appearing due to large noise and hence assume an exact source condition for the true unknown $u^\dagger$ in the following.

In this example, two factors simplify our analysis remarkably. First, the symmetric Bregman distance coincides with the squared norm (as discussed in Example 2.7)

$$D_R^{\mu_u, \mu_v}(u, v) = \|u - v\|_{L^2(\mathbb{T}^d)}^2.$$

Secondly, the term $e_{\alpha, \zeta}(K^*N)$ can be explicitly estimated since

$$
\begin{aligned}
e_{\alpha, \zeta}(K^*N) &= \inf_{w \in L^2(\mathbb{T}^d)} \left(\zeta R^\star \left(\frac{K^*W - K^*N}{\zeta} \right) + \frac{\alpha}{2} \|W\|_{L^2(\mathbb{T}^d)}^2 \right) \\
&= \frac{1}{2\zeta} \inf_{w \in L^2(\mathbb{T}^d)} \left(\|K^*W - K^*N\|_{L^2(\mathbb{T}^d)}^2 + \alpha\zeta \|W\|_{L^2(\mathbb{T}^d)}^2 \right). \quad (4.3)
\end{aligned}
$$

Let us record the following short calculation as a lemma. For precise notation, let use denote $\mathcal{R}_\beta = (K^*K + \beta I)^{-1} : L^2(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$, $\beta > 0$, to highlight the restriction of K^* (and K) to $X = Y = L^2(\mathbb{T}^d)$.

Lemma 4.1. *Consider K as a bounded linear operator $K : L^2(\mathbb{T}^d) \to H^t(\mathbb{T}^d)$ for $t > d/2$. Then it follows that*

$$e_{\alpha, \zeta}(K^*N) = \frac{\alpha}{2} \langle N, K\mathcal{R}_{\alpha\zeta} K^*N \rangle_{H^{-t}(\mathbb{T}^d) \times H^t(\mathbb{T}^d)}$$

and

$$\mathbb{E} e_{\alpha, \zeta}(K^*N) = \frac{\alpha}{2} \mathrm{Tr}_{L^2(\mathbb{T}^d)}(K\mathcal{R}_{\alpha\zeta} K^*). \quad (4.4)$$

Proof. The minimizing estimator of problem (4.3) is given by $W_{\alpha\zeta} = K\mathcal{R}_{\alpha\zeta} K^*N$. Hence we can write

$$\|K^*W_{\alpha\zeta} - K^*N\|_{L^2(\mathbb{T}^d)}^2 + \alpha\zeta \|W_{\alpha\zeta}\|_{L^2(\mathbb{T}^d)}^2 = \alpha\zeta \langle N, K\mathcal{R}_{\alpha\zeta} K^*N \rangle_{H^{-t}(\mathbb{T}^d) \times H^t(\mathbb{T}^d)},$$

where $t > d/2$ It is well-known that N as white noise has a series representation $N = \sum_{j=1}^\infty N_j \psi_j$ almost surely, where $N_j \sim \mathcal{N}(0, 1)$ are i.i.d. and $\{\psi_j\}_{j=1}^\infty$ constitutes any orthonormal basis of $L^2(\mathbb{T}^d)$. The claim (4.4) now follows easily by applying the series representation together with independence of N_i and N_j for $i \neq j$. $\square$

Let us mention that the quantity on the right-hand side of the estimate (4.4),

$$\mathrm{Tr}_{L^2(\mathbb{T}^d)}(K\mathcal{R}_{\alpha\zeta} K^*) = \mathrm{Tr}_{L^2(\mathbb{T}^d)}((KK^* + \alpha\zeta I)^{-1}KK^*),$$

is known as the *effective dimension* in literature (cf. [67]). In the finite dimensional case it is between zero (as $\alpha\zeta \to \infty$) and the rank of KK^* (as $\alpha\zeta \to 0$). In the following we use a conservative estimate of the effective dimension in order to illustrate the results,

optimal estimates can be achieved under special assumptions, which is beyond our scope (cf. e.g. [48]). Our analysis in the non-Gaussian case indicates that $\mathbb{E}e_{\alpha,\zeta}(K^*N)$ is the basis for understanding a generalization of effective dimension for such, its analysis is a possibly important question for future research.

Theorem 4.2. *Assume that $K : L^2(\mathbb{T}^d) \to H^t(\mathbb{T}^d)$, where $t > d/2$, is a Hilbert–Schmidt operator in $L^2(\mathbb{T}^d)$ and $R(u) = \frac{1}{2}\|u\|_{L^2(\mathbb{T}^d)}^2$. When the true unknown $u^\dagger$ fulfills the exact source condition $\mu^\dagger = K^*w^\dagger$, where $w^\dagger \in L^2$, we obtain the convergence rate*

$$E_B(U_\alpha^\delta, u^\dagger) = \mathbb{E}\left\|U_\alpha^\delta - u^\dagger\right\|_{L^2(\mathbb{T}^d)}^2 \lesssim \delta^{2/3},$$

with choice $\alpha \simeq \delta^{2/3}$.

Proof. Considering (1.4) where we have now $u = \mu \in \partial R(u)$. Therefore, we can write

$$K^*(KU_\alpha^\delta - (f + \delta N)) + \alpha U_\alpha^\delta = 0$$

and consequently

$$K^*(KU_\alpha^\delta - Ku^\dagger) + \alpha(U_\alpha^\delta - u^\dagger) = \delta K^*N - \alpha K^*w^\dagger \tag{4.5}$$

where $u^\dagger = \mu^\dagger = K^*w^\dagger$ for $w^\dagger \in L^2(\mathbb{T}^d)$. Taking duality product of $U_\alpha^\delta - u^\dagger$ and equation (4.5) yields

$$\|KU_\alpha^\delta - Ku^\dagger\|_{L^2(\mathbb{T}^d)}^2 + \alpha\|U_\alpha^\delta - u^\dagger\|_{L^2(\mathbb{T}^d)}^2 = \delta\langle K^*N, U_\alpha^\delta - u^\dagger\rangle_{L^2(\mathbb{T}^d)} + \alpha\langle w^\dagger, K(u^\dagger - U_\alpha^\delta)\rangle_{L^2(\mathbb{T}^d)}. \tag{4.6}$$

We will approximate the right hand side terms separately. Following the idea behind the estimate (2.11) we bound the first term on the right hand side of (4.6) by

$$\delta\langle K^*N, U_\alpha^\delta - u^\dagger\rangle_{L^2(\mathbb{T}^d)} = \delta\langle K^*N - K^*W, U_\alpha^\delta - u^\dagger\rangle_{L^2(\mathbb{T}^d)} + \delta\langle W, K(U_\alpha^\delta - u^\dagger)\rangle_{L^2(\mathbb{T}^d)}$$

$$\leq \frac{\delta\zeta_2}{2}\|K^*N - K^*W\|_{L^2(\mathbb{T}^d)}^2 + \frac{\delta}{2\zeta_2}\|U_\alpha^\delta - u^\dagger\|_{L^2(\mathbb{T}^d)}^2 + \frac{\delta^2}{2}\|W\|_{L^2(\mathbb{T}^d)}^2 + \frac{1}{2}\|K(U_\alpha^\delta - u^\dagger)\|_{L^2(\mathbb{T}^d)}^2$$

for any $\zeta_2 > 0$. For the last term in (4.6) we have

$$\alpha\langle w^\dagger, K(u^\dagger - U_\alpha^\delta)\rangle_{L^2(\mathbb{T}^d)} \leq \frac{\alpha^2}{2}\|w^\dagger\|_{L^2(\mathbb{T}^d)}^2 + \frac{1}{2}\|K(U_\alpha^\delta - u^\dagger)\|_{L^2(\mathbb{T}^d)}^2.$$

Since $w = W(\omega) \in L^2(\mathbb{T}^d)$ is arbitrary, using the estimates above we get

$$\begin{aligned} \|U_\alpha^\delta - u^\dagger\|_{L^2(\mathbb{T}^d)}^2 &\leq \frac{1}{\alpha - \frac{\delta}{2\zeta_2}}\left\{\inf_{w\in L^2(\mathbb{T}^d)}\left(\frac{\delta\zeta_2}{2}\|K^*N - K^*w\|_{L^2(\mathbb{T}^d)}^2 + \frac{\delta^2}{2}\|w\|_{L^2(\mathbb{T}^d)}^2\right) + \frac{\alpha^2}{2}\|w^\dagger\|_{L^2(\mathbb{T}^d)}^2\right\} \\ &\leq \frac{2\delta}{\alpha}e_{\delta,\frac{\alpha}{\delta}}(K^*N) + \alpha\|w^\dagger\|_{L^2(\mathbb{T}^d)}^2. \end{aligned} \tag{4.7}$$

Above, we obtained the last estimate by choosing $\zeta_2 = \frac{\delta}{\alpha}$.

