SIAM J. NUMER. ANAL.
Vol. 58, No. 1, pp. 21–47

OPTIMAL CONVERGENCE RATES FOR TIKHONOV REGULARIZATION IN BESOV SPACES*

FREDERIC WEIDLING†, BENJAMIN SPRUNG†, AND THORSTEN HOHAGE†

Abstract. This paper deals with Tikhonov regularization for linear and nonlinear ill-posed operator equations with wavelet Besov norm penalties. We show order optimal rates of convergence for finitely smoothing operators and for the backwards heat equation for a range of Besov spaces using variational source conditions. We also derive order optimal rates for a white noise model with the help of variational source conditions and concentration inequalities for sharp negative Besov norms of the noise.

Key words. inverse problems, white noise, regularization, variational source conditions

AMS subject classifications. 65J20, 65J22, 65N21

DOI. 10.1137/18M1178098

1. Introduction.

We consider ill-posed operator equations

$$F(f^\dagger) = g^\dagger$$

with a noisy right-hand side and a forward operator $F\colon \mathcal{D} \subset \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X}$ is some Besov space $B^s_{p,q}$; $\mathcal{D} \subset \mathcal{X}$ a nonempty, closed, convex set; and $\mathcal{Y}$ an L^2 space; $f^\dagger \in \mathcal{D}$ denotes the true solution. We treat two noise models: the standard deterministic error model where the observed data g^{obs} satisfy

$$g^{\mathrm{obs}} = g^\dagger + \xi, \qquad \|\xi \,|\, \mathcal{Y}\| \le \delta, \tag{1.1}$$

with a deterministic noise level $\delta > 0$ and statistical models

$$g^{\mathrm{obs}} = g^\dagger + \varepsilon Z \tag{1.2}$$

with a statistical noise level $\varepsilon > 0$ and some noise process Z on $\mathcal{Y}$ with white noise as the most prominent example. (As we will have to deal with norms involving many, sometimes nested indices, we use the notation $\|\cdot \,|\, \mathcal{Y}\|$ instead of $\|\cdot\|_{\mathcal{Y}}$ throughout the paper.)

One of the most common approaches to compute a stable approximation of $f^\dagger$ given g^{obs} is Tikhonov regularization of the form

$$\widehat{f}_\alpha \in \operatorname*{arg\,min}_{f \in \mathcal{D}} \left[\tfrac{1}{t} \left\| F(f) - g^{\mathrm{obs}} \,\middle|\, \mathcal{Y} \right\|^t + \alpha \mathcal{R}(f) \right] \tag{1.3}$$

with some $t \ge 1$ and a regularization parameter $\alpha > 0$. In this paper we study the case that $\mathcal{R}(f)$ is given by a norm power of a Besov norm of f. Such penalties

*Received by the editors April 3, 2018; accepted for publication (in revised form) October 7, 2019; published electronically January 7, 2020.

https://doi.org/10.1137/18M1178098

Funding: The work of the authors was supported by Deutsche Forschungsgemeinschaft (DFG) through project B01 of RTG 2088. The work of the second author was supported by DFG through project C09 of CRC 755.

†Institut für Numerische und Angewandte Mathematik, Universität Göttingen, Lotzestraße 16-18, 37083 Göttingen, Germany (f.weidling@math.uni-goettingen.de, b.sprung@math.uni-goettingen.de, hohage@math.uni-goettingen.de).

with wavelet Besov norms with small index p are frequently used to enforce sparsity (see, e.g., [11, 33]). As white noise on a Hilbert space $\mathcal{Y}$ does not belong to $\mathcal{Y}$ with probability 1, we use $t = 2$ in this case, expand the square, and omit the term $\frac{1}{2}\|g^{\mathrm{obs}} \,|\, \mathcal{Y}\|^2$, which has no influence on the minimizer. This yields

$$\widehat{f}_\alpha \in \operatorname*{arg\,min}_{f\in\mathcal{D}} \left[\tfrac{1}{2}\|F(f) \,|\, \mathcal{Y}\|^2 - \left\langle g^{\mathrm{obs}}, F(f)\right\rangle + \alpha\mathcal{R}(f)\right]. \tag{1.4}$$

A main goal of regularization theory are bounds on the distance of regularized estimators of the true solution in terms of the noise level δ or ε, respectively. In the case that $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces, such error bounds can be obtained by spectral theory (see, e.g., [15] for the deterministic case and [3] for the stochastic case). Concerning convergence results for deterministic regularization with Besov norms, we refer the reader to [11] for convergence without rates, to [32] for situations in which the rate $O(\sqrt{\delta})$ based on the source condition $\partial\mathcal{R}(f^\dagger)\cap\operatorname{ran}(T^*) \neq \emptyset$ can be achieved (possibly with respect to negative Sobolev norms), and to [33] for situations in which the rate $O(\delta^{2/3})$ under the source condition $\partial\mathcal{R}(f^\dagger)\cap\operatorname{ran}(T^*T) \neq \emptyset$ occurs. For statistical inverse problems, minimax optimal rates under Besov smoothness assumptions have been shown for methods based on wavelet shrinkage (see [9, 12, 30, 31]). Variational regularization has the advantage that no assumptions on the operator are required; it even works for nonlinear operators.

In the last decade it has become popular to formulate source conditions in the form of variational inequalities and to derive convergence rates for Tikhonov regularization under such variational source conditions [16, 20, 22, 34, 35]. Recently, the authors have developed a method for the verification of variational source conditions in Hilbert spaces under standard smoothness conditions. This method has been successfully applied to a number of interesting inverse problems, both linear and nonlinear [26, 27, 36, 42].

In this paper we extend our technique for the verification of variational source conditions to a Banach space setting. In particular, this allows us to derive variational source conditions for a certain class of operators if the true solution belongs to some Besov space $B^s_{p,q}$. This leads to optimal convergence rates for $p \in (1, 2]$ and $q \geq 2$. An important step is a new characterization of subgradient smoothness (see Theorem 3.6). As a second main novelty, we introduce a new technique to treat white noise and other stochastic noise models in nonquadratic variational regularization in an optimal way (Theorem 2.6). We obtain not only order optimal convergence rates of the expected error but also an estimate of the distribution of the error in terms of the distribution of a negative Besov norm of the noise process. At least in the case of white noise concentration, inequalities for such negative Besov norms are known.

The remainder of this paper is organized as follows: In section 2 we recall the definition of variational source conditions and the corresponding deterministic rates. Moreover, we formulate our new technique to derive error bounds for a statistical noise model in an abstract functional analytic setting. Then we introduce a strategy for the verification of a variational source condition including a characterization of subgradient smoothness in section 3. In sections 4 and 5 we present our results on convergence rates for finitely smoothing operators and for the backwards heat equation, respectively. Moreover, we show some numerical results for finitely smoothing operators confirming our theoretical error bounds. The paper has two appendices: one collecting properties of Besov spaces used in this paper and the other giving details on our numerical experiments.

2. Variational source conditions.

2.1. Basic definitions. Variational source conditions and the general error bounds in this section will be formulated in terms of the *Bregman distance* $\Delta_{\mathcal{R}}(f, f^\dagger)$ between $f, f^\dagger \in \mathcal{X}$ with respect to some convex functional $\mathcal{R}\colon \mathcal{X} \to (-\infty, \infty]$ defined by

$$\Delta_{\mathcal{R}}(f, f^\dagger) := \mathcal{R}(f) - \mathcal{R}(f^\dagger) - \langle f^*, f - f^\dagger \rangle,$$

where for the rest of this paper

$$f^* \in \partial\mathcal{R}(f^\dagger).$$

The use of Bregman distances in inverse problems was first proposed in [5, 13]. Here $\partial\mathcal{R}(f^\dagger)$ denotes the subdifferential of $\mathcal{R}$ at $f^\dagger$ (see [14]). In general $\Delta_{\mathcal{R}}(f, f^\dagger)$ depends on the choice of the subgradient $f^* \in \partial\mathcal{R}(f^\dagger)$, but in this paper we will only consider differentiable penalty functionals $\mathcal{R}$ such that $\partial\mathcal{R}(f^\dagger)$ is a singleton (see [14, Prop. I.5.3]). In this case $\Delta_{\mathcal{R}}$ is the second-order Taylor reminder, which can often be related to more familiar distance measures.

Example 2.1.

i. Let $\mathcal{X}$ be a Hilbert space; then, choosing $\mathcal{R}(f) = \frac{1}{2}\|f - f_0 \,|\, \mathcal{X}\|^2$ for some $f_0 \in \mathcal{X}$, one obtains that $\Delta_{\mathcal{R}}(f, f^\dagger) = \frac{1}{2}\big\|f - f^\dagger \,\big|\, \mathcal{X}\big\|^2$.

ii. If $\mathcal{X}$ is an r-convex Banach space, then by [4, Lemma 2.7] there exists a constant $C_{\mathcal{X}} > 0$ such that

$$\frac{C_{\mathcal{X}}}{r}\big\|f - f^\dagger \,\big|\, \mathcal{X}\big\|^r \leq \Delta_{\frac{1}{r}\|\cdot\,|\,\mathcal{X}\|^r}(f, f^\dagger), \qquad f, f^\dagger \in \mathcal{X}. \tag{2.1}$$

A variational source condition as first proposed in [22] for $\psi = \sqrt{\cdot}$ is an abstract smoothness condition for $f^\dagger$ of the form

$$\forall f \in \mathcal{D}\colon \quad \big\langle f^*, f^\dagger - f \big\rangle \leq \frac{1}{2}\Delta_{\mathcal{R}}(f, f^\dagger) + \psi\big(\|F(f^\dagger) - F(f) \,|\, \mathcal{Y}\|^t\big). \tag{2.2}$$

Here $\psi : [0, \infty) \to [0, \infty)$ is a concave index function; i.e., ψ is concave, continuous, and increasing and $\psi(0) = 0$. Section 3 is devoted to the interpretation of such conditions.

In [43, Ass. 1] the notion of an *effective noise level* $\mathbf{err}\colon \mathcal{Y} \to \mathbb{R}$ was introduced, and $\mathbf{err}(F(\widehat{f}_\alpha))$ bounds the effect of data noise on $\widehat{f}_\alpha$; see Proposition 2.3 and section 2.3 below. It is defined for some fixed $C_{\mathrm{err}} \geq 1$ by

$$\mathbf{err}(g) : = \mathcal{S}(g^\dagger) - \mathcal{S}(g) + \frac{1}{C_{\mathrm{err}} t}\big\|g - g^\dagger \,\big|\, \mathcal{Y}\big\|^t \tag{2.3}$$

with the data fidelity term $\mathcal{S}(g) := \frac{1}{t}\big\|g - g^{\mathrm{obs}} \,\big|\, \mathcal{Y}\big\|^t$ in case of (1.3) and $\mathcal{S}(g) := \frac{1}{2}\|g \,|\, \mathcal{Y}\|^2 - \big\langle g^{\mathrm{obs}}, g\big\rangle$ in case of (1.4), where we set $t := 2$. In the deterministic case we choose $C_{\mathrm{err}} = 2^{t-1}$ in (2.3), in which case one can show that $\mathbf{err}(g) \leq \overline{\mathbf{err}} := \frac{2}{t}\delta^t$; see [43, Ex. 3.1]. For bounds of $\mathbf{err}$ in the case of Poisson and impulsive noise, we refer the reader to [28, 43] and for bounds of $\mathbf{err}$ for the noise model (1.2) to section 2.3.