In order to derive a convergence rate we point out that $\mathcal{R}_\beta$ is a self-adjoint semipositive-definite bounded linear operator satisfying $\left\|\mathcal{R}_\beta^{1/2}\right\|_{L^2(\mathbb{T}^d)\to L^2(\mathbb{T}^d)} \le \frac{1}{\sqrt{\beta}}$ and consequently

$$\mathbb{E}e_{\delta,\frac{\alpha}{\delta}}(K^*N) = \frac{\delta}{2}\mathrm{Tr}_{L^2(\mathbb{T}^d)}(K\mathcal{R}_\alpha K^*) \le \frac{\delta}{2\alpha}\mathrm{Tr}_{L^2(\mathbb{T}^d)}(KK^*). \qquad (4.8)$$

Since K^* is a Hilbert-Schmidt operator, we have $\mathrm{Tr}_{L^2(\mathbb{T}^d)}(KK^*) < \infty$.

Now it follows from equations (4.7) and (4.8) that

$$\mathbb{E}\|U_\alpha^\delta - u^\dagger\|_{L^2(\mathbb{T}^d)}^2 \le \frac{\delta^2}{\alpha^2}\mathrm{Tr}_{L^2(\mathbb{T}^d)}(KK^*) + \alpha\|w^\dagger\|_{L^2(\mathbb{T}^d)}^2. \qquad (4.9)$$

The bound in (4.9) is optimised by choosing $\alpha \simeq \delta^{2/3}$, which also yields the claim. $\qquad\square$

From the previous theorem we see that the assumption of finite trace of KK^* : $L^2(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$ is indeed equivalent to the condition

$$\mathbb{E}(e_{\delta,\gamma}(K^*N)) < \infty \qquad (4.10)$$

for some $\delta, \gamma > 0$ as well as to the condition

$$\mathbb{E}(\|K^*N\|_{L^2(\mathbb{T}^d)}^2) < \infty, \qquad (4.11)$$

which appears to be a natural requirement.

One can observe better convergence rates if faster decay of eigenvalues of KK^* is assumed.

Theorem 4.3. *Suppose that* $R(u) = \frac{1}{2}\|u\|_{L^2(\mathbb{T}^d)}^2$. *Moreover, assume that* $\{\lambda_j\}_{j=1}^\infty$ *are eigenvalues of* $KK^* : L^2(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$ *and there exists* $0 < m \le 1$ *such that*

$$\sum_{j=1}^\infty \lambda_j^m < \infty.$$

Then, when the true unknown $u^\dagger$ *fulfills the exact source condition* $\mu^\dagger = K^*w^\dagger$, *where* $w^\dagger \in L^2$, *it follows that for* $\alpha \simeq \delta^\kappa$, *where* $\kappa = \frac{2}{2+m}$, *we obtain*

$$E_B(U_\alpha^\delta, u^\dagger) = \mathbb{E}\left\|U_\alpha^\delta - u^\dagger\right\|_{L^2(\mathbb{T}^d)}^2 \lesssim \delta^{\frac{2}{2+m}}.$$

Proof. Suppose p and q are Hölder conjugates such that $m = \frac{1}{q}$. By applying Young's inequality to

$$p\alpha^{p/q}\lambda_j \le \alpha^p + \lambda_j^p \le (\alpha + \lambda_j)^p, \quad \alpha, \lambda_j \ge 0,$$

we obtain

$$(p\alpha^{p/q}\lambda_j)^{1/p} \le \alpha + \lambda_j. \qquad (4.12)$$

28

This yields

$$Tr_{L^2(\mathbb{T}^d)}(K\mathcal{R}_\alpha K^*) = \sum_{j=1}^\infty \frac{\lambda_j}{\lambda_j + \alpha} \leq \sum_{j=1}^\infty \frac{\lambda_j}{(p\alpha^{p/q}\lambda_j)^{1/p}} \leq \frac{1}{p^{1/p}\alpha^{1/q}} \sum_{j=1}^\infty \lambda_j^{1/q}.$$

Therefore, by equation (4.7) the frequentist risk is bounded by

$$\mathbb{E}\|U_\alpha^\delta - u^\dagger\|_{L^2(\mathbb{T}^d)}^2 \leq \frac{\delta^2}{p^{1/p}\alpha^{1/q+1}} \sum_{j=1}^\infty \lambda_j^{1/q} + \alpha\|w^\dagger\|_{L^2(\mathbb{T}^d)}^2.$$

The proof is concluded by optimizing $\alpha \simeq \delta^\kappa$. $\qquad\qquad\qquad\qquad\qquad\square$

As mentioned before the mean integrated squared error (MISE) of an estimator U_α^δ is defined

$$R(U_\alpha^\delta, u^\dagger) = \mathbb{E}\|U_\alpha^\delta - u^\dagger\|_{L^2}^2(\mathbb{T}^d). \qquad\qquad (4.13)$$

The minimax risk $r_\delta(H^r(\mathbb{T}^d), M)$ on the Sobolev space $H^r(\mathbb{T}^d)$ is then given by

$$r_\delta(H^r(\mathbb{T}^d), M) = \inf_{U_\alpha^\delta} \sup_{\|u^\dagger\|_{H^r(\mathbb{T}^d)} < M} R(U_\alpha^\delta, u^\dagger),$$

where the infimum is taken over all estimators of the form $U_\alpha^\delta = g(F^\delta)$. Here we have denoted $g \in \mathcal{B}(H^{-d/2-\epsilon}(\mathbb{T}^d), H^r(\mathbb{T}^d))$ where $\mathcal{B}(H^{-d/2-\epsilon}(\mathbb{T}^d), H^r(\mathbb{T}^d))$ is the set of Borel measurable functions from $H^{-d/2-\epsilon}(\mathbb{T}^d)$ to $H^r(\mathbb{T}^d)$. Next we will compare the convergence results of Theorems 4.2 and 4.3 to the known minimax convergence rates for same problems.

Remark 4.4. As an example of a group of operators that fills the conditions in Theorem 4.3 we can take bijective elliptic pseudodifferential operators that are $t > \frac{d}{2m}$ (where $m = 1$ in the case described in Theorem 4.2) orders smoothing e.g. $K = (I - \Delta)^{-\frac{t}{2}}$. We assume the exact source condition in the Theorems 4.2 and 4.3, that is, $u^\dagger = \mu^\dagger = K^*w^\dagger$, where $w^\dagger \in L^2$, hence we can conclude $u^\dagger \in H^r(\mathbb{T}^d)$, where $r = t$. This means we assume the minimum extra smoothness from $u^\dagger$, that is, the smoothness of the unknown and the order of smoothing of the forward operator are the same.

Since $r = t$ we can rewrite the convergence rate κ in form

$$\kappa = \frac{2}{2+m} = \frac{2r}{r+t+tm}.$$

Note that $tm = d/2 + \epsilon$ and hence the convergence rates achieved in Theorems 4.2 and 4.3 agree, up to $\epsilon > 0$ arbitrarily small, with the minimax convergence rate, see e.g. [12, 39].

4.3 Besov penalty

Suppose that functions $\{\psi_\ell\}_{\ell=1}^\infty$ form an orthonormal wavelet basis for $L^2(\mathbb{T})$ on the one-dimensional torus $\mathbb{T}$, where we have utilized global indexing. We can characterize the periodic Besov space $B_{pq}^s(\mathbb{T})$ using the given basis in the following way: the series

$$u(x) = \sum_{\ell=1}^\infty u_\ell \psi_\ell(x)$$

belongs to $B_{pq}^s(\mathbb{T})$ if and only if

$$2^{js}2^{j(\frac{1}{2}-\frac{1}{p})}\left(\sum_{\ell=2^j}^{2^{j+1}-1} |u_\ell|^p\right)^{1/p} \in \ell^q(\mathbb{N}). \tag{4.14}$$

We assume that the basis is r-regular for r large enough in order to provide a basis for a Besov space with smoothness s [19]. Here we are concerned with the special case $p = q$ and use abbreviation $B_p^s = B_{pp}^s$. It is well-known that an equivalent norm to (4.14) is given by

$$\left\|\sum_{\ell=1}^\infty u_\ell \psi_\ell\right\|_{B_p^s(\mathbb{T})} = \left(\sum_{\ell=1}^\infty \ell^{p(s+\frac{1}{2})-1}|u_\ell|^p\right)^{1/p}.$$

4.3.1 Case $p = 1$

Suppose that $X = Y = L^2(\mathbb{T})$ with orthogonal basis $\{\psi_\ell\}_\ell$. The noise N is assumed to have same statistics as in previous section. Here, we consider a regularization term given by

$$R(u) = \|u\|_{B_1^s(\mathbb{T})} = \sum_{\ell=1}^\infty \ell^{s-1/2}|u_\ell| \tag{4.15}$$

for $s \geq \frac{1}{2}$, where $u = \sum_{\ell=1}^\infty u_\ell \psi_\ell$. The dual norm in (3.1) is simply the norm of $B_\infty^{-s}(\mathbb{T})$, i.e.,

$$S(q) = \|q\|_{B_\infty^{-s}(\mathbb{T})} = \sup_{\ell \in \mathbb{N}} \ell^{1/2-s}|q_\ell|$$

for $q = \sum_{\ell=1}^\infty q_\ell \psi_\ell \in B_\infty^{-s}(\mathbb{T})$. Notice carefully that for parameters $s \geq \frac{1}{2}$ the functional R satisfies conditions (R1)-(R4) with the weak topology, since there is a continuous embedding from $B_1^s(\mathbb{T})$ to $L^2(\mathbb{T})$.

We notice that an arbitrary approximate source condition of type (3.4) requires a sparse structure of the true unknown as pointed out by the following lemma. Therefore, it does not cover a general class of unknowns for this 1-homogeneous example.

Lemma 4.5. *Let us assume that R is given by (4.15) and $K : L^2(\mathbb{T}) \to L^2(\mathbb{T})$ is such that for all $\ell \in \mathbb{N}$ there exists a function $w^{(\ell)} \in Y$ such that*

$$K^* w^{(\ell)} = e^{(\ell)}, \tag{4.16}$$

30

where $e^{(\ell)} = (0, \dots, 0, 1, 0, \dots)$ *is the infinite unit sequence with* 1 *at the ℓ-th position and* 0 *else. Then the following two statements are equivalent:*

(i) *The subgradient $\mu^\dagger$ satisfies the approximate source condition in Assumption 3.2 with some $r_1 \geq 0$.*

(ii) *The unknown $u^\dagger$ is non-zero only in a finite set of coefficient.*

Proof. Since R is defined as in (4.15) we see that the the subgradient is given by

$$\mu^\dagger_\ell = \begin{cases} \ell^{s-\frac{1}{2}} & \text{when } u^\dagger_\ell > 0 \\ -\ell^{s-\frac{1}{2}} & \text{when } u^\dagger_\ell < 0 \end{cases}$$

and $\mu^\dagger_\ell \in (-\ell^{s-\frac{1}{2}}, \ell^{s-\frac{1}{2}})$ when $u^\dagger_\ell = 0$.

Since $K^* : L^2 \to L^2$ we have $(K^*w)_\ell \to 0$, when $\ell \to \infty$. If $u^\dagger$ has infinitely many non-zero coefficients

$$S(\mu^\dagger - K^*w) = \sup_{\ell \in \mathbb{N}} \ell^{1/2-s} |\mu^\dagger_\ell - (K^*w)_\ell| \geq 1$$

and the approximate source condition (3.4) can not be satisfied. On the other hand if the unknown is non-zero only in a finite set of coefficient, that is, there exists such L that $u^\dagger_\ell = 0$ for $\ell > L$ we can choose subgradient $\mu^\dagger$ so that $\mu^\dagger_\ell = 0$ for $\ell > L$. Using assumption (4.16) we can then choose $w \in L^2$ so that $(K^*w)_\ell = \mu^\dagger_\ell$. Hence $S(\mu^\dagger - K^*w) = 0$ and the source condition (3.4) is fulfilled with any $r_1 \geq 0$. $\square$

As an example of group of operators that satisfies assumption (4.16) we can take operators with diagonal structure. The more general meaning of the assumption (4.16) has been studied for example in [27] and the references therein.

In addition to the above, we make an assumption on the smoothness of K and K^* by requiring that there exists a constant $C > 0$ and $t > \frac{1}{2}$ such that both satisfy

$$\frac{1}{C} \|\psi\|_{B_2^r} \leq \|K\psi\|_{B_2^{t+r}} \leq C \|\psi\|_{B_2^r} \tag{4.17}$$

(similar for K^*) for $r \in \mathbb{R}$ and $\psi \in B_2^r(\mathbb{T})$. The above smoothness condition enables a straightforward study of the noise terms, which allows us to deduce contraction rates.

Under the given assumptions we can write (recall equation (3.7))

$$e_{\delta,\zeta}(\eta) = \frac{\delta}{2} \inf_{w \in \mathcal{W}} \|w\|_Y^2,$$

with a fixed realization $\eta = K^*n = K^*N(\omega)$, where

$$\mathcal{W} = \left\{ w \in L^2(\mathbb{T}) \;\middle|\; \sup_{\ell \in \mathbb{N}} \ell^{1/2-s} \left| \langle n - w, K\psi_\ell \rangle_{B_2^{-t} \times B_2^t} \right| \leq \zeta \right\}.$$

Lemma 4.6. *Let us assume that* $K : L^2(\mathbb{T}) \to L^2(\mathbb{T})$ *satisfies condition* (4.17) *with a parameter* $t > \frac{1}{2}$ *and* R *is defined by* (4.15) *for* $s \geq \frac{1}{2}$. *Then it holds that*

$$\mathbb{E}e_{\delta,\zeta}(K^*N) \lesssim \delta\zeta^{-\frac{2}{2s+2t-1}}.$$

Proof. From condition (4.17) it follows that

$$e_{\delta,\zeta}(\eta) = \frac{\delta}{2}\inf_{w \in \mathcal{W}}\|w\|_{L^2}^2 \leq \frac{C\delta}{2}\inf_{w \in \mathcal{W}}\|K^*w\|_{B_2^t}^2$$

almost surely. The condition $w \in \mathcal{W}$ for a fixed realization $n = N(\omega)$ is equivalent to

$$|(K^*n)_\ell - (K^*w)_\ell| \leq \zeta\ell^{s-\frac{1}{2}} =: D_\ell$$

uniformly for all $\ell \in \mathbb{N}$ and hence

$$\inf_{w \in \mathcal{W}}\|K^*w\|_{B_2^t}^2 = \sum_{\ell=1}^{\infty}\inf_{|(K^*n)_\ell - (K^*w)_\ell| \leq D_\ell}\ell^{2t}|(K^*w)_\ell|^2 \leq \sum_{\ell=1}^{\infty}\ell^{2t}\max(|(K^*n)_\ell| - D_\ell, 0)^2.$$

Now $(K^*N)_\ell = \langle K^*N, \psi_\ell \rangle$ is a normally distributed random variable with zero mean and variance $\sigma_\ell^2 = \|K\psi_\ell\|_{L^2}^2 \simeq \|\psi_\ell\|_{B_2^{-t}}^2 \simeq \ell^{-2t}$ (according to (4.17)). Therefore, we have

$$\begin{aligned}
\mathbb{E}\inf_{w \in \mathcal{W}}\|w\|_{L^2}^2 &\leq \mathbb{E}\sum_{\ell=1}^{\infty}\ell^{2t}\max(|(K^*N)_\ell| - D_\ell, 0)^2 \\
&\lesssim \sum_{\ell=1}^{\infty}\frac{\ell^{2t}}{\sigma_\ell}\int_{D_\ell}^{\infty}(x - D_\ell)^2\exp\left(-\frac{x^2}{2\sigma_\ell^2}\right)dx \\
&\leq \sum_{\ell=1}^{\infty}\frac{\ell^{2t}}{\sigma_\ell}\int_0^{\infty}x^2\exp\left(-\frac{x^2}{2\sigma_\ell^2}\right)dx \cdot \exp\left(-\frac{D_\ell^2}{2\sigma_\ell^2}\right) \\
&\simeq \sum_{\ell=1}^{\infty}\sigma_\ell^2\ell^{2t}\exp\left(-\frac{D_\ell^2}{2\sigma_\ell^2}\right) \simeq \sum_{\ell=1}^{\infty}\exp\left(-\frac{1}{2}\zeta^2\ell^{2s+2t-1}\right).
\end{aligned}$$

Due to our assumptions on s and t we notice that the last sum converges. The sum above can be approximated as follows

$$\begin{aligned}
\sum_{\ell=1}^{\infty}\exp\left(-\frac{1}{2}\zeta^2\ell^{2s+2t-1}\right) &\simeq \int_0^{\infty}\exp\left(-\frac{1}{2}\left(\zeta^{\frac{2}{2s+2t-1}}x\right)^{2s+2t-1}\right)dx \\
&= \zeta^{\frac{-2}{2s+2t-1}}\int_0^{\infty}\exp\left(-\frac{1}{2}y^{2s+2t-1}\right)dy \simeq \zeta^{\frac{-2}{2s+2t-1}},
\end{aligned}$$

where we applied a change of variable $y = \zeta^{\frac{2}{2s+2t-1}}x$. This yields the claim. $\square$