2.2. Convergence rates for deterministic errors. For the convergence rate theorem we will assume that ψ is concave and define for convenience $\varphi_\psi(\tau) := (-\psi)^*\left(-\frac{1}{\tau}\right)$ which governs the bias. Here ψ^* denotes the Fenchel conjugate for a convex function ψ given by $\psi^*(v) = \sup_u[\langle v, u\rangle - \psi(u)]$. Some corresponding calculus rules can be found, e.g., in [14].

Example 2.2. The most common examples of index functions ψ appearing in variational source conditions are of either Hölder or logarithmic type. If one extends these functions by $\psi(\tau) = -\infty$ for $\tau < 0$, one can calculate φ_ψ and obtain (see, e.g., [16])

$$\text{(2.4a)} \qquad \text{Hölder type:} \qquad \psi(\tau) = \tau^\mu \quad \Longrightarrow \quad \varphi_\psi(\tau) = c_\mu \tau^{\frac{\mu}{1-\mu}}$$

$$\text{(2.4b)} \qquad \text{logarithmic:} \qquad \begin{aligned} &\psi(\tau) = (-\ln \tau)^{-p}(1 + o(1)) \text{ as } \tau \to 0, \\ \Longrightarrow\ &\varphi_\psi(\tau) = (-\ln \tau)^{-p}(1 + o(1)) \text{ as } \tau \to 0. \end{aligned}$$

Under the assumption of a variational source condition, one can prove the following convergence rates in terms of the effective noise model.

PROPOSITION 2.3 (see [28, Thm. 2.3]). *Assume the variational source condition* (2.2) *holds true, and let* $\widehat{f}_\alpha$ *be a global minimizer of the Tikhonov functionals in* (1.3) *or* (1.4).

i. *Then* $\widehat{f}_\alpha$ *satisfies the following error bounds:*

$$\text{(2.5a)} \qquad \frac{1}{2}\Delta_{\mathcal{R}}(\widehat{f}_\alpha, f^\dagger) \le \frac{\mathbf{err}(F(\widehat{f}_\alpha))}{2\alpha} + \varphi_\psi(2C_{\mathrm{err}}\alpha),$$

$$\text{(2.5b)} \qquad \left\| F(\widehat{f}_\alpha) - g^\dagger \,\middle|\, \mathcal{Y} \right\|^2 \le 2C_{\mathrm{err}}\,\mathbf{err}(F(\widehat{f}_\alpha)) + 4C_{\mathrm{err}}\alpha\varphi_\psi(4C_{\mathrm{err}}\alpha).$$

ii. *If there exists some constant* $\overline{\mathbf{err}} \ge \mathbf{err}(F(f))$ *for all* f, *the infimum of the right-hand side of* (2.5a) *with* $\mathbf{err}(F(\widehat{f}_\alpha))$ *replaced by* $\overline{\mathbf{err}}$ *is attained if and only if* $\alpha = \bar{\alpha}$, *where* $\bar{\alpha}$ *is a solution to*

$$\text{(2.6a)} \qquad \frac{-1}{2C_{\mathrm{err}}\bar{\alpha}} \in \partial(-\psi)(2C_{\mathrm{err}}\overline{\mathbf{err}}).$$

In this case one obtains the convergence rate

$$\text{(2.6b)} \qquad \frac{1}{2}\Delta_{\mathcal{R}}(\widehat{f}_{\bar{\alpha}}, f^\dagger) \le C_{\mathrm{err}}\psi(\overline{\mathbf{err}}).$$

Actually in [28] only the deterministic case of (1.3) was considered, but the proof is the same for (1.4).

2.3. Convergence rates for random noise. From now on we assume that the reader has some basic knowledge about Besov spaces which can be found in Appendix A. In the following let Ω be either a bounded Lipschitz domain or the d-dimensional torus $\mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$ and $\mathscr{D}'(\Omega)$ the space of distributions on Ω. In this subsection we consider noise models (1.2) with a random variable Z in $\mathscr{D}'(\Omega)$. To prove convergence rates we need a deviation inequality for Z which is given by the following assumption.

Assumption 2.4. Assume that for all $\tilde{p} \in (1, \infty)$ we have $Z \in B^{-d/2}_{\tilde{p},\infty}(\Omega)$ almost surely and that there exist constants $C_Z, M_Z, \mu > 0$ such that

$$\forall t > 0\colon \quad \mathbb{P}\left(\left\| Z \,\middle|\, B^{-d/2}_{\tilde{p},\infty} \right\| > M_Z + t \right) \le \exp(-C_Z t^\mu).$$

If $Z = W$ is a Gaussian white noise and $\Omega = \mathbb{T}^d$, then Assumption 2.4 holds true with $\mu = 2$, M_Z being the median of $\|W \mid B^{-d/2}_{\tilde{p}',\infty}\|$ and C_Z depending on $\tilde{p}$ and d (see [41, Thm. 3.4, Cor. 3.7] or [18, remark after Thm. 4.4.3]). In the following let p' denote the Hölder conjugate for any number $1 \le p \le \infty$.

For random noise we choose $C_{\mathrm{err}} = 1$ in (2.3) and obtain $\mathbf{err}(g) = \varepsilon\langle Z, g - g^\dagger\rangle$; hence, it is natural to estimate the error functional for $p \in (1, 2]$ via

$$\mathbf{err}(g) = \varepsilon\big\langle Z, g - g^\dagger\big\rangle \le \varepsilon\Big\|Z \,\Big|\, B^{-d/2}_{p',\infty}\Big\|\Big\|g - g^\dagger \,\Big|\, B^{d/2}_{p,1}\Big\|, \tag{2.7}$$

and the challenge is to find a good control of the second factor. We will show that this can be done via an interpolation approach which for our two model problems results again in optimal rates (for $q \ge 2$). To formulate a general error bound, we will assume for the moment that the second factor can be estimated as follows.

Assumption 2.5. There exist constants $C, \beta, \gamma > 0$ such that the inequality

$$\Big\|F(f_1) - F(f_2) \,\Big|\, B^{d/2}_{p,1}\Big\| \le C\big\|F(f_1) - F(f_2) \,\big|\, L^2\big\|^\beta \Delta_{\mathcal{R}}(f_1, f_2)^\gamma \tag{2.8}$$

holds true for all $f_1, f_2 \in \mathcal{D}$.

Often this assumption can be verified by Remark 3.2 below.

THEOREM 2.6. *Let a variational source condition* (2.2) *and Assumption* 2.5 *be fulfilled, and let* $\widehat{f}_\alpha$ *be a global minimizer of the Tikhonov functional in* (1.4).

i. *If* $0 < \beta < 2$, *then the effective noise level at* $F(\widehat{f}_\alpha)$ *is bounded by*

$$\mathbf{err}\Big(F(\widehat{f}_\alpha)\Big) \le C\Big\|\varepsilon Z \,\Big|\, B^{-d/2}_{p',\infty}\Big\|^{\frac{2}{2-\beta}} \Delta_{\mathcal{R}}(f, f^\dagger)^{\frac{2\gamma}{2-\beta}} + 2\alpha\varphi_\psi(4\alpha).$$

ii. *If in addition* $0 < \gamma < \frac{1}{2}(2-\beta)$, *this implies the error bound*

$$\frac{1}{2}\Delta_{\mathcal{R}}(\widehat{f}_\alpha, f^\dagger) \le C\alpha^{-\frac{2-\beta}{2-\beta-2\gamma}}\Big\|\varepsilon Z \,\Big|\, B^{-d/2}_{p',\infty}\Big\|^{\frac{2}{(2-\beta)-2\gamma}} + 4\varphi_\psi(4\alpha).$$

Proof. For (i) note that due to Assumption 2.5 we obtain

$$\begin{aligned}\mathbf{err}\Big(F(\widehat{f}_\alpha)\Big) &= \Big\langle \varepsilon Z, F(\widehat{f}_\alpha) - g^\dagger\Big\rangle \le \Big\|\varepsilon Z \,\Big|\, B^{-d/2}_{p',\infty}\Big\|\Big\|F(\widehat{f}_\alpha) - g^\dagger \,\Big|\, B^{d/2}_{2,1}\Big\| \\ &\le C\Big\|\varepsilon Z \,\Big|\, B^{-d/2}_{p',\infty}\Big\|\Big\|F(\widehat{f}_\alpha) - g^\dagger \,\Big|\, L^2\Big\|^\beta \Delta_{\mathcal{R}}(\widehat{f}_\alpha, f^\dagger)^\gamma.\end{aligned}$$

By the image space convergence rate (2.5b) of Proposition 2.3 we can estimate

$$\begin{aligned}\mathbf{err}\Big(F(\widehat{f}_\alpha)\Big) &\le C\Big\|\varepsilon Z \,\Big|\, B^{-d/2}_{p',\infty}\Big\|\Big[2\,\mathbf{err}\Big(F(\widehat{f}_\alpha)\Big) + 4\alpha\varphi_\psi(4\alpha)\Big]^{\beta/2}\Delta_{\mathcal{R}}(\widehat{f}_\alpha, f^\dagger)^\gamma \\ &\le C\Big\|\varepsilon Z \,\Big|\, B^{-d/2}_{p',\infty}\Big\|^{\frac{2}{2-\beta}}\Delta_{\mathcal{R}}(\widehat{f}_\alpha, f^\dagger)^{\frac{2\gamma}{2-\beta}} + \frac{1}{2}\Big[\mathbf{err}\Big(F(\widehat{f}_\alpha)\Big) + 2\alpha\varphi_\psi(4\alpha)\Big]\end{aligned}$$

by Young's inequality with a generic constant $C > 0$. Rearranging terms yields the bound on the effective noise level.

To prove (ii), note that due to (2.5a) in Proposition 2.3 we have

$$\Delta_{\mathcal{R}}(\widehat{f}_\alpha, f^\dagger) \le \frac{\mathbf{err}(F(\widehat{f}_\alpha))}{\alpha} + 2\varphi_\psi(2\alpha) \le \frac{\mathbf{err}(F(\widehat{f}_\alpha))}{\alpha} + 2\varphi_\psi(4\alpha).$$

Together with the first part we obtain

$$\begin{aligned}\Delta_{\mathcal{R}}(\widehat{f}_\alpha, f^\dagger) &\le C\alpha^{-1}\Delta_{\mathcal{R}}(\widehat{f}_\alpha, f^\dagger)^{\frac{2\gamma}{2-\beta}}\Big\|\varepsilon Z \,\Big|\, B^{-d/2}_{p',\infty}\Big\|^{\frac{2}{2-\beta}} + 4\varphi_\psi(4\alpha) \\ &\le \frac{1}{2}\Delta_{\mathcal{R}}(\widehat{f}_\alpha, f^\dagger) + C\alpha^{-\frac{2-\beta}{2-\beta-2\gamma}}\Big\|\varepsilon Z \,\Big|\, B^{-d/2}_{p',\infty}\Big\|^{\frac{2}{2-\beta-2\gamma}} + 4\varphi_\psi(4\alpha),\end{aligned}$$

which proves the claim. ∎

3. Verification of variational source conditions. In [27, Thm. 2.1] two of the authors formulated a strategy for the verification of variational source conditions in terms of orthogonal projection operators. In this section we will extend this strategy to Banach space settings. It turns out that the smoothness of subgradients of the solution rather than the smoothness of the solution itself determines the convergence rate. Therefore, a crucial step will be the analysis of the smoothness of subgradients in Besov spaces.