Theorem 4.7. *Let us assume that $K : L^2(\mathbb{T}) \to L^2(\mathbb{T})$ satisfies conditions (4.16) and (4.17) with parameter $t > \frac{1}{2}$, R is defined by (4.15) for $s \geq \frac{1}{2}$ and $u^\dagger$ is supported on a finite number of coefficients. Then $\mu^\dagger$ satisfies an approximate source condition in Assumption 3.2 with any $r_1 \geq 0$. For the choice $\alpha \simeq \delta^\kappa$, where*

$$\kappa = \frac{2s + 2t}{2s + 2t + 1},$$

we obtain the convergence rate

$$\mathbb{E} D_R^{\mu_\alpha^\delta, \mu^\dagger}(U_\alpha^\delta, u^\dagger) \lesssim \delta^\kappa.$$

Proof. First, in equation (3.8) we apply

$$
\begin{aligned}
\mathbb{E} D_R^{\mu_\alpha^\delta, \mu^\dagger}(U_\alpha^\delta, u^\dagger) &\leq \inf_{\zeta_1 \in \mathbb{R}_+} \left(\zeta_1 \left(1 + \frac{\delta}{\alpha} \mathbb{E} e_{\delta, \frac{\alpha\gamma}{\delta}}(K^*N) \right) + e_{\alpha, \zeta_1}(\mu^\dagger) \right) \\
&\quad + \inf_{\zeta_2 \in \mathbb{R}_+} \left(\frac{\delta}{\alpha} \zeta_2 \left(1 + \frac{\delta}{\alpha} \mathbb{E} e_{\delta, \frac{\alpha\gamma}{\delta}}(K^*N) \right) + \frac{\delta}{\alpha} \mathbb{E} e_{\delta, \zeta_2}(K^*N) \right) \\
&=: \ \widetilde{M}_1 + \widetilde{M}_2. \qquad\qquad (4.18)
\end{aligned}
$$

Notice that by Lemma 4.6 and assumption $\alpha = \delta^\kappa$, $\kappa \leq 1$, we have

$$\mathbb{E} e_{\delta, \frac{\alpha\gamma}{\delta}}(K^*N) \lesssim \gamma^{-s'} \delta^{1+(1-\kappa)s'} \lesssim 1$$

for a constant γ, where we denote $s' = \frac{2}{2s+2t-1} > 0$ for convenience. Therefore, we see as in the proof of Theorem 3.4 that

$$\widetilde{M}_1 \lesssim \inf_{\zeta_1 \in \mathbb{R}_+} \left\{ \zeta_1 + \delta^\kappa \zeta_1^{-r_1} \right\} \simeq \delta^{\frac{\kappa}{1+r_1}}$$

and

$$\widetilde{M}_2 \lesssim \inf_{\zeta_2 \in \mathbb{R}_+} \left\{ \delta^{1-\kappa} \zeta_2 + \delta^{2-\kappa} \zeta_2^{-s'} \right\} \simeq \delta^{(1-\kappa)\frac{s'}{1+s'} + (2-\kappa)\frac{1}{s'+1}} = \delta^{\frac{(1-\kappa)s' + 2 - \kappa}{1+s'}}.$$

The convergence rate is minimized for κ which satisfies

$$\frac{\kappa}{1+r_1} = \frac{(1-\kappa)s' + 2 - \kappa}{1+s'}.$$

Since r_1 can be chosen arbitrarily small, we conclude that

$$\kappa = \frac{1}{2} \cdot \frac{2 + s'}{1 + s'} = \frac{2s + 2t}{2s + 2t + 1}.$$

This concludes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

4.3.2 Case $1 < p \leq 2$

Let us set $X = L^2(\mathbb{T})$. We consider here the special case when the forward operator K in (1.1) can be diagonalized in the basis $\{\phi_\ell\}_\ell$, i.e., $\langle \phi_\ell, K\phi_{\ell'} \rangle = 0$, whenever $\ell \neq \ell'$. It follows that we can reduce our model to a countable number of independent equations

$$f_\ell = k_\ell u_\ell + \delta N_\ell$$

for $\ell \in \mathbb{N}$, where $f_\ell = \langle f, \phi_\ell \rangle$, $u_\ell = \langle u, \phi_\ell \rangle$, $k_\ell = \langle \phi_\ell, K\phi_\ell \rangle$ and the random variables $N_\ell = \langle N, \phi_\ell \rangle$ are normally distributed i.i.d. Similar to the case $p = 1$ we assume that K satisfies (4.17), which corresponds to assuming $k_\ell \simeq \ell^{-t}$ asymptotically with respect to ℓ.

Suppose that the regularization functional R is given by $R(u) = \frac{1}{p}\|u\|^p_{B^s_p(\mathbb{T})}$ for $1 < p < 2$ and $s \geq \frac{1}{p} - \frac{1}{2}$ so that $B^s_p(\mathbb{T})$ can be embedded continuously to X. The convex conjugate $R^\star$ satisfies

$$R^\star(u) = \frac{1}{q}\sum_{\ell=1}^{\infty} \ell^{q(-s+\frac{1}{2})-1}|u_\ell|^q = \frac{1}{q}\|u\|^q_{B^{-s}_q(\mathbb{T})},$$

where p and q are Hölder conjugates.

For the convenience of the reader, we assume that $\mu^\dagger$ satisfies the accurate source condition, i.e.,

$$\mu^\dagger = K^* w.$$

We can then write

$$\mathbb{E}D^{\mu^\delta_\alpha, \mu^\dagger}_R(u^\delta_\alpha, u^\dagger) \lesssim \mathbb{E}M_1 + \mathbb{E}M_2 \tag{4.19}$$

where $\mathbb{E}M_1 = \alpha$ and

$$M_2 = \inf_{\zeta_2 \in \mathbb{R}_+} \left(\zeta_2^{\frac{2}{2-p}} \left(\frac{\delta}{\alpha}\right)^{\frac{2}{2-p}} \left(1 + \frac{\delta}{\alpha}e_{\frac{\delta}{2}, \frac{\alpha\gamma}{\delta}}(K^*N)\right) + \frac{\delta}{\alpha}e_{\delta, \zeta_2}(K^*N) \right).$$

Lemma 4.8. *Let us assume that K and R are as above, $s \geq \frac{1}{p} - \frac{1}{2}$ and $t > \frac{1}{2}$. Then we can estimate*

$$\mathbb{E}e_{\delta, \zeta}(K^*N) \lesssim \delta^{1-1/r}\zeta^{\frac{1-q}{r}},$$

where $r = q(t + s - \frac{1}{2}) + 1 > \frac{q}{2}$.

Proof. By definition we have

$$
\begin{aligned}
e_{\delta, \zeta}(K^*N) &= \inf_{w \in Y}\left(\zeta R^\star\left(\frac{K^*(w-N)}{\zeta}\right) + \frac{\delta}{2}\|w\|^2_Y \right) \\
&= \frac{1}{2q}\inf_{w \in Y}\sum_{\ell=1}^{\infty}\left(2\zeta^{1-q}\ell^{q(-s+\frac{1}{2})-1}k_\ell^q|w_\ell - N_\ell|^q + \delta w_\ell^2 \right).
\end{aligned}
$$

Let us now abbreviate $a_\ell = 2\zeta^{1-q}\ell^{q(-s+\frac{1}{2})-1}k_\ell^q$ and consider an upper bound for the infimum by elements in Y that are supported only on the first L basis vectors. We find that

$$e_{\delta,\zeta}(K^*N) \lesssim \sum_{\ell=1}^{L} \min\{a_\ell|N_\ell|^q, \delta N_\ell^2\} + \sum_{\ell=L+1}^{\infty} a_\ell|N_\ell|^q$$

and, therefore,

$$\mathbb{E}e_{\delta,\zeta}(K^*N) \lesssim \sum_{\ell=1}^{L} \mathbb{E}\min\{a_\ell|N_\ell|^q, \delta N_\ell^2\} + \sum_{\ell=L+1}^{\infty} a_\ell \tag{4.20}$$

since $\mathbb{E}|N_\ell|^q \simeq 1$. In order to evaluate the expectation in (4.20) we need the following integral identity

$$\int_D^\infty x^2 \exp\left(-\frac{x^2}{2}\right) dx = D\exp\left(-\frac{D^2}{2}\right) + \sqrt{\frac{\pi}{2}}\operatorname{erfc}\left(\frac{D}{\sqrt{2}}\right) \tag{4.21}$$

and the estimate

$$\int_0^D x^q \exp\left(-\frac{x^2}{2}\right) dx = g(D) \lesssim \begin{cases} D^{q+1} & \text{when } 0 \le D \le 1 \\ 1 & D > 1. \end{cases} \tag{4.22}$$

Next we define D_ℓ to satisfy

$$a_\ell D_\ell^q = \delta D_\ell^2, \quad \text{i.e.} \quad D_\ell = \left(\frac{\delta}{a_\ell}\right)^{\frac{1}{q-2}}.$$

The expectation in (4.20) satisfies

$$\begin{aligned}
\mathbb{E}\min\{a_\ell|N_\ell|^q, \delta N_\ell^2\} &= \int_{-\infty}^{\infty} \min\{a_\ell|x|^q, \delta x^2\}\exp\left(-\frac{x^2}{2}\right) dx \\
&= 2a_\ell \int_0^{D_\ell} |x|^q \exp\left(-\frac{x^2}{2}\right) dx + 2\delta \int_{D_\ell}^{\infty} x^2 \exp\left(-\frac{x^2}{2}\right) dx \\
&\lesssim a_\ell g(D_\ell) + \delta(D_\ell+1)f(D_\ell),
\end{aligned}$$

where

$$f(D_\ell) \simeq \begin{cases} 1, & \text{when } 0 \le D_\ell \le 1 \text{ and} \\ \exp(-\frac{D_\ell^2}{2}) & \text{otherwise.} \end{cases}$$

For small values of D_ℓ, i.e. $\delta \le a_\ell$ we have

$$\mathbb{E}\min\{a_\ell|N_\ell|^q, \delta N_\ell^2\} \lesssim \frac{\delta^{q+1}}{a_\ell^q} + \left(\frac{\delta}{a_\ell}\right)^{\frac{1}{q-2}} + \delta$$

35

and for $\delta > a_\ell$ it holds that

$$\mathbb{E}\min\{a_\ell|N_\ell|^q, \delta N_\ell^2\} \lesssim a_\ell + \delta\left(\left(\frac{\delta}{a_\ell}\right)^{\frac{1}{q-2}} + 1\right)\exp\left(\left(\frac{\delta}{a_\ell}\right)^{\frac{1}{q-2}}\right).$$

Since L in (4.20) is arbitrary, we have

$$
\begin{aligned}
\mathbb{E}e_{\delta,\zeta}(K^*N) &\lesssim \sum_{\ell=1}^{\infty}\mathbb{E}\min\{a_\ell|N_\ell|^q, \delta N_\ell^2\}\\
&= \sum_{\ell=1}^{\widetilde{L}}\left(\frac{\delta^{q+1}}{a_\ell^q} + \left(\frac{\delta}{a_\ell}\right)^{\frac{1}{q-2}} + \delta\right) + \sum_{\ell=\widetilde{L}}^{\infty}\left(a_\ell + \delta\left(\left(\frac{\delta}{a_\ell}\right)^{\frac{1}{q-2}} + 1\right)\exp\left(\left(\frac{\delta}{a_\ell}\right)^{\frac{1}{q-2}}\right)\right)\\
&\lesssim \delta\widetilde{L} + \sum_{\ell=\widetilde{L}}^{\infty}a_\ell + \delta\sum_{\ell=\widetilde{L}}^{\infty}\left(\left(\frac{\delta}{a_\ell}\right)^{\frac{1}{q-2}} + 1\right)\exp\left(\left(\frac{\delta}{a_\ell}\right)^{\frac{1}{q-2}}\right), \quad (4.23)
\end{aligned}
$$

where we have chosen $\widetilde{L}$ so that $a_{\widetilde{L}+1} < \delta \le a_{\widetilde{L}}$.

Recall now that due to our assumptions we have

$$a_\ell \simeq \zeta^{1-q}\ell^{q(-t-s+\frac{1}{2})-1} = \zeta^{1-q}\ell^{-r},$$

where we write $r = q(t+s-\frac{1}{2})+1 > \frac{q}{2}$. Since $\delta \simeq a_{\widetilde{L}}$, our choice for $\widetilde{L}$ indicates that

$$\widetilde{L} \simeq \delta^{-\frac{1}{r}}\zeta^{-\frac{q-1}{r}}.$$

Now we are able to estimate all terms in (4.23). First, we have

$$\sum_{\ell=\widetilde{L}}^{\infty}a_\ell \simeq \zeta^{1-q}\int_{\widetilde{L}}^{\infty}\ell^{-r}d\ell = \zeta^{1-q}\frac{(\delta^{-\frac{1}{r}}\zeta^{\frac{1-q}{r}})^{1-r}}{r-1} \simeq \delta\widetilde{L}.$$

Second, by denoting $\theta = \left(\frac{\delta}{\zeta^{1-q}}\right)^{\frac{1}{q-2}}$ we obtain

$$
\begin{aligned}
\sum_{\ell=\widetilde{L}}^{\infty}\left(\frac{\delta}{a_\ell}\right)^{\frac{1}{q-2}}f\left(\left(\frac{\delta}{a_\ell}\right)^{\frac{1}{q-2}}\right) &\simeq \theta\int_{\widetilde{L}}^{\infty}x^{\frac{r}{q-2}}\exp\left(-\frac{\theta^2}{2}x^{\frac{2r}{q-2}}\right)dx\\
&= \theta^{\frac{2-q}{r}}\int_{1}^{\infty}y^{\frac{r}{q-2}}\exp\left(\frac{-y^{\frac{2r}{q-2}}}{2}\right)dy \simeq \delta^{-1/r}\zeta^{\frac{1-q}{r}} \simeq \widetilde{L},
\end{aligned}
$$

where we applied a change of variables $y = \theta^{\frac{q-2}{r}}x$. Third, we notice similarly to the second case that

$$\sum_{\ell=\widetilde{L}}^{\infty}f\left(\left(\frac{\delta}{a_\ell}\right)^{\frac{1}{q-2}}\right) \simeq \int_{\widetilde{L}}^{\infty}\exp\left(-\frac{\theta^2}{2}x^{\frac{2r}{q-2}}\right)dx \simeq \theta^{\frac{2-q}{r}}\int_{1}^{\infty}\exp\left(-\frac{y^{\frac{2r}{q-2}}}{2}\right)dy \simeq \widetilde{L}.$$

36

Finally, by applying the three estimates above to (4.23) we conclude that

$$\mathbb{E}e_{\delta,\zeta}(K^*N) \lesssim \delta\widetilde{L} \simeq \delta^{1-1/r}\zeta^{\frac{1-q}{r}},$$

where $r = q(t + s - 1/2) + 1$, which yields the claim. $\qquad\square$

Theorem 4.9. *Let us assume that K and R are given as above, $s \geq \frac{1}{p} - \frac{1}{2}$ and $t > \frac{1}{2}$. For the choice $\alpha \simeq \delta^\kappa$, where*

$$\kappa = \frac{4(s + t)}{4(s + t) + 1}$$

we obtain the convergence rate

$$\mathbb{E}D_R^{\mu_\alpha^\delta,\mu^\dagger}(U_\alpha^\delta, u^\dagger) \lesssim \delta^\kappa.$$

Proof. By combing Lemma 4.8 with inequality (4.19) we have

$$\mathbb{E}D_R^{\mu_\alpha^\delta,\mu^\dagger}(u_\alpha^\delta, u^\dagger) \quad \lesssim \quad \alpha + \inf_{\zeta>0}\left(\zeta^{\frac{2}{2-p}}\left(\frac{\delta}{\alpha}\right)^{\frac{2}{2-p}}\left(1 + \frac{\delta}{\alpha}(\frac{\delta}{2})^{1-\frac{1}{r}}\left(\frac{\alpha\gamma}{\delta}\right)^{\frac{1-q}{r}}\right) + \frac{\delta}{\alpha}\delta^{1-1/r}\zeta^{\frac{1-q}{r}}\right)$$

$$=: \quad \widetilde{M}_1 + \widetilde{M}_2. \qquad\qquad (4.24)$$

By setting $\alpha \simeq \delta^\kappa$ we can write

$$\widetilde{M}_2 \lesssim \inf_{\zeta>0}\left(\zeta^\nu\delta^{\nu(1-\kappa)}\left(1 + \delta^{2+\frac{q-2}{r}-\kappa(1+\frac{q-1}{r})}\right) + \delta^{2-\kappa-\frac{1}{r}}\zeta^{\frac{1-q}{r}}\right)$$

$$= \left(1 + \delta^{2+\frac{q-2}{r}-\kappa(1+\frac{q-1}{r})}\right)^{\frac{q-1}{\nu r+q-1}}\delta^{\frac{\nu(1-\kappa)(q-1)+\nu r(2-\kappa-\frac{1}{r})}{\nu r+q-1}},$$

where we have denoted $\nu = \frac{2}{2-p}$. By our assumption $r > 1$, that is, $2+\frac{q-2}{r}-\kappa(1+\frac{q-1}{r}) > 0$ and we obtain

$$M_2 \lesssim \delta^{\frac{\nu(1-\kappa)(q-1)+\nu r(2-\kappa-\frac{1}{r})}{\nu r+q-1}}.$$