3.1. Preliminaries. One of the main difficulties when trying to prove a variational source condition is the Bregman distance appearing on the right-hand side of (2.2) since its properties depend very much on the specific choice of the regularization functional. Hence, we are going to assume that it can be estimated from below by some norm power.

Assumption 3.1. There exist constants $C_\Delta > 0$ and $r > 1$ such that

$$C_\Delta \|f_1 - f_2 \,|\, \mathcal{X}\|^r \leq \Delta_{\mathcal{R}}(f_2, f_1) \qquad \text{for all } f_1, f_2 \in \mathcal{X}.$$

This assumption is satisfied in particular if $\mathcal{X}$ is convex of power type and $\mathcal{R}$ is a norm power; see Example 2.1. However, the case of $\mathcal{R}(\cdot) = \|\cdot \,|\, \ell^1\|$ shows that this is not always the case; if, e.g., $f_1, f_2 > 0$, then $\Delta_R(f_2, f_1) = 0$, but $\|f_1 - f_2 \,|\, \ell^1\|$ might be arbitrary large.

Remark 3.2. Under Assumption 3.1, (2.8) in Assumption 2.5 is fulfilled if

$$\left\| F(f_1) - F(f_2) \,\middle|\, B^{d/2}_{p,1} \right\| \leq C \left\| F(f_1) - F(f_2) \,\middle|\, L^2 \right\|^\beta \|f_1 - f_2 \,|\, \mathcal{X}\|^{\gamma r}. \tag{3.1}$$

Note that for linear operators F, one necessarily has $\beta + \gamma r = 1$. For nonlinear operators F, Assumption 2.5 can also be verified by standard interpolation inequalities if F maps Lipschitz continuously into some space of higher regularity.

3.2. Basic strategy. Our main tool for the derivation of variational source conditions will be the following generalization of Theorem 2.1 in [27].

THEOREM 3.3. *Let $\mathcal{X}$ and $\mathcal{Y}$ be Banach spaces and $\mathcal{R}$ a penalty term such that Assumption* 3.1 *is fulfilled. Let $f^\dagger \in \mathcal{D}$ and $f^* \in \partial\mathcal{R}(f^\dagger)$. Suppose that there exists a family of operators $P_j \colon \mathcal{X}^* \to \mathcal{X}^*$ for $j \in J$ an index set such that for some functions $\kappa, \sigma \colon J \to (0, \infty)$ and a constant $\gamma \geq 0$ the following holds true for all $j \in J$:*

$$\|(I - P_j) f^* \,|\, \mathcal{X}^*\| \leq \kappa(j), \tag{3.2a}$$

$$\inf_{j \in J} \kappa(j) = 0, \tag{3.2b}$$

$$\begin{gathered} \langle P_j f^*, f^\dagger - f \rangle \leq \sigma(j) \left\| F(f^\dagger) - F(f) \,\middle|\, \mathcal{Y} \right\| + \gamma \kappa(j) \left\| f^\dagger - f \,\middle|\, \mathcal{X} \right\| \\ \textit{for all } f \in \mathcal{D} \textit{ with } \left\| f^\dagger - f \,\middle|\, \mathcal{X} \right\| \leq \left(\tfrac{2}{C_\Delta} \|f^* \,|\, \mathcal{X}^*\| \right)^{\frac{r'}{r}}. \end{gathered} \tag{3.2c}$$

Then $f^\dagger$ fulfills a variational source condition (2.2) *with the concave index function*

$$\psi_{\mathrm{vsc}}(\tau) = \inf_{j \in J} \left[\sigma(j) \tau^{1/t} + \frac{1}{r'} \left(\tfrac{2}{C_\Delta} \right)^{r'/r} (1 + \gamma)^{r'} \kappa(j)^{r'} \right]. \tag{3.3}$$

Condition (3.2a) describes the smoothness of the solution (actually rather the smoothness of the subdifferential, but in the examples considered later one of the two uniquely determines the other; see Theorem 3.6), whereas (3.2c) describes the local ill-posedness of the problem.

Example 3.4. In order to illustrate these interpretations, consider the case that $\mathcal{X},\mathcal{Y}$ are Hilbert spaces and F an injective compact operator, and let $(f_j,g_j,\sigma_j)_{j\in\mathbb{N}}$ be the corresponding singular system. Set $P_jf=\sum_{k\le j}\langle f,f_j\rangle f_j$. Then we obtain that $f^*=f^\dagger$ and (3.2a) reads

$$\left\|(I-P_j)f^\dagger \,\middle|\, \mathcal{X}\right\| = \left(\sum_{k>j}|\langle f^\dagger,f_j\rangle|^2\right)^{1/2} =: \kappa(j);$$

i.e., κ measures the decay rate of the coefficients of $f^\dagger$ in the system $(f_j)_{j\in\mathbb{N}}$. In the case where f_j are trigonometric polynomials, this measures classical smoothness. Denoting by $Q_jg=\sum_{k\le j}\langle g,g_j\rangle g_j$ we obtain an inequality of the form (3.2c) via

$$\begin{aligned}\langle P_jf^\dagger,f^\dagger-f\rangle &\le \|P_jf^\dagger\,|\,\mathcal{X}\|\|P_j(f^\dagger-f)\,|\,\mathcal{X}\| \le \|P_jf^\dagger\,|\,\mathcal{X}\|\frac{1}{\sigma_j}\|Q_jT(f^\dagger-f)\,|\,\mathcal{Y}\|\\ &= \left(\sum_{k\le j}\frac{|\langle f^\dagger,f_j\rangle|^2}{\sigma_j^2}\right)^{1/2}\|T(f^\dagger-f)\,|\,\mathcal{Y}\| = \sigma(j)\|T(f^\dagger-f)\,|\,\mathcal{Y}\|\end{aligned}$$

with $\sigma(j):=(\sum_{k\le j}|\langle f^\dagger,f_j\rangle|^2)^{1/2}/\sigma_j\le\|f^\dagger\,|\,\mathcal{X}\|/\sigma_j$; i.e., σ measures the decay rate of the singular values of F relative to the decay rate of the coefficients of $f^\dagger$ in the singular system.

Proof of Theorem 3.3. First assume that f does not satisfy the condition in the second line of (3.2c) or equivalently that $\|f^*\,|\,\mathcal{X}^*\|\le\frac{C_\Delta}{2}\|f^\dagger-f\,|\,\mathcal{X}\|^{r-1}$. Then

$$\left\langle f^*,f^\dagger-f\right\rangle \le \|f^*\,|\,\mathcal{X}^*\|\|f^\dagger-f\,|\,\mathcal{X}\| \le \tfrac{C_\Delta}{2}\|f^\dagger-f\,|\,\mathcal{X}\|^r \le \tfrac12\Delta_{\mathcal{R}}(f,f^\dagger);$$

that is, the variational source condition holds true even with $\psi\equiv0$. Otherwise, using (3.2a), (3.2c), and Young's inequality we get for each $j\in J$ that

$$\begin{aligned}&\left\langle f^*,f^\dagger-f\right\rangle\\ &=\left\langle P_jf^*,f^\dagger-f\right\rangle+\left\langle (I-P_j)f^*,f^\dagger-f\right\rangle\\ &\le\sigma(j)\left\|F(f^\dagger)-F(f)\,\middle|\,\mathcal{Y}\right\|+(1+\gamma)\kappa(j)\left\|f^\dagger-f\,\middle|\,\mathcal{X}\right\|\\ &\le\sigma(j)\left\|F(f^\dagger)-F(f)\,\middle|\,\mathcal{Y}\right\|+\tfrac{1}{r'}\left(\tfrac{2}{C_\Delta}\right)^{r'/r}(1+\gamma)^{r'}\kappa(j)^{r'}+\tfrac{C_\Delta}{2r}\left\|f^\dagger-f\,\middle|\,\mathcal{X}\right\|^r\\ &\le\sigma(j)\left\|F(f^\dagger)-F(f)\,\middle|\,\mathcal{Y}\right\|+\tfrac{1}{r'}\left(\tfrac{2}{C_\Delta}\right)^{r'/r}(1+\gamma)^{r'}\kappa(j)^{r'}+\tfrac12\Delta_{\mathcal{R}}(f,f^\dagger).\end{aligned}$$

Taking the infimum over the right-hand side with respect to $j\in J$ yields (3.3) with $\tau=\|F(f^\dagger)-F(f)\,|\,\mathcal{Y}\|^t$.

Note that ψ is defined as an infimum over concave and increasing functions and hence is also increasing and concave. By (3.2b) we obtain that $\psi(0)=0$, and hence ψ is indeed an index function. ∎

For linear operators and under the conditions below, one can choose $\gamma=0$, and the additional restriction that one needs (3.2c) only for $\left\|f^\dagger-f\,\middle|\,\mathcal{X}\right\|$ small is not necessary as already seen in the specific example. We have included both complications here since they may be needed for some nonlinear operators; see, e.g., [26]. We will present two different applications of the strategies in sections 4 and 5 and discuss how they may generalize to a similar setting in Remarks 4.9 and 5.2.

3.3. Besov spaces. Recall that we assume Ω to be either a bounded Lipschitz domain or $\mathbb{T}^d$. We now introduce two equivalent norms on $B^s_{p,q}$; further details on these spaces can be found in Appendix A.

3.3.1. Wavelet systems. For any $j \in \mathbb{N}_0$ let I_j be a countable index set, let $I := \{(j,l) : j \in \mathbb{N}_0, l \in I_j\}$, and let $(\phi_{j,l})_{(j,l)\in I} \subset L^2(\Omega)$ be an orthonormal system. Assume that for some $\sigma \in \mathbb{N}$ the system fulfills $\phi_{j,l} \in C^\sigma(\overline{\Omega}, \mathbb{R})$. For $f \in (C^\sigma(\overline{\Omega}, \mathbb{R}))'$ define the linear mapping

$$\mathcal{W}f := \lambda := (\lambda_{j,l})_{(j,l)\in I}, \qquad \text{where} \qquad \lambda_{j,l} := \int_\Omega f(x)\phi_{j,l}(x)\,\mathrm{d}x. \tag{3.4a}$$

We assume that this system also generates Besov spaces in the following way: For all $p, q \in [1,\infty]$ and $s \in \mathbb{R}$ with $|s| < \sigma$ we get that $f \in \mathcal{D}'(\Omega)$ belongs to $B^s_{p,q}(\Omega)$ if and only if

$$\|f \mid B^s_{p,q}\|_{\mathcal{W}} := \left[\sum_{j\in\mathbb{N}_0} 2^{jsq} 2^{jd(\frac{1}{2}-\frac{1}{p})q} \left(\sum_{l\in I_j} |\lambda_{j,l}|^p\right)^{\frac{q}{p}}\right]^{\frac{1}{q}} \tag{3.4b}$$

(with the usual modifications if $p = \infty$ or $q = \infty$) is finite. In this case, (3.4b) is an equivalent norm on $B^s_{p,q}$. Furthermore, this usually implies that

$$\text{the decomposition} \qquad f := \mathcal{W}^*\lambda = \sum_{(j,l)\in I} \lambda_{j,l}\phi_{j,l} \qquad \text{is unique} \tag{3.4c}$$

with unconditional convergence in $\mathcal{D}'(\Omega)$ and local convergence in $B^u_{p,q}$ for all $u < s$ (even in $B^s_{p,q}$ if $p \neq \infty$ and $q \neq \infty$).