By optimizing the convergence rate in (4.24) we have

$$\mathbb{E}D_R^{\mu_\alpha^\delta,\mu^\dagger}(u_\alpha^\delta, u^\dagger) \lesssim \delta^{\frac{\nu(q+2(r-1))}{\nu(q+2r-1)+q-1}}.$$

Since $r = q(t + s - 1/2) + 1$ we can write

$$\frac{\nu(q+2(r-1))}{\nu(q+2r-1)+q-1} = \frac{\nu(2q(t+s))}{\nu(2q(t+s)+1)+q-1}.$$

Furthermore since $\nu = \frac{2}{2-p}$ and $\frac{1}{p} + \frac{1}{q} = 1$ we see that

$$\frac{4q}{2-p} = \frac{4p}{(2-p)(p-1)} \quad \text{and} \quad \frac{2}{2-p} + q - 1 = \frac{p}{(2-p)(p-1)},$$

which yields the claim

$$\mathbb{E}D_R^{\mu_\alpha^\delta,\mu^\dagger}(u_\alpha^\delta, u^\dagger) \lesssim \delta^{\frac{4(s+t)}{4(s+t)+1}}.$$

$\qquad\square$

Remark 4.10. If we assume $p = 2$ then we can use the inequalities of Theorem 3.10 instead of Theorem 3.7 to attain the same result. Note that if $p = 2$ and the exact source condition is assumed then $\mu^\dagger = (I - \Delta)^s u^\dagger = K^* w$, where $w \in L^2(\mathbb{T})$. This means that $u^\dagger \in H^r(\mathbb{T})$, with $r = 2s + t$, and we can write

$$\mathbb{E} D_R^{\mu_\alpha^\delta, \mu^\dagger}(u_\alpha^\delta, u^\dagger) = \mathbb{E}\|u_\alpha^\delta - u^\dagger\|_{H^s(\mathbb{T})}^2 \leq \delta^{\frac{4(r-s)}{2r+2t+1}},$$

which is the minimax rate [12].

Remark 4.11. We point out that minimax rates for linear statistical inverse problems in wavelet basis have been studied for estimators based on Galerkin methods and non-linear thresholding algorithms (see [21, 16, 40] and references therein). In the first two papers the authors construct a finite-dimensional estimator u_δ for any $\delta > 0$ such that

$$\sup_{u \in B} \mathbb{E}\|u - u_\delta\|_{L^2}^2 \lesssim \left(\delta\sqrt{|\log \delta|}\right)^{\frac{4s}{2s+2t+1}}, \tag{4.25}$$

the forward operator K is t times smoothing (similar to (4.17)) and

$$B = \{u \mid \|u\|_{B_p^s} \leq C\}. \tag{4.26}$$

Such rates are also known to be optimal [16]. Compared to (4.26) our method builds upon a more general source condition. We do not necessarily require that the true solution is in the range of K^* (if the range is defined as K^*Y and not K^* on a larger space including the noise). However, there is interplay between smoothness of K and our source condition. In addition, the rate in (4.25) is achieved in a L^2-norm, whereas the symmetric Bregman distance of B_1^s-norm in Theorem 4.7 is not a norm, since it is not strictly positive and does not satisfy a triangle inequality. On the other hand the Bregman distance estimate can be used to provide structural properties related to sparsity, e.g. a bound on the norm of the wavelet coefficients of the reconstruction outside the support of the coefficients of $u^\dagger$ (cf. [10]). We also mention that for the related approach wavelet soft-thresholding, where first a reconstruction $K^{-1}f^\delta$ is computed in a very large Besov space and then projected back by soft thresholding of the wavelet coefficients (cf. [21]), our approach can be used to provide analogous rates as in [16] by applying the estimates to the variational regularization

$$J_\alpha^\delta(u) = \frac{1}{2}\|u\|_{L^2} - \langle (K^*)^{-1}u, f^\delta \rangle + \alpha R(u),$$

where $R(u)$ is the associated Besov norm (weighted ℓ^1-norm on wavelet coefficients). This is indeed a special case of our approach with definition $\|f\|_Y = \|K^{-1}f\|_{L^2}$, that is, Y is the space of elements where the latter norm is finite. Note that this corresponds naturally to a large noise case that cannot be treated with the existing theory. We also mention that some extensions to the case of K not being injective are possible. From our analysis and the form of the functional it becomes apparent that a necessary condition is that the true solution is in the range of K^* and the associated subgradient is in L^2,

38

which is indeed a rather weak condition. Our approach then yields L^2-estimates as in [16] (corresponding to an estimate for $\|K(u-u^\dagger)\|_Y^2$, but in addition we also obtain an estimate in the Bregman distance providing information about the sparsity. It remains an interesting future question to provide more comparison between the Galerkin approach and variational methods.

4.4 Total Variation-type Regularization

In the following we discuss the case of total variation regularization

$$R(u) = \sup_{\varphi \in C_0^\infty(\mathbb{T}^d), \|\varphi\|_\infty \leq 1} \int_{\mathbb{T}^d} \nabla \cdot \varphi u \; dx, \tag{4.27}$$

or related regularizations such as infimal convolutions with higher order total variation (cf. [9] and references therein) in the case of spatial dimension $d \leq 2$, when there is an embedding into $X_0 = L^2(\mathbb{T}^d)$. Thus, it is natural to use Hilbert space embedding in this case. We assume that K can be extended to a $t > d/2 + \epsilon$ times smoothing bijective bounded linear operator in Sobolev scale. We also assume that N is white noise taking values in $H^{-d/2-\epsilon}$ as in the previous sections. We will use the estimate (3.15) for realizations of K^*N (noting $L = K^*K$) and write

$$e_{\delta,\zeta}(K^*N) \leq C\|L^{-\nu}K^*N\|_{L^2(\mathbb{T}^d)}^{\frac{1}{\nu}}\zeta^{-\frac{1-2\nu}{\nu}}\delta$$

to obtain an estimate for the expectation $\mathbb{E}(e_{\delta,\zeta}(K^*N))$. Subsequently one could use similar reasoning as in the previous section respectively Section 3.1 to obtain full rates, which we leave to the reader.

The key question for the finite expectation of $e_{\delta,\zeta}(K^*N)$ is the choice of ν such that

$$\mathbb{E}\|L^{-\nu}K^*N\|_{L^2(\mathbb{T}^d)}^{\frac{1}{\nu}} < \infty.$$

Note that by Fernique's theorem any moment of white noise is finite in $H^{-d/2-\epsilon}(\mathbb{T}^d)$ for any $\epsilon > 0$ [17]. Thus, we do not need to worry about the exponent $\frac{1}{\nu}$ in the expectation, but rather optimize ν to have

$$\|L^{-\nu}K^*N\|_{L^2(\mathbb{T}^d)} \leq C\|N\|_{H^{-d/2-\epsilon}(\mathbb{T}^d)}.$$

With the above smoothing assumptions, we see that K^* maps from $H^{-d/2-\epsilon}(\mathbb{T}^d)$ to $H^{t-d/2-\epsilon}(\mathbb{T}^d)$, hence we achieve the above rate estimate if $L^{-\nu}$ is bounded from $H^{t-d/2-\epsilon}(\mathbb{T}^d)$ to $L^2(\mathbb{T}^d)$. For K being the inverse of a translation invariant differential or pseudo-differential operator one obtains that $L^\nu : L^2(\mathbb{T}^d) \to H^{2t\nu}(\mathbb{T}^d)$. In the following we write $K \in \Psi^\rho$ for a pseudodifferential operator K if its symbol is in $\mathcal{S}^\rho(\mathbb{T}^d; \mathbb{T}^d)$ [63]. The condition above means that we should choose $\nu = \frac{t-d/2-\epsilon}{2t}$.

As a specific example consider the pseudodifferential operator $K = (-\Delta + I)^{-1}$. Then K is a twice smoothing bijective operator between $H^{-d/2-\epsilon}(\mathbb{T}^d)$ and $H^{2-d/2-\epsilon}(\mathbb{T}^d)$, which gives us $\nu = \frac{2-d/2-\epsilon}{4}$, i.e., one can choose ν arbitrarily close to $\frac{4-d}{8}$.