Example 3.5. The properties above are satisfied in particular in the following cases:

i. Let $\Omega = \mathbb{T}^d$, and let $(\tilde{\phi}_{j,l})_{(j,l)\in\tilde{I}}$ be either the d-dimensional Daubechies wavelet system of order $n \in \mathbb{N}$ or the Meyer wavelet system for $\mathbb{R}^d$, where $\tilde{I} = \{(j,l) : j \in \mathbb{N}_0, l \in \tilde{I}_j\}$ with $I_0 := \mathbb{Z}^d$ and $I_j = \{1,\dots,2^d-1\}\times\mathbb{Z}^d$. Define periodization $\phi_{j,l}(x) := \sum_{z\in\mathbb{Z}^d} \tilde{\phi}_{j,l}(x-z)$ for $x \in \mathbb{T}^d$ of $\tilde{\phi}_{j,l}$ for $l \in I_j \subset \tilde{I}_j$ with $I_0 := \{0\}$ and $I_j := \{1,\dots,2^d-1\} \times \{z \in \mathbb{Z}^d \colon 0 \le z_i < 2^j\}$. Then $(\phi_{j,l})_{(j,l)\in I}$ is an orthonormal system in $L^2(\mathbb{T}^d)$. Furthermore, for fixed $\sigma \in \mathbb{N}$ the system fulfills (3.4), where in case of the Daubechies wavelet system we have to choose n large enough (one has $\phi_{j,l} \in C^1$ for $n \ge 3$ and $\phi_{j,l} \in C^2$ for $n \ge 7$, while for large n the asymptotic formula $\phi_{j,l} \in C^\sigma$ for $\sigma > 0.2n$ holds true; see [10, sect. 7]).
ii. If $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain, there are different ways to define Besov spaces that differ mainly in how to treat boundary values. Let

$$B^s_{p,q}(\Omega) := \begin{cases} \{f \in \mathcal{D}'(\Omega) : f = h|_\Omega, h \in B^s_{p,q}(\mathbb{R}^d)\} & \text{if } s \le 0, \\ \{f \in \mathcal{D}'(\Omega) : f = h|_\Omega, h \in B^s_{p,q}(\mathbb{R}^d), \operatorname{supp} h \subset \overline{\Omega}\} & \text{if } s > 0, \end{cases}$$

and

$$\|f \mid B^s_{p,q}(\Omega)\| := \inf \|h \mid B^s_{p,q}(\mathbb{R}^d)\|,$$

where the infimum is taken over all extensions h as above. For $B^s_{p,q}(\Omega)$ defined like this, an explicit construction of an orthonormal system based on the Daubechies orthonormal wavelet system and a Whitney decomposition fulfilling (3.4) is carried out in [38, Thms. 2.33 and 3.23]. For computationally more feasible constructions of orthogonal wavelets on the interval (or boxes via tensor product constructions) we refer the reader to [8] and with respect to boundary conditions to [2].

Motivated by the above example we will call $(\phi_{j,l})_{(j,l)\in I}$ a wavelet system and $\mathcal{W}$ the wavelet transform. In the following we will always assume that the wavelet system is such that (3.4c) holds true, and hence we have a norm on $B^s_{p,q}(\Omega)$ given by (3.4b).

3.3.2. Fourier system. In case that $\Omega = \mathbb{T}^d$, an equivalent norm on $B^s_{p,q}$ for $p \in (1,\infty)$, $q \in [1,\infty]$ and $s \in \mathbb{R}$ is given by the following: Denote by $\chi_0(x)$ the characteristic function of the unit square in $\mathbb{R}^d$ and by the system $(\chi_j)_{j\in\mathbb{N}_0}$ the corresponding dyadic resolution of unity, that is,

$$\chi_0(x) = \begin{cases} 1 & |x|_\infty \le 1 \\ 0 & \text{else} \end{cases} \quad \text{and} \quad \chi_j(x) := \chi_0(2^{-j}x) - \chi_0(2^{-j+1}x) \tag{3.5a}$$

for $j \in \mathbb{N}$. Furthermore, we will denote by

$$\mathcal{F}f := \hat{f} := \left(\hat{f}(z)\right)_{z\in\mathbb{Z}^d}, \quad \text{where} \quad \hat{f}(z) := \int_{\mathbb{T}^d} f(x)\overline{e_z(x)}\,\mathrm{d}x \quad \text{with} \quad e_z(x) := \mathrm{e}^{2\pi \mathrm{i} x\cdot z},$$

the Fourier transform on $\mathbb{T}^d$. Further define

$$I_j := \{z \in \mathbb{Z}^d \colon \chi_j(2\pi z) = 1\}. \tag{3.5b}$$

A norm on $B^s_{p,q}$ is then defined by

$$\left\|f \,\middle|\, B^s_{p,q}\right\| := \left[\sum_{j\in\mathbb{N}_0} 2^{jsq} \left\| \sum_{l\in I_j} \hat{f}(l) e_l \,\middle|\, L^p(\mathbb{T}^d) \right\|^q \right]^{\frac{1}{q}} \tag{3.5c}$$

(see [38, sect. 1.3]) with the usual modification for $q = \infty$. Note that while it is "standard" to introduce these spaces via a dyadic resolution of unity as we do above, it is usually assumed that this resolution is also smooth (see, e.g., [39, sect. 2.3]). However, this is not required for the range of the parameter $p \in (1,\infty)$ which we are considering here (see, e.g., [39, sect. 2.5.4]).

3.4. Subgradient smoothness. If $\mathcal{X}$ is not a Hilbert space, then the mapping $f \mapsto f^*$ is no longer the identity mapping. While the continuity properties of this mapping have been studied for some time (see, e.g., [6] and references therein), much less is known on the question whether additional smoothness of f yields additional smoothness of f^*. Although not stated explicitly in this form, the results in [32, 33] essentially show that for $B^{s_2}_{p_2,p_2} \subset B^{s_1}_{p_1,p_1}$ and $f^* \in \partial \frac{1}{p_1} \|f \,|\, B^{s_1}_{p_1,p_1}\|^{p_1}_{\mathcal{W}}$ for a smooth enough wavelet system the relation

$$f \in B^{s_2}_{p_2,p_2} \quad \Longrightarrow \quad f^* \in B^{-s_1}_{p_1',p_1'} \cap B^{s_3}_{p_3,p_3} \tag{3.6}$$

holds true where $p_3 = \frac{p_2}{p_1 - 1}$ and $s_3 = -s_1 + (s_2 - s_1)(p_1 - 1)$. The proof of (3.6) can be carried out along the lines of the proof of the next theorem, showing that it is even an equivalence result. For our new result, we restrict to the case that $p_1 = p_2$, allowing for different fine indices q_j.

THEOREM 3.6. *Let $p, q_1 \in (1,\infty)$, $q_2 \in [1,\infty]$, $s_1, s_2 \in \mathbb{R}$, and $r > 0$. Set $s_3 = -s_1 + (s_2 - s_1)(q_1 - 1)$ and $q_3 = \frac{q_2}{q_1 - 1}$, and assume that the chosen wavelet system fulfills the assumptions in section 3.3 with $\sigma > \max\{|s_1|, |s_2|, |s_3|\}$. Let $f^* \in \partial \frac{1}{r} \| f \mid B^{s_1}_{p,q_1} \|^r_{\mathcal{W}}$; then $f \in B^{s_1}_{p,q_1} \cap B^{s_2}_{p,q_2}$ if and only if $f^* \in B^{-s_1}_{p',q_1'} \cap B^{s_3}_{p',q_3}$. Furthermore,*

$$\left\| f^* \,\middle|\, B^{s_3}_{p',q_3} \right\|_{\mathcal{W}} = \| f \mid B^{s_1}_{p,q_1} \|^{r-q_1}_{\mathcal{W}} \| f \mid B^{s_2}_{p,q_2} \|^{q_1 - 1}_{\mathcal{W}}.$$

Proof. We obtain $f^* \in B^{-s_1}_{p',q_1'}$ directly by the mapping properties of the subdifferential.

Denote by $\lambda = \mathcal{W} f$ the wavelet decomposition of f. For the given range of the parameters p, q_1, r, the norm is differentiable, therefore one obtains that $f^* \in \partial \frac{1}{r} \| f \mid B^{s_1}_{p,q_1} \|^r$ if and only if $f^* = \mathcal{W}^* \mu = \sum_{(j,l)\in I} \mu_{j,l} \phi_{j,l}$, where

$$\mu_{j,l} = \| f \mid B^{s_1}_{p,q_1} \|^{r-q_1} 2^{j s_1 q_1} 2^{jd(\frac{1}{2} - \frac{1}{p}) q_1} \left[\sum_{m \in I_j} |\lambda_{j,m}|^p \right]^{\frac{1}{p}(q_1 - p)} \frac{\lambda_{j,l}}{|\lambda_{j,l}|^{2-p}}.$$

For $q_3 \neq \infty$ (the case $q_3 = \infty$ follows along the same lines) we get

$$\begin{aligned}
&\left\| f^* \,\middle|\, B^{s_3}_{p',q_3} \right\|^{q_3} / \| f \mid B^{s_1}_{p,q_1} \|^{q_3(r-q_1)} \\
&= \sum_{j \in \mathbb{N}_0} 2^{j(s_3 + s_1 q_1) q_3} 2^{jd(\frac{1}{2} - \frac{1}{p'} + q_1(\frac{1}{2} - \frac{1}{p})) q_3} \left[\sum_{l \in I_j} |\lambda_{j,l}|^p \right]^{\frac{q_3}{p}(q_1 - p) + \frac{q_3}{p'}} \\
&= \sum_{j \in \mathbb{N}_0} 2^{j s_2 q_2} 2^{jd(\frac{1}{2} - \frac{1}{p}) q_2} \left[\sum_{l \in I_j} |\lambda_{j,l}|^p \right]^{\frac{q_2}{p}} \\
&= \| f \mid B^{s_2}_{p,q_2} \|^{q_2};
\end{aligned}$$

hence, taking the q_3-root proves the claim.

For the "only if" part note that by duality $f^* \in \partial \frac{1}{r} \| f \mid B^s_{p,q} \|^r$ implies that $f \in \partial \frac{1}{r'} \| f^* \mid B^{-s}_{p',q'} \|^{r'}$; hence, we also have that the implication $f^* \in B^{-s_1}_{p',q_1'} \cap B^{s_3}_{p',q_3}$ implies $f \in B^{s_1}_{p,q_1} \cap B^{s_2}_{p,q_2}$. ∎

The interesting case of the theorem above is if either $s_1 < s_2$ or if $s_1 = s_2$ and $q_2 < q_1$(and hence $B^{s_1}_{p,q_1} \cap B^{s_2}_{p,q_2} = B^{s_2}_{p,q_2}$ in both cases) since in these cases $B^{s_3}_{p',q_3}$ is a proper subspace of $B^{-s_1}_{p',q_1'}$ and not the other way around. Otherwise—i.e., if $B^{s_1}_{p,q_1} \cap B^{s_2}_{p,q_2} = B^{s_1}_{p,q_1}$—the explicit expression of the norm might still be useful. We would like to highlight one special case of the above theorem. If $q_2 = \infty$, that is, f is in the largest space with smoothness s, then we also obtain $q_3 = \infty$. This is interesting because Besov spaces $B^s_{2,\infty}$ are known to be maximal sets for L^2-regularization for certain problems (see [27]).

From now on we will always assume sufficient smoothness of the wavelet system in the sense of the previous theorem without further mentioning.

3.5. Bernstein- and Jackson-type inequalities.

Assumption 3.7. Let $(P_j)_{j\in\mathbb{N}_0}\colon \mathcal{D}'(\Omega)\to\mathcal{D}'(\Omega)$ be a family of linear operators such that for all $p\in(1,\infty)$, q, $\tilde q\in[1,\infty]$ and s, $t\in\mathbb{R}$ with $\max\{|s|,|t|\}\le\sigma$ the norm bounds

$$\begin{aligned}&\text{if } t>s: && \left\|P_jf \,\middle|\, B^t_{p,\tilde q}\right\|\le c_1 2^{j(t-s)}\left\|f \,\middle|\, B^s_{p,q}\right\|\\ &\text{if } t<s: && \left\|(I-P_j)f \,\middle|\, B^t_{p,\tilde q}\right\|\le c_2 2^{j(t-s)}\left\|f \,\middle|\, B^s_{p,q}\right\|\end{aligned}$$

with constant $c_1=c_1(p,q,\tilde q,s,t)>0$ and $c_2=c_1(p,q,\tilde q,s,t)>0$, called Bernstein and Jackson inequality, respectively, hold true.

Example 3.8. The following are possible choices of the operators $(P_j)_{j\in\mathbb{N}_0}$.

i. Let Ω be a Lipschitz domain or the torus as in section 3.3 and $\mathcal{W}$ the wavelet transform defined in (3.4a). Set for $j\in\mathbb{N}_0$

$$P_jf := \mathcal{W}^*Q_j\mathcal{W}f,$$

$$\text{where}\qquad (Q_j\lambda)_{k',l'} := \begin{cases}\lambda_{k',l'} & k'\le j\\ 0 & \text{else}\end{cases}\qquad \text{for all } (k',l')\in\mathbb{N}_0\times I_k. \tag{3.7a}$$

Due to (3.4b) this immediately implies a Bernstein and Jackson inequality of the desired form:

(a) *Bernstein inequality:* For $t>s$, $\tilde q\in[1,\infty]$ and $f\in B^s_{p,q}$ we get

$$\begin{aligned}\left\|P_jf \,\middle|\, B^t_{p,\tilde q}\right\|^{\tilde q} &= \sum_{k\le j}2^{k(t-s)\tilde q}\left(2^{ks}2^{kd(\frac12-\frac1p)}\left(\sum_{l\in I_k}|\lambda_{k,l}|^p\right)^{\frac1p}\right)^{\tilde q}\\ &\le \sum_{k\le j}2^{k(t-s)\tilde q}\left\|f \,\middle|\, B^s_{p,\infty}\right\|^{\tilde q}\le c2^{j(t-s)\tilde q}\left\|f \,\middle|\, B^s_{p,q}\right\|^{\tilde q}\end{aligned}$$

for some constant c depending on t, s, and $\tilde q$ only.

(b) *Jackson inequality:* For $t<s$, $\tilde q\in[1,\infty]$ and $f\in B^s_{p,q}$ we obtain

$$\begin{aligned}\left\|(I-P_j)f \,\middle|\, B^t_{p,\tilde q}\right\|^{\tilde q} &= \sum_{k>j}2^{k(t-s)\tilde q}\left(2^{ks}2^{kd(\frac12-\frac1p)}\left(\sum_{l\in I_k}|\lambda_{k,l}|^p\right)^{\frac1p}\right)^{\tilde q}\\ &\le \sum_{k>j}2^{k(t-s)\tilde q}\left\|f \,\middle|\, B^s_{p,\infty}\right\|^{\tilde q}\le c2^{j(t-s)\tilde q}\left\|f \,\middle|\, B^s_{p,q}\right\|^{\tilde q},\end{aligned}$$

where the constant c depends again on t, s, and $\tilde q$ only.

ii. Let $\Omega=\mathbb{T}^d$, and let the norm of $B^s_{p,q}$ be given by (3.5). Then set

$$P_j := \mathcal{F}^*\left(\sum_{k\le j}\chi_k(2\pi\cdot)\right)\mathcal{F} \tag{3.7b}$$

for $j\in\mathbb{N}_0$ to get the following:

(a) *Bernstein inequality:* For $t>s$, $\tilde q\in[1,\infty]$ and $f\in B^s_{p,q}$ we get

$$\begin{aligned}\left\|P_jf \,\middle|\, B^t_{p,\tilde q}\right\|^{\tilde q} &= \sum_{k\le j}2^{k(t-s)\tilde q}\left(2^{ks\tilde q}\left\|\sum_{l\in I_k}\hat f(l)e_l \,\middle|\, L^p(\mathbb{T}^d)\right\|\right)^{\tilde q}\\ &\le \sum_{k\le j}2^{k(t-s)\tilde q}\left\|f \,\middle|\, B^s_{p,\infty}\right\|^{\tilde q}\le c2^{j(t-s)\tilde q}\left\|f \,\middle|\, B^s_{p,q}\right\|^{\tilde q}\end{aligned}$$

for some constant c depending on t, s, and $\tilde q$ only.

(b) *Jackson inequality:* For $t < s$, $\tilde{q} \in [1,\infty]$ and $f \in B^s_{p,q}$ we obtain

$$\begin{aligned}\left\|(I-P_j)f \,\middle|\, B^t_{p,\tilde{q}}\right\|^{\tilde{q}} &= \sum_{k>j} 2^{k(t-s)\tilde{q}} \left(2^{ks\tilde{q}} \left\| \sum_{l\in I_k} \hat{f}(l) e_l \,\middle|\, L^p(\mathbb{T}^d)\right\|\right)^{\tilde{q}} \\ &\le \sum_{k>j} 2^{k(t-s)\tilde{q}} \left\|f \,\middle|\, B^s_{p,\infty}\right\|^{\tilde{q}} \le c 2^{j(t-s)\tilde{q}} \left\|f \,\middle|\, B^s_{p,q}\right\|^{\tilde{q}},\end{aligned}$$

where the constant c depends again on t, s, and $\tilde{q}$ only.

Further possibilities include projections onto spline or finite element subspaces.

COROLLARY 3.9. *Let $1 < p, q < \infty$, $r = \max\{2, p, q\}$, and $s > 0$. Let $f^\dagger \in B^s_{p,\infty}$ with $\|f^\dagger \,|\, B^s_{p,\infty}\|_{\mathcal{W}} \le \varrho$ for some $\varrho > 0$, and $f^* \in \partial \frac{1}{r}\|f^\dagger \,|\, B^0_{p,q}\|^r_{\mathcal{W}}$. Then $\|f^* \,|\, B^{s(q-1)}_{p',\infty}\| \le c\varrho^{r-1}$, and if $(P_j)_{j\in\mathbb{N}}$ is chosen according to Assumption 3.7 and $a > s(q-1)$, there exists some constant $c > 0$ such that*

$$\left\|P_j f^* \,\middle|\, B^a_{p',q'}\right\| \le c\varrho^{r-1} 2^{j(a-s(q-1))} \quad \text{and} \quad \left\|(I-P_j) f^* \,\middle|\, B^0_{p',q'}\right\| \le c\varrho^{r-1} 2^{-js(q-1)}.$$

Proof. By Theorem 3.6 we get $f^* \in B^{s(q-1)}_{p',\infty}$ together with a norm bound. Inserting this bound into Assumption 3.7, where P_j and $I - P_j$ are applied to f^*, gives the desired inequalities. ∎

We will see in section 4 that for a wide class of applications, these inequalities are enough to verify variational source conditions.

3.6. Deterministic lower bounds. In order to see whether the convergence rates implied by variational source conditions are of optimal order, we need to find a lower bound on these rates. Such a bound is provided by the modulus of continuity, and this lower bound is known to be sharp in Hilbert spaces (see [40]).

DEFINITION 3.10. *Let $F: \mathcal{X} \to \mathcal{Y}$ be continuous and injective, and let $\mathcal{K} \subset \mathcal{X}$ be compact. Then the* modulus of continuity *$\omega(\delta, \mathcal{K})$ of $(F|_{\mathcal{K}})^{-1}$ is defined by*

$$\omega(\delta,\mathcal{K}) := \sup\{\|f_1 - f_2 \,|\, \mathcal{X}\| \colon f_1, f_2 \in \mathcal{K}, \|F(f_1) - F(f_2) \,|\, \mathcal{Y}\| \le \delta\}.$$

LEMMA 3.11 (cf. [15, Rem. 3.12]). *The worst-case error of any (linear or nonlinear) reconstruction method $R: \mathcal{Y} \to \mathcal{X}$ on $\mathcal{K}$ satisfies the lower bound*

$$\sup\left\{\left\|f - R(g^{\mathrm{obs}}) \,\middle|\, \mathcal{X}\right\| \colon f \in \mathcal{K}, g^{\mathrm{obs}} \in \mathcal{Y}, \left\|F(f) - g^{\mathrm{obs}} \,\middle|\, \mathcal{Y}\right\| \le \delta\right\} \ge \frac{1}{2}\omega(2\delta, \mathcal{K}). \tag{3.8}$$

Proof. Consider $f_1, f_2 \in \mathcal{K}$ such that $\|F(f_1) - F(f_2) \,|\, \mathcal{Y}\| \le 2\delta$. Then $g^{\mathrm{obs}} := \frac{1}{2}(F(f_1) + F(f_2))$ satisfies $\left\|F(f_j) - g^{\mathrm{obs}} \,\middle|\, \mathcal{Y}\right\| \le \delta$. Hence, the left-hand side $\Delta_R(\delta, \mathcal{K})$ of (3.8) fulfills

$$\Delta_R(\delta,\mathcal{K}) \ge \max_{j\in\{1,2\}} \left\|f_j - R(g^{\mathrm{obs}}) \,\middle|\, \mathcal{X}\right\| \ge \frac{1}{2}\sum_{j=1}^{2} \left\|f_j - R(g^{\mathrm{obs}}) \,\middle|\, \mathcal{X}\right\| \ge \frac{1}{2}\|f_1 - f_2 \,|\, \mathcal{X}\|.$$

Taking the supremum over all f_1, f_2 with the given properties shows (3.8). ∎

We will prove the following adaptation of [11, Prop. 4.6], which estimates the decay of the modulus of continuity if the data have a structure that is compatible with the structure of the Besov space. For this purpose let $(\tilde{\phi}_{j,l})_{(j,l)\in\tilde{I}} \subset L^2(\Omega)$ for

$\tilde{I} = \{(j,l) : j \in \mathbb{N}_0, l \in \tilde{I}_j\}$ with some countable index sets $(\tilde{I}_j)_{j\in\mathbb{N}_0}$ be an orthonormal system, which might be different from $(\phi_{j,l})_{(j,l)\in I}$. Further assume that this system defines an equivalent norm on Besov spaces by

$$\begin{aligned} \|f \,|\, B^s_{p,q}\| &:= \left[\sum_{j\in\mathbb{N}_0} 2^{jsq} \left\| \sum_{l\in I_j} \tilde{\lambda}_{j,l}(f)\tilde{\phi}_{j,l} \,\middle|\, L^p(\Omega)\right\|^q\right]^{\frac{1}{q}} \\ \text{with} \qquad \tilde{\lambda}_{j,l}(f) &:= \int_\Omega f(x)\tilde{\phi}_{j,l}(x)\,\mathrm{d}x \end{aligned} \tag{3.9}$$

for all $p \in (1,\infty)$, $q \in [1,\infty]$ and all $|s| \le \tilde{\sigma}$.