Theorem 4.12. *Let us assume that $K \in \Psi^{-t}$ where $t > d/2 + \epsilon$ with some $\epsilon > 0$, that is, K is of order t smoothing pseudodifferential operator. Regularization functional R is defined by (4.27) and $\mu^\dagger$ satisfies the approximate source condition of order $r_1 \geq 0$ in Assumption 3.2. Then for the choice $\alpha \simeq \delta^\kappa$ where*

$$\kappa = \begin{cases} \frac{1+r_1}{(2+r_1)(1-\nu)} & \text{for } r_1 \leq \frac{d+2\epsilon}{t} \text{ and} \\ 1 & \text{else} \end{cases}$$

we obtain the convergence rate

$$\mathbb{E}D_R^{\mu_\alpha^\delta,\mu^\dagger}(U_\alpha^\delta, u^\dagger) \lesssim \begin{cases} \delta^{\frac{1}{(2+r_1)(1-\nu)}} \leq \delta^{\frac{1}{2+r_1}} & \text{for } r_1 \leq \frac{d+2\epsilon}{t} \text{ and} \\ \delta^{\frac{1}{1+r_1}} & \text{else} \end{cases}$$

where $\nu = \frac{t-d/2-\epsilon}{2t}$.

Proof. Recall that

$$\mathbb{E}D_R^{\mu_\alpha^\delta,\mu^\dagger}(U_\alpha^\delta, u^\dagger) \leq \widetilde{M}_1 + \widetilde{M}_2. \tag{4.28}$$

where terms $\widetilde{M}_1$ and $\widetilde{M}_2$ are given in equation (4.18). We have for a constant γ and $\alpha \simeq \delta^\kappa$, $\kappa \leq 1$ that

$$\mathbb{E}e_{\delta,\frac{\alpha\gamma}{\delta}}(K^*N) \lesssim \gamma^{2-\frac{1}{\nu}}\delta^{\kappa+(\frac{1}{\nu}-1)(1-\kappa)} \lesssim 1.$$

since $\frac{1}{\nu} > 2$. Therefore, we obtain

$$\widetilde{M}_1 \lesssim \inf_{\zeta_1 \in \mathbb{R}_+} \left\{\zeta_1 + \delta^\kappa \zeta_1^{-r_1}\right\} \simeq \delta^{\frac{\kappa}{1+r_1}}$$

and

$$\widetilde{M}_2 \lesssim \inf_{\zeta_2 \in \mathbb{R}_+} \left\{\delta^{1-\kappa}\left(\zeta_2 + \delta\zeta_2^{-\frac{1-2\nu}{\nu}}\right)\right\} \simeq \delta^{\frac{1-\kappa(1-\nu)}{1-\nu}}.$$

If $r_1 \leq \frac{d+2\epsilon}{t}$ the convergence rate is minimized with κ which satisfies

$$\kappa = \frac{1+r_1}{(2+r_1)(1-\nu)}.$$

For $r_1 \geq \frac{d+2\epsilon}{t}$ we choose $\kappa = 1$ and consequently the claim holds. $\square$

5 Outlook to the Bayesian approach

In the Bayesian approach to inverse problems the model equation (1.2) is often written in the form

$$F_\delta = KU + \delta N \tag{5.1}$$

where, in addition to the observational random noise N, we describe our prior beliefs about the unknown in terms of the probability distribution of the random variable $U : \Omega \to X$. The solution to the inverse problem is then the probability distribution of

U conditioned on a measurement outcome F_δ. The posterior distribution now provides means for uncertainty quantification.

The analysis of small noise limit, in Bayesian case also known as the theory of posterior consistency, has attracted a lot of interest in the last decade. Posterior convergence rates were first studied in [29, 62]. In those two papers Gaussian noise and prior are assumed and the interest is on the convergence of the approximated solution U_δ^α, generated by a 'true' $u^\dagger$, to the same truth $u^\dagger$. Similar convergence or the contraction of the whole posterior distribution is further studied e.g. in papers [2, 18, 41, 54, 57, 65]. In [42, 43] Bayesian cost estimator similar to (5.2) in Gaussian case is considered.

A widely used approach to extract information from a posteriori distribution is to find so-called maximum a posteriori (MAP) estimator. In finite dimensional problems, the MAP estimate maximizes a posteriori probability density function and is, loosely speaking, the most probable solution to the problem (5.1). In the infinite-dimensional case, the MAP estimator is less understood. In certain probabilistic models, the MAP estimate is known to minimize a problem of type (1.3). We refer to our earlier work in [33, 34] and other authors in [18, 22] for more discussion on the topic. We point out that, in general, the connection between the estimator induced by (1.3) and the MAP estimate is not well-established. Despite this deficit, understanding the Bayes cost in such a case based on Bregman distance would be highly interesting for practical problems.

Our results in Theorem 2.9 now directly yields that

$$
\mathbb{E}_{N,U}(D_R^{\mu_\alpha^\delta,\mu}(U_\alpha^\delta,U)) \leq \mathbb{E}_N\Big(\mathbb{E}_U\Big(\inf_{(\zeta_1,\zeta_2)\in(\mathbb{R}^+)^2}(\zeta_1 + \frac{\delta}{\alpha}\zeta_2)^{1/(1-\theta)}C_\theta(U_\alpha^\delta,U)^{1/(1-\theta)} +
$$
$$
\frac{1}{1-\theta}e_{\alpha,\zeta_1}(M) + \frac{\delta}{\alpha(1-\theta)}e_{\delta,\zeta_2}(K^*N)\Big)\Big), \qquad (5.2)
$$

where $M : \Omega \to X^*$ formally satisfies $M(\omega) \in \partial R(U(\omega))$. The Bayes cost for the MAP estimate, however, is not a straightforward matter since the subgradient set $\partial R(U)$ is not necessarily well-defined. Consider a Gaussian prior U in a Hilbert space X with zero-mean and covariance $C_U : X \to X$. In such a case, the functional R induced by the prior satisfies $R(u) = \left\|C_U^{-1/2}u\right\|_X^2$, i.e., R coincides with the norm of the Cameron–Martin space. On the other hand, realizations of U are in the Cameron–Martin space with probability zero. Similarly, expectation over R and Bregman distance in (5.2) are not bounded.

It is know from the earlier work [43] by the last author that in Gaussian setting the Bregman distance based Bayes cost can be estimated using a weaker norm than the one induced by the prior. Hence an intriguing question for future work is to characterize functional R for which the Bayes cost (and the bound) in (5.2) makes sense.

Let us finally comment that in a purely Bayesian approach the prior information should be independent of the measurement F_δ. For instance, MAP estimate of problem (1.2) for a δ-independent prior and a noise distribution δN with white noise N formally correspond to an estimator (1.3) where α is replaced by $\alpha\delta^2$ for a constant α. In literature this principle is occasionally omitted and general a priori rules $\alpha = \alpha(\delta)$ are considered. Such an approach resembling the frequentist method leads to 'priors' that are scaled

with respect to the noise level δ and hence no longer independent of the measurement. With general $\alpha(\delta)$ the minimisation problem (1.3) can not be seen as a proper MAP estimate. However, it is a useful estimator to study since with constant α we often do not get convergence in the original space.

Acknowledgements

A   Convex Conjugates

For completeness we recall the convex conjugate $R^\star : X^* \to \mathbb{R} \cup \{\infty\}$ defined via

$$R^\star(q) = \sup_{u \in X} \left(\langle q, u \rangle_{X^* \times X} - R(u) \right). \tag{A.1}$$

Note that by definition of $R^\star(q)$ we obtain the following well-known generalization of Young's inequality

$$\langle q, u \rangle_{X^* \times X} \leq R(u) + R^\star(q), \tag{A.2}$$

for all $u \in X$ and $q \in X^*$, which we employ at several instances throughout the paper.

References

[1] F. u. Abramovich and B. Silverman. Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, 85(1):115–129, 1998.

[2] S. Agapiou, S. Larsson, and A. M. Stuart. Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems. *Stochastic Processes and their Applications*, 123(10):3828–3860, 2013.

[3] M. Benning and M. Burger. Error estimates for general fidelities. *Electronic Transactions on Numerical Analysis*, 38(44-68):77, 2011.

[4] T. Bonesky, K. S. Kazimierski, P. Maass, F. Schöpfer, and T. Schuster. Minimization of Tikhonov functionals in Banach spaces. In *Abstract and Applied Analysis*, volume 2008. Hindawi Publishing Corporation, 2008.

[5] L. M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3):200–217, 1967.

[6] A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM review*, 51(1):34–81, 2009.

[7] M. Burger. Bregman distances in inverse problems and partial differential equation. *arxiv:1505.05191*, 2015.

[8] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse problems*, 20(5):1411, 2004.

[9] M. Burger and S. Osher. A guide to the TV zoo. In *Level Set and PDE Based Reconstruction Methods in Imaging*, pages 1–70. Springer International Publishing, 2013.

[10] M. Burger, E. Resmerita, and L. He. Error estimation for Bregman iterations and inverse scale space methods in image restoration. *Computing*, 81(2-3):109–135, 2007.

[11] V. Caselles, A. Chambolle, and M. Novaga. Total variation in imaging. *Handbook of Mathematical Methods in Imaging*, pages 1455–1499, 2015.