Proposition 3.12. *Let $1 < p < \infty$, $1 \le q, \tilde{q} \le \infty$, $s > 0$. In the setting of Lemma 3.11 set $\mathcal{X} := B^0_{p,q}$, $\mathcal{K} := \{f \in B^s_{p,\tilde{q}} \colon \|f \,|\, B^s_{p,\tilde{q}}\| \le \varrho\}$, $\mathcal{Y} := L^2$. Assume that the following holds true:*

i. $F(0) = 0$.
ii. *There exists $(b_{j,l})_{(j,l)\in\tilde{I}} \subset (0,\infty)$ such that*

$$\|F(f_1) - F(f_2) \,|\, L^2\|^2 \le \sum_{(j,l)\in\tilde{I}} b_{j,l}\left|\tilde{\lambda}_{j,l}(f_1) - \tilde{\lambda}_{j,l}(f_2)\right|^2.$$

iii. *There exists $(f_j)_{j\in\mathbb{N}_0} \subset \mathcal{X}$ such that $\tilde{\lambda}_{k,l}(f_j) = 0$ for all $k \neq j$ and for two constants $c_1, c_2 > 0$ the estimates*

$$\frac{1}{c_1} \le \left\|f_j \,\middle|\, L^2\right\| \le c_1 \quad \text{and} \quad \frac{1}{c_2} \le \|f_j \,|\, L^p\| \le c_2$$

hold true for all $j \in \mathbb{N}_0$.

Then the modulus of continuity satisfies

$$\omega(\delta,\mathcal{K}) \ge \frac{1}{c_2}\sup_{j\in\mathbb{N}_0}\left\{\min\left\{\frac{1}{c_1}\left(\max_{l\in\tilde{I}_j} b_{j,l}\right)^{-\frac{1}{2}}\delta, \frac{1}{c_2}2^{-js}\varrho\right\}\right\}.$$

Proof. For $j \in \mathbb{N}_0$ set

$$w_j := \min\left\{\frac{1}{c_1}\left(\max_{l\in\tilde{I}_j} b_{j,l}\right)^{-\frac{1}{2}}\delta, \frac{1}{c_2}2^{-js}\varrho\right\}.$$

Straightforward computations show that

$$\left\|F(w_jf_j) - F(0) \,\middle|\, L^2\right\| \le \delta, \quad \left\|w_jf_j \,\middle|\, B^s_{p,\tilde{q}}\right\| \le \varrho, \quad \text{and} \quad \left\|w_jf_j \,\middle|\, B^0_{p,q}\right\| \ge \frac{w_j}{c_2}.$$

In particular, $w_jf_j \in \mathcal{K}$ for all $j \in \mathbb{N}_0$ and obviously also $0 \in \mathcal{K}$. This yields

$$\omega(\delta,\mathcal{K}) \ge \sup_{j\in\mathbb{N}_0}\left\|w_jf_j \,\middle|\, B^0_{p,q}\right\| \ge \frac{1}{c_2}\sup_{j\in\mathbb{N}_0}\left\{\min\left\{\frac{1}{c_1}\left(\max_{l\in\tilde{I}_j} b_{j,l}\right)^{-\frac{1}{2}}\delta, \frac{1}{c_2}2^{-js}\varrho\right\}\right\}. \qquad \square$$

Note that the norm defined by the Fourier expansion in (3.5) is of the form (3.9), while setting $(\tilde{\phi}_{j,l})_{(j,l)\in\tilde{I}} := (\phi_{j,l})_{(j,l)\in I}$ with $(\phi_{j,l})_{(j,l)\in I}$ as in (3.4) leads to an equivalent norm since there exists a constant $c > 0$ independent of j such that

$$\frac{1}{c} 2^{jd(\frac{1}{2}-\frac{1}{p})} \|\lambda_{j,\cdot} \,|\, \ell^p(I_j)\| \leq \left\| \sum_{l \in I_j} \lambda_{j,l} \phi_{j,l} \,\middle|\, L^p(\Omega) \right\| \leq c 2^{jd(\frac{1}{2}-\frac{1}{p})} \|\lambda_{j,\cdot} \,|\, \ell^p(I_j)\|; \tag{3.10}$$

see [7, Thm. 3.9.2].

Example 3.13. Proposition 3.12(iii) holds true for both types of Besov norms considered in this paper:

i. Assume that there is a constant $c > 0$ such that

$$c 2^{jd} \leq |I_j|$$

(which is fulfilled for the cases presented in Example 3.5). Then, choosing $\Gamma_j \subset I_j$ with $|\Gamma_j| \sim 2^{jd}$, we get by (3.10) that for $f_j := \sum_{l \in \Gamma_j} 2^{-j\frac{d}{2}} \phi_{j,l}$ Proposition 3.12(iii) holds true for some constants $c_1, c_2 > 0$.

ii. In the Fourier setting (3.5) we get $\|e_z \,|\, L^p\| = 1$ for all $z \in \mathbb{Z}^d$ and $p \in [1, \infty]$. Hence, we can set $f_j = e_l$ for some $l \in I_j$ in order to get that Proposition 3.12(iii) is fulfilled with $c_1 = c_2 = 1$.

4. Finitely smoothing operators. In this section we assume that the forward operator F_a is a-times smoothing for some $a > 0$. More precisely, we assume that

$$p \in (1, 2],\ q \in (1, \infty),\ a > \frac{d}{p} - \frac{d}{2}, \tag{4.1a}$$

$$\big\|F_a(f_1) - F_a(f_2) \,\big|\, L^2\big\| \leq L \big\|f_1 - f_2 \,\big|\, B^{-a}_{2,2}\big\|, \tag{4.1b}$$

$$\big\|f_1 - f_2 \,\big|\, B^{-a}_{2,2}\big\| \leq L \big\|F_a(f_1) - F_a(f_2) \,\big|\, L^2\big\|, \tag{4.1c}$$

$$\big\|F_a(f_1) - F_a(f_2) \,\big|\, B^{a}_{p,q}\big\| \leq L \big\|f_1 - f_2 \,\big|\, B^{0}_{p,q}\big\|, \tag{4.1d}$$

for some $L > 0$ and all $f_1, f_2 \in \mathcal{D} \subset B^0_{p,q}$.

Example 4.1. Equations (4.1) are, e.g., satisfied for the following examples: Set $F_a = (I - \Delta)^{-a/2}$ (or more generally, let F_a be an injective elliptic pseudodifferential operator of order $-a$) on $\mathbb{T}^d$; then $F_a : B^s_{p,q}(\mathbb{T}^d) \to B^{s+a}_{p,q}(\mathbb{T}^d)$ is bounded and boundedly invertible for all $s \in \mathbb{R}$. In the case of bounded domains, boundary conditions have to be taken into account and incorporated in the Besov spaces. For an extensive discussion of the mapping properties of elliptic differential operators in Besov spaces, we refer the reader to [37, Chap. 5].

Example 4.2. In [25, Lemma 2.9] it was shown that (4.1b) and (4.1c) follow from the same equations with F_a replaced by $F'_a[f^\dagger]$ and the *range invariance condition* on F_a. The latter condition is typically only satisfied on bounded subsets of $\mathcal{X}$. In particular, this covers the following example of the identification of a reaction coefficient f. Here upper and lower bounds on the unknown reaction coefficient are often given by physical or chemical a priori knowledge.

Let Ω be a bounded Lipschitz domain in $\mathbb{R}^d$ for $d \in \{1, 2, 3\}$ and $h_1 \in C^\infty(\Omega)$ and $h_2 \in C^\infty(\partial\Omega)$ be positive with $h_2(x) \geq \kappa > 0$ for all $x \in \partial\Omega$. Let $0 \leq \underline{f} < \overline{f} < \infty$. For $f \in \mathcal{D} := \{f \in L^\infty : \underline{f} \leq f \leq \overline{f}, \operatorname{supp}(f) \subset \Omega\}$ define $F(f) = u$, where u solves

$$\begin{aligned} \big(-\Delta + f\big) u &= h_1 && \text{in } \Omega, \\ u &= h_2 && \text{on } \partial\Omega. \end{aligned}$$

Note that by the maximum principle, the assumptions on the right-hand sides h_1 and

h_2, and $f \geq 0$, we have $u \geq \kappa$. Therefore, as shown in [21, Ex. 4.2], F satisfies the range invariance condition in $\mathcal{D}$ if $\overline{f} - \underline{f}$ is sufficiently small. We obtain (4.1b) and (4.1c) for $a = 2$. The verification of (4.1d), which is only needed for random noise, is outside the scope of this paper.

Due to the assumptions (4.1a) and (4.1b) and the continuous embedding $B^0_{p,q} \hookrightarrow B^{-a}_{p,q}$ (see (A.2)), $F_a\colon B^0_{p,q}(\mathbb{T}^d) \to L^2(\mathbb{T}^d)$ is well-defined and continuous. This allows us to choose

$$\mathcal{X} = B^0_{p,q}(\mathbb{T}^d), \qquad \mathcal{R}(\cdot) = \frac{1}{r}\|\cdot \,|\, B^0_{p,q}\|^r_{\mathcal{W}}, \qquad \text{and} \qquad \mathcal{Y} = B^0_{2,2}(\mathbb{T}^d) = L^2(\mathbb{T}^d), \tag{4.2}$$

with $r = \max\{2, p, q\}$ the modulus of convexity (see [29]) for arbitrary $p \in (1, 2]$ and $q \in (1, \infty)$. The penalty term in the Tikhonov functional given by the Besov norm will be expressed via wavelet coefficients as defined in (3.4); hence, most constants will depend implicitly on the specific choice of the wavelet system, which we will not mention further.

4.1. Deterministic convergence rates.

We will first derive convergence rates for the deterministic error model (1.1). We use the strategy of Theorem 3.3 to obtain a variational source condition first and then apply Proposition 2.3(ii).

THEOREM 4.3. *Assume* (1.1), (4.1a)–(4.1c) *and* (4.2), *and suppose that* $f^\dagger \in B^s_{p,\infty}$ *for some* $s \in (0, \frac{a}{q-1})$ *with* $\|f^\dagger \,|\, B^s_{p,\infty}\| \leq \varrho$. *Then there exists a constant* $c > 0$ *such that a variational source condition with*

$$\psi(\tau) = c\varrho^\nu \tau^\mu \quad \textit{where} \quad \nu = \begin{cases} \frac{qa}{a+s}, & q \geq 2, \\ \frac{2a}{a+s(q-1)}, & q \leq 2, \end{cases} \quad \textit{and} \quad \mu = \begin{cases} \frac{q}{2}\frac{s}{a+s}, & q \geq 2, \\ \frac{s(q-1)}{a+s(q-1)}, & q \leq 2, \end{cases}$$

holds true. Moreover, the Tikhonov functional in (1.4) *with* $F = F_a$ *has a minimizer* $\widehat{f}_\alpha$, *and* $\widehat{f}_\alpha$ *is unique if* F_a *is linear. If* α *is chosen by* (2.6a), *then every minimizer* $\widehat{f}_\alpha$ *satisfies the error bound*

$$\left\|\widehat{f}_\alpha - f^\dagger \,\middle|\, B^0_{p,q}\right\| \leq \begin{cases} c\varrho^{\frac{a}{a+s}} \delta^{\frac{s}{a+s}}, & q \geq 2, \\ c\varrho^{\frac{a}{a+s(q-1)}} \delta^{\frac{s(q-1)}{a+s(q-1)}}, & q \leq 2, \end{cases} \tag{4.3}$$

with a constant c *independent of* $f^\dagger$, $\widehat{f}_\alpha$, ϱ, *and* δ.

Proof. We apply Theorem 3.3 with the choice P_j as in (3.7a). Then we see by Corollary 3.9 and our assumptions that we can choose $\kappa(j) = c\varrho^{r-1}2^{-js(q-1)}$.

To verify (3.2c) denote by S_a the operator $S_a := \mathcal{W}^*\tilde{S}_a\mathcal{W}f$, where $(\tilde{S}_a\lambda)_{j,l} = 2^{ja}\lambda_{j,l}$ for all $(j,l) \in I$. Using the relation to Lebesgue spaces (A.3) and (4.1c), we obtain the estimate

$$\begin{aligned} \left\langle P_j f^*, f^\dagger - f\right\rangle &= \left\langle S_a P_j f^*, S_{-a}(f^\dagger - f)\right\rangle \\ &\leq c\left\|S_a P_j f^* \,\middle|\, L^{p'}\right\| \left\|S_{-a}(f^\dagger - f) \,|\, L^2\right\| \left\|1 \,\middle|\, L^{\frac{2p}{2-p}}(\mathbb{T}^d)\right\| \\ &\leq c\left\|S_a P_j f^* \,|\, B^0_{p',2}\right\| \left\|f^\dagger - f \,|\, B^{-a}_{2,2}\right\| \\ &\leq cL\left\|P_j f^* \,|\, B^a_{p',2}\right\| \left\|F_a(f^\dagger) - F_a(f) \,|\, L^2\right\|. \end{aligned} \tag{4.4}$$

By Corollary 3.9 we can hence choose $\gamma = 0$ and

$$\sigma(j) = c\varrho^{r-1}2^{j(a-s(q-1))}$$

in (3.2c) with $c > 0$ depending on the wavelet system and the parameters s, p, q, a.

Now Theorem 3.3 implies that a variational source condition holds true with

$$\psi_{\mathrm{vsc}}(\tau) = \inf_{j\in\mathbb{N}_0} c\Big[\varrho^{r-1}2^{j(a-s(q-1))}\sqrt{\tau} + \varrho^r 2^{-js(q-1)r'}\Big].$$

Choosing j such that $2^j \sim (\varrho/\sqrt{t})^\tau$ with $\tau = \frac{1}{s(q-1)(r'-1)+a}$ we can estimate

$$\psi_{\mathrm{vsc}}(\tau) \le c\varrho^{r-\frac{s(q-1)r'}{a+s(q-1)(r'-1)}}\tau^{\frac{1}{2}\frac{s(q-1)r'}{s(q-1)(r'-1)+a}}.$$

Now use that for $q \le 2$ we have $r = r' = 2$ and that for $q \ge 2$ we have $r = q$ and $r' = q'$.

The existence of $\widehat{f}_\alpha$ follows from standard results (see, e.g., [34, Thm. 3.22]) using the compactness of the embedding $B^0_{p,q} \hookrightarrow B^{-a}_{2,2}$ (see (A.2)) as well as (4.1a) and (4.1b). Uniqueness of $\widehat{f}_\alpha$ for linear operators is obvious by strict convexity. From Proposition 2.3 we obtain $\Delta_{\mathcal{R}}(\widehat{f}_\alpha, f^\dagger) \le \psi_{\mathrm{vsc}}(\delta^2)$, and via Example 2.1(ii) (with $r = \max(2, q)$ as discussed after (4.2)) this yields the convergence rate (4.3). ∎

In practice the parameters s and ϱ describing the smoothness of $f^\dagger$ are usually unknown, of course, and hence the a priori rule (2.6a) is not implementable. Therefore, a posteriori rules such as the discrepancy principle are used, under which the same error bounds can be shown without prior knowledge of s and ϱ (see, e.g., [23]).

4.2. Extensions. In this subsection we discuss extensions of the results of the previous subsection resulting from different penalty terms and data-fidelity terms, respectively.

THEOREM 4.4. *Let the assumptions of Theorem* 4.3 *hold true, but in* (4.2) *set* $\mathcal{X} = B^{\tilde{s}}_{p,q}$ *and* $\mathcal{R}(\cdot) = \frac{1}{r}\|\cdot \,|\, B^{\tilde{s}}_{p,q}\|^r_{\mathcal{W}}$ *for* $\tilde{s} \in \mathbb{R}$. *Further, replace the last inequality in* (4.1a) *by* $a^* := a + \tilde{s} > \frac{d}{p} - \frac{d}{2}$. *Let* $f^\dagger \in B^s_{p,\infty}$ *for some* $s \in \mathbb{R}$ *such that* $s^* := s - \tilde{s} \in (0, \frac{a^*}{q-1})$. *Then the Tikhonov minimizer* $\widehat{f}_\alpha$ *in* (1.4) *exists, and for* α *chosen by* (2.6a) *and* $q \ge 2$ *it satisfies*

$$\Big\|\widehat{f}_\alpha - f^\dagger \,\Big|\, B^{\tilde{s}}_{p,q}\Big\| \le c\varrho^{\frac{a^*}{a^*+s^*}}\delta^{\frac{s^*}{a^*+s^*}}.$$

Proof. The proof is analogous to the proof of Theorem 4.3. Here we obtain $\kappa(j) = c\varrho^{r-1}2^{-js^*(q-1)}$ and $\sigma(j) = c\varrho^{r-1}2^{j(a^*-s^*(q-1))}$. ∎

Remark 4.5. Suppose the constraint $f^\dagger \in \mathcal{D}$ is incorporated in the penalty term $\mathcal{R}$ by replacing it by $\widetilde{\mathcal{R}}(f) := \mathcal{R}(f) + \chi_{\mathcal{D}}(f)$, where $\chi_{\mathcal{D}}(f) := 0$ if $f \in \mathcal{D}$ and $\chi_{\mathcal{D}}(f) := \infty$ otherwise. Then $\partial\widetilde{\mathcal{R}}(f^\dagger) = \partial\mathcal{R}(f^\dagger) + \partial\chi(f^\dagger)$ by the sum rule. $\partial\chi(f^\dagger)$ coincides with the normal cone at $f^\dagger$ and differs from $\{0\}$ if $f^\dagger$ belongs to the boundary of $\mathcal{D}$. In this case $\partial\widetilde{\mathcal{R}}(f^\dagger)$ may contain elements of higher smoothness than $\partial\mathcal{R}(f^\dagger)$, leading to faster rates of convergence (see [17] and [15, sect. 5.4]).

THEOREM 4.6. *The error bound* (4.3) *in Theorem* 4.3 *remains true if we replace* $\mathcal{Y} = L^2$ *by* $\mathcal{Y} = \mathcal{X} = B^0_{p,q}$ *for* $p \in (1, 2]$ *and* $q \in (1, \infty)$, *if we replace the Tikhonov function* (1.4) *by* (1.3) *with arbitrary* $t \ge 1$, *and if we replace Assumption* (4.1) *by*

$$\frac{1}{L}\big\|f_1 - f_2 \,\big|\, B^{-a}_{p,q}\big\| \le \big\|F_a(f_1) - F_a(f_2) \,\big|\, B^0_{p,q}\big\| \le L\big\|f_1 - f_2 \,\big|\, B^{-a}_{p,q}\big\|.$$

Proof. In the proof of Theorem 4.3 we have to adapt the estimate (4.4) in the following way:

$$\big\langle P_j f^*, f^\dagger - f\big\rangle \le c\big\|P_j f^* \,\big|\, B^a_{p',q'}\big\|\big\|f^\dagger - f \,\big|\, B^{-a}_{p,q}\big\|.$$

The norm $\|P_j f^* \,|\, B^a_{p',q'}\|$ can be estimated as before. The exponents μ in the source condition are $\mu = \frac{q}{t}\frac{s}{a+s}$ for $q \geq 2$ and $\mu = \frac{2}{t}\frac{s(q-1)}{a+s(q-1)}$ for $q \in (1,2]$ in this case. ∎

Corollary 4.7. *Under the assumptions of Theorem* 4.3 *with* $q = 2$*, choose* $j := \lfloor -\frac{1}{2a}\ln_2 \alpha \rfloor$. *Then*

$$\left\| P_j \widehat{f}_\alpha - f^\dagger \,\middle|\, L^p \right\| \leq c \left\| P_j \widehat{f}_\alpha - f^\dagger \,\middle|\, B^0_{p,p} \right\| \leq c \varrho^{\frac{a}{s+a}} \delta^{\frac{s}{s+a}} \left(\ln \delta^{-1}\right)^{\frac{1}{p}-\frac{1}{2}}. \tag{4.5}$$

Proof. Setting $h = \widehat{f}_\alpha - f^\dagger$ we get from Hölder's inequality that

$$\|P_j h \,|\, B^0_{p,p}\| = \left(\sum_{j=0}^{j_*} 1 \cdot \|h_j \,|\, L^p\|^p \right)^{1/p} \leq j^{1/p-1/2} \|P_j h \,|\, B^0_{p,2}\|.$$

As $\alpha \sim (\delta/\varrho)^{2a/(s+a)}$ by (2.6a) we have $\|(I - P_j) f^\dagger \,|\, B^0_{p,p}\| \leq c 2^{-js} \|f^\dagger \,|\, B^s_{p,\infty}\| \leq c\varrho\sqrt{\alpha}^{s/a} = c\varrho^{a/(s+a)}\delta^{s/(s+a)}$. Combining both inequalities yields the assertion. ∎

Remark 4.8. In the setting of Theorem 4.6 the projection P_j in (4.5) can be omitted. This follows after some computations comparing the value of the Tikhonov functional for $\widehat{f}_\alpha$ and $f^*_\alpha := P_j \widehat{f}_\alpha + (I - P_j) f^\dagger$.

Remark 4.9. The basic idea of Theorem 4.3 can be generalized as follows: Assume that there exist Banach spaces $\mathcal{X}_{\mathrm{Lip}}$ and $\mathcal{X}_{\mathrm{s}}$ such that the embeddings $\mathcal{X} \hookrightarrow \mathcal{X}_{\mathrm{s}} \hookrightarrow \mathcal{X}_{\mathrm{Lip}}$ are continuous. Let $(\mathcal{X}'_j)_{j\in\mathbb{N}}$ be a sequence of finite dimensional subspaces such that

$$\overline{\bigcup_{j\in\mathbb{N}} \mathcal{X}'_j}^{\|\cdot\,|\,\mathcal{Z}\|} = \mathcal{Z} \qquad \text{for } \mathcal{Z} \in \{\mathcal{X}', \mathcal{X}'_{\mathrm{s}}, \mathcal{X}'_{\mathrm{Lip}}\},$$

and let P_j be a projection onto $\mathcal{X}'_j$. Assume that $f^\dagger$ is such that $f^* \in \mathcal{X}'_{\mathrm{s}}$ and that for all elements $h \in \mathcal{X}'_{\mathrm{s}}$ the generalized Bernstein and Jackson inequalities

$$\begin{aligned} \|P_j h \,|\, \mathcal{X}'_{\mathrm{Lip}}\| &\leq \tilde{\sigma}(j)\|h \,|\, \mathcal{X}'_{\mathrm{s}}\| \\ \text{and} \qquad \|(I - P_j) h \,|\, \mathcal{X}'\| &\leq \tilde{\kappa}(j)\|h \,|\, \mathcal{X}'_{\mathrm{s}}\| \end{aligned}$$

hold true. If the operator F fulfills the Lipschitz estimate

$$\|f_1 - f_2 \,|\, \mathcal{X}_{\mathrm{Lip}}\| \leq L\|F(f_1) - F(f_2) \,|\, \mathcal{Y}\| \qquad \text{for all } f_1, f_2 \in \mathcal{X},$$

then (3.2) is fulfilled with $\kappa(j) = \tilde{\kappa}(j)\|f^* \,|\, \mathcal{X}'_{\mathrm{s}}\|$, $\sigma(j) = L\tilde{\sigma}(j)\|f^* \,|\, \mathcal{X}'_{\mathrm{s}}\|$, and $\gamma = 0$.