[12] L. Cavalier. Nonparametric statistical inverse problems. *Inverse Problems*, 24(3):034004, 2008.

[13] L. Cavalier and A. Tsybakov. Sharp adaptation for inverse problems with random noise. *Probability Theory and Related Fields*, 123(3):323–354, 2002.

[14] A. Chambolle, V. Duval, G. Peyré, and C. Poon. Geometric properties of solutions to the total variation denoising problem. *Inverse Problems*, 33(1):015002, 2016.

[15] A. Cohen, M. Hoffmann, and M. Reiss. Adaptive wavelet Galerkin methods for linear inverse problems. *SIAM Journal on Numerical Analysis*, 42(4):1479–1501, 2004.

[16] A. Cohen, M. Hoffmann, and M. Reiss. Adaptive wavelet galerkin methods for linear inverse problems. *SIAM Journal on Numerical Analysis*, 42(4):1479–1501, 2004.

[17] G. Da Prato and J. Zabczyk. *Stochastic equations in infinite dimensions*, volume 152. Cambridge university press, 2014.

[18] M. Dashti, K. J. H. Law, A. M. Stuart, and J. Voss. MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Problems*, 29(9):095017, 27, 2013.

[19] I. Daubechies. *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[20] I. Daubechies, G. Teschke, and L. Vese. Iteratively solving linear inverse problems under general convex constraints. *Inverse Problems and Imaging*, 1(1):29, 2007.

[21] D. L. Donoho. Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition. *Applied and computational harmonic analysis*, 2(2):101–126, 1995.

[22] M. M. Dunlop and A. M. Stuart. The Bayesian formulation of EIT: Analysis and algorithms. *arXiv preprint arXiv:1508.04106*, 2015.

[23] H. Egger. Regularization of inverse problems with large noise. In *Journal of Physics: Conference Series*, volume 124, page 012022. IOP Publishing, 2008.

[24] P. Eggermont, V. LaRiccia, and M. Nashed. On weakly bounded noise in ill-posed problems. *Inverse Problems*, 25(11):115018, 2009.

[25] I. Ekeland and R. Temam. *Convex analysis and variational problems*. SIAM, 1976.

[26] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

[27] J. Flemming and M. Hegland. Convergence rates in 1-regularization when the basis is not smooth enough. *Applicable Analysis*, 94(3):464–476, 2015.

[28] J. Flemming, B. Hofmann, and I. Veselić. On ℓ^1-regularization in light of nashed's ill-posedness concept. *Computational Methods in Applied Mathematics*, 15(3):279–289, 2015.

[29] S. Ghosal, J. K. Ghosh, and A. W. Van Der Vaart. Convergence rates of posterior distributions. *Annals of Statistics*, 28(2):500–531, 2000.

[30] A. Goldenshluger and S. V. Pereverzev. Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Probability Theory and Related Fields*, 118(2):169–186, 2000.

[31] M. Grasmair. Linear convergence rates for Tikhonov regularization with positively homogeneous functionals. *Inverse Problems*, 27(7):075014, 2011.

[32] T. Hein. Convergence rates for regularization of ill-posed problems in Banach spaces by approximate source conditions. *Inverse Problems*, 24(4):045007, 2008.

[33] T. Helin and M. Burger. Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems. *Inverse Problems*, 31(8):085009, 2015.

[34] T. Helin and M. Lassas. Hierarchical models in statistical inverse problems and the Mumford-Shah functional. *Inverse Problems*, 27(1):015008, 32, 2011.

[35] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms I: fundamentals*, volume 305. Springer Science & Business Media, 2013.

[36] B. Hofmann. Approximate source conditions in Tikhonov–Phillips regularization and consequences for inverse problems with multiplication operators. *Mathematical Methods in the Applied Sciences*, 29(3):351–371, 2006.

[37] B. Hofmann, B. Kaltenbacher, C. Poeschl, and O. Scherzer. A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Problems*, 23(3):987, 2007.

[38] B. Hofmann and M. Yamamoto. Convergence rates for Tikhonov regularization based on range inclusions. *Inverse Problems*, 21(3):805, 2005.

[39] T. Hohage and F. Weidling. Characterizations of variational source conditions, converse results, and maxisets of spectral regularization methods. *arXiv preprint arXiv:1603.05133*, 2016.

[40] T. Hohage and F. Werner. Inverse problems with poisson data: statistical regularization theory, applications and algorithms. *Inverse Problems*, 32(9):093001, 2016.

[41] T.-M. Huang. Convergence rates for posterior distributions and adaptive estimation. *The Annals of Statistics*, 32(4):1556–1593, 2004.

[42] H. Kekkonen, M. Lassas, and S. Siltanen. Analysis of regularized inversion of data corrupted by white Gaussian noise. *Inverse Problems*, 30(4):045009, 2014.

[43] H. Kekkonen, M. Lassas, and S. Siltanen. Posterior consistency and convergence rates for Bayesian inversion with hypoelliptic operators. *arXiv preprint arXiv:1507.01772*, 2015.

[44] G. Kerkyacharian, D. Picard, L. Birgé, P. Hall, O. Lepski, E. Mammen, A. Tsybakov, G. Kerkyacharian, and D. Picard. Thresholding algorithms, maxisets and well-concentrated bases. *Test*, 9(2):283–344, 2000.

[45] S. Kindermann. Convex tikhonov regularization in banach spaces: New results on convergence rates. *Journal of Inverse and Ill-posed Problems*, 24(3):341–350, 2016.

[46] B. Knapik, A. van Der Vaart, J. Van Zanten, et al. Bayesian inverse problems with gaussian priors. *The Annals of Statistics*, 39(5):2626–2657, 2011.

[47] D. A. Lorenz. Convergence rates and source conditions for Tikhonov regularization with sparsity constraints. *Journal of Inverse and Ill-Posed Problems*, 16(5):463–478, 2008.

[48] S. Lu and P. Mathé. Discrepancy based model selection in statistical inverse problems. *Journal of Complexity*, 30(3):290–308, 2014.

[49] B. A. Mair and F. H. Ruymgaart. Statistical inverse estimation in Hilbert scales. *SIAM Journal on Applied Mathematics*, 56(5):1424–1444, 1996.

[50] P. Mathé and B. Hofmann. Direct and inverse results in variable Hilbert scales. *Journal of Approximation Theory*, 154(2):77–89, 2008.

[51] P. Mathé and S. V. Pereverzev. Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19(3):789, 2003.

[52] P. Mathé and U. Tautenhahn. Interpolation in variable Hilbert scales with application to inverse problems. *Inverse Problems*, 22(6):2271, 2006.

[53] P. Mathé and U. Tautenhahn. Regularization under general noise assumptions. *Inverse Problems*, 27(3):035016, 2011.

[54] F. Monard, R. Nickl, and G. P. Paternain. Efficient nonparametric Bayesian inference for X-ray transforms. *arXiv preprint arXiv:1708.06332*, 2017.

[55] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005.

[56] R. Ramlau and G. Teschke. A tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints. *Numerische Mathematik*, 104(2):177–203, 2006.

[57] K. Ray. Bayesian inverse problems with non-conjugate priors. *Electronic Journal of Statistics*, 7:2516–2549, 2013.

[58] E. Resmerita. Regularization of ill-posed problems in Banach spaces: convergence rates. *Inverse Problems*, 21(4):1303, 2005.

[59] E. Resmerita and O. Scherzer. Error estimates for non-quadratic regularization and the relation to enhancement. *Inverse Problems*, 22(3):801, 2006.

[60] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[61] T. Schuster, B. Kaltenbacher, B. Hofmann, and K. S. Kazimierski. *Regularization methods in Banach spaces*, volume 10. Walter de Gruyter, 2012.

[62] X. Shen and L. Wasserman. Rates of convergence of posterior distributions. *Annals of Statistics*, pages 687–714, 2001.

[63] M. E. Taylor. *Pseudodifferential operators*. Springer, 1996.

[64] M. C. Veraar. Regularity of Gaussian white noise on the d-dimensional torus. In *Marcinkiewicz centenary volume*, volume 95 of *Banach Center Publ.*, pages 385–398. Polish Acad. Sci. Inst. Math., Warsaw, 2011.

[65] S. J. Vollmer. Posterior consistency for Bayesian inverse problems through stability and regression results. *Inverse Problems*, 29(12):125011, 2013.

[66] Z.-B. Xu and G. F. Roach. Characteristic inequalities of uniformly convex and uniformly smooth banach spaces. *Journal of Mathematical Analysis and Applications*, 157(1):189–210, 1991.

[67] T. Zhang. Learning bounds for kernel regression using effective data dimensionality. *Neural Computation*, 17(9):2077–2098, 2005.