4.3. Statistical convergence rates. As a variational source condition is fulfilled and the regularization functional fulfills Assumption 3.1, we only need to show that the operator also fulfills Assumption 2.5. We then obtain convergence rates via Theorems 2.6 and 4.3.

Lemma 4.10. *Suppose that* $a > d/2$ *and* (4.1d) *hold true. Then the operator* F_a *fulfills Assumption* 2.5 *and* (3.1) *with* $\beta = 1 - \frac{d}{2a}$ *and* $\gamma = \frac{d}{2ar}$. *Moreover, there exists* $c > 0$ *such that*

$$\left\| g \,\middle|\, B^{d/2}_{p,1} \right\| \leq c \left\| g \,\middle|\, L^2 \right\|^{1-\frac{d}{2a}} \left\| g \,\middle|\, B^a_{p,q} \right\|^{\frac{d}{2a}} \qquad \textit{for all } g \in B^a_{p,q}. \tag{4.6}$$

Proof. By K-interpolation theory (see [39, sect. 2.4.2], [37, sect. 1.3]) there exists a constant $c > 0$ such that

$$\left\| g \,\middle|\, B^{d/2}_{p,1} \right\| \leq c \left\| g \,\middle|\, B^0_{p,2} \right\|^{1-\frac{d}{2a}} \left\| g \,\middle|\, B^a_{p,q} \right\|^{\frac{d}{2a}},$$

and since $\|\cdot \,|\, B^0_{p,2}(\mathbb{T}^d)\| \leq \|\cdot \,|\, L^2(\mathbb{T}^d)\|$ for $p \leq 2$, this implies (4.6). Using (4.1d) and Assumption 3.1 we obtain

$$\left\| F_a(f_1) - F_a(f_2) \,\middle|\, B^a_{p,q} \right\| \leq L \left\| f_1 - f_2 \,\middle|\, B^0_{p,q} \right\| \leq L \big(C_\Delta^{-1} \Delta_{\mathcal{R}}(f_1, f_2) \big)^{\frac{1}{r}}.$$

This together with (4.6) for $g = F_a(f_1) - F_a(f_2)$ yields Assumption 2.5 and (3.1). ◻

Lemma 4.10 can be used to derive not only convergence rates but also the existence of a minimizer.

PROPOSITION 4.11. *Suppose that $a > d/2$ and F_a satisfies* (4.1a)–(4.1b). *Then for the noise model* (1.2) *with Z satisfying Assumption* 2.4, *the Tikhonov functional in* (1.4) *with $F = F_a$ has a global minimizer $\widehat{f}_\alpha$ almost surely.*

Proof. Note that the data fidelity term in (1.4) is not bounded from below in general, and therefore standard results in the literature such as [34, Thm. 3.22] do not apply. However, with the help of Lemma 4.10 we can show coercivity of the entire Tikhonov functional in $B^0_{p,q}$ if $N := \|g^{\mathrm{obs}} \,|\, B^{-d/2}_{p',\infty}\| < \infty$ (which is true with probability 1 by Assumption 2.4). To this end we bound the mixed term as follows using (4.6) and Hölder's inequality $xy \leq c x^{4a/(2a+d)} + \frac{1}{2} y^{4a/(2a-d)}$:

$$\begin{aligned}
\langle g^{\mathrm{obs}}, F(f) \rangle &\leq N \left\| F(f) \,\middle|\, B^{d/2}_{p,1} \right\| \leq CN \left\| F(f) \,\middle|\, B^a_{p,q} \right\|^{d/2a} \left\| F(f) \,\middle|\, L^2 \right\|^{1-d/2a} \\
&\leq cN^{\frac{4a}{2a+d}} \left(\left\| F(f) - F(0) \,\middle|\, B^a_{p,q} \right\| + \left\| F(0) \,\middle|\, B^a_{p,q} \right\| \right)^{2\frac{d}{2a+d}} + \tfrac{1}{2} \left\| F(f) \,\middle|\, L^2 \right\|^2 \\
&\leq cN^{\frac{4a}{2a+d}} \left\| f \,\middle|\, B^0_{p,q} \right\|^{2\frac{d}{2a+d}} + A + \tfrac{1}{2} \left\| F(f) \,\middle|\, L^2 \right\|^2,
\end{aligned}$$

where $A := cN^{4a/(2a+d)} \|F(0) \,|\, B^{d/2}_{p,1}\|^{2d/(2a+d)}$ and c is a generic constant. Plugging this into the Tikhonov functional yields

$$\begin{aligned}
&\tfrac{1}{2} \left\| F(f) \,\middle|\, L^2 \right\|^2 - \langle g^{\mathrm{obs}}, F(f) \rangle + \alpha \left\| f \,\middle|\, B^0_{p,q} \right\|^r \\
&\geq - cN^{\frac{4a}{2a+d}} \left\| f \,\middle|\, B^0_{p,q} \right\|^{2\frac{d}{2a+d}} + \alpha \left\| f \,\middle|\, B^0_{p,q} \right\|^r + A,
\end{aligned}$$

and as $r \geq 2$ the right-hand side tends to ∞ as $\|f \,|\, B^0_{p,q}\| \to \infty$. This shows that a minimizing sequence (f_n) of the Tikhonov functional must be bounded in $B^0_{p,q}$. As $B^0_{p,q}$ is reflexive, by the Banach–Alaoglu theorem there exists a subsequence f_{n_k} and $f \in B^0_{p,q}$ such that $f_{n_k} \rightharpoonup f$ for $k \to \infty$. Since the embedding $B^0_{p,q} \hookrightarrow B^{-a}_{2,2}$ is compact, we have $\lim_{k\to\infty} \|f_{n_k} - f \,|\, B^{-a}_{2,2}\| = 0$ and by (4.1b) also $\lim_{k\to\infty} \|F(f_{n_k}) - F(f) \,|\, L^2\| = 0$. Now it follows from (3.1) and the boundedness of $\|f_{n_k} \,|\, B^0_{p,q}\|$ that $\|F(f_n) - F(f) \,|\, B^{d/2}_{p,1}\|$ tends to 0 as $k \to \infty$. Together with the weak lower semicontinuity of $\|\cdot \,|\, B^0_{p,q}\|^r$ it follows that f minimizes the Tikhonov functional. ◻

Together with Theorem 4.3 we find the following.

THEOREM 4.12. *Assume* (1.2) *with Z satisfying Assumption* 2.4, (4.1), *and* (4.2) *with $a > d/2$ and $q \geq 2$. Moreover, suppose that $f^\dagger \in B^s_{p,\infty}$ for some $s \in (0, \frac{a}{q-1})$*

with $\|f^\dagger \,|\, B^s_{p,\infty}\| \leq \varrho$. *Then* $\widehat{f}_\alpha$ *(as in Proposition* 4.11*) for an optimal choice of* α *as specified in the proof satisfies the error bound*

$$\mathbb{P}\left[\left\|\widehat{f}_\alpha - f^\dagger \,\middle|\, B^0_{p,q}\right\| > (c+t)\,\varrho^{\frac{a+d/2}{a+s+d/2}}\varepsilon^{\frac{s}{a+s+d/2}}\right] \leq \exp\Big(-C_Z t^{\mu\left(\frac{q}{2}+\frac{(q-2)d}{4a}\right)}\Big)$$

for all $t > 0$. *In particular, for any* $\sigma \geq 1$ *we have*

$$\mathbb{E}\left(\left\|\widehat{f}_\alpha - f^\dagger \,\middle|\, B^0_{p,q}\right\|^\sigma\right)^{1/\sigma} \leq C\varrho^{\frac{a+d/2}{a+s+d/2}}\varepsilon^{\frac{s}{a+s+d/2}}.$$

Here c *and* C *are positive constants independent of* $f^\dagger$, $\widehat{f}_\alpha$, ε, *and* ϱ.

Proof. As $q \geq 2$, by Theorem 4.3 a variational source condition with $\psi(\tau) = c\varrho^{\frac{2a}{a+s}}\tau^{\frac{s}{a+s}}$ holds true, and hence we can use Example 2.2, (2.4a) with the calculus rules for Fenchel duals to obtain that

$$\varphi_\psi(\tau) = C\varrho^{\frac{2qa}{2(a+s)-qs}}\tau^{\frac{qs}{2(a+s)-qs}}.$$

Thus, Theorem 2.6(ii) together with Lemma 4.10 gives

$$\Delta_{\mathcal{R}}(\widehat{f}_\alpha, f^\dagger) \leq C\alpha^{-\frac{q+qd/(2a)}{q+d(q-2)/(2a)}}\left\|\varepsilon Z \,\middle|\, B^{-d/2}_{p',\infty}\right\|^{\frac{2q}{q+d(q-2)/(2a)}} + C\varrho^{\frac{2qa}{2(a+s)-qs}}\alpha^{\frac{qs}{2(a+s)-qs}}.$$

By the choice $\alpha \sim \varrho^{\frac{qa+d\frac{q}{2}(1-\frac{2}{q}))}{a+s+d/2}}\varepsilon^{\frac{2(a+s)-qs}{a+s+d/2}}$ and using (2.1), we find

$$\left\|\widehat{f}_\alpha - f^\dagger \,\middle|\, B^0_{p,q}\right\| \leq C\varrho^{\frac{a+d/2}{a+s+d/2}}\varepsilon^{\frac{s}{a+s+d/2}}\left(1 + \left\|Z \,\middle|\, B^{-d/2}_{p',\infty}\right\|^{\frac{2}{q+d(q-2)/(2a)}}\right),$$

so the deviation inequality given by Assumption 2.4 completes the proof. ∎

4.4. Lower bounds.

THEOREM 4.13. *Suppose that* F_a *satisfies* (4.1b). *Then there cannot exist a reconstruction method* $R\colon L^2 \to B^0_{p,q}$ *for the operators* F_a *satisfying the worst-case error bound*

$$\sup\left\{\left\|f - R(g^{\mathrm{obs}}) \,\middle|\, B^0_{p,q}\right\| : \|f \,|\, B^s_{p,\infty}\| \leq \varrho, \left\|F_a(f) - g^{\mathrm{obs}} \,\middle|\, L^2\right\| \leq \delta\right\} = o\big(\varrho^{\frac{a}{s+a}}\delta^{\frac{s}{s+a}}\big).$$

Hence, for $q \geq 2$ *the rate in Theorem* 4.3 *is optimal up to the value of the constant.*

Proof. The assumptions of Proposition 3.12 are fulfilled with $b_{j,l} = c2^{-2ja}$ for some $c > 0$ by (4.1b). Hence, we obtain

$$\omega(\delta, \mathcal{K}) \geq c\max_{k\in\mathbb{N}_0}\left\{\min\left\{2^{-ks}\varrho, 2^{ka}\delta\right\}\right\} \geq c\varrho^{\frac{a}{s+a}}\delta^{\frac{s}{s+a}},$$

where we have chosen $k \in \mathbb{N}_0$ such that the terms are balanced, i.e., $2^k \sim (\frac{\varrho}{\delta})^{1/(s+a)}$. Now the claim follows by Lemma 3.11. ∎

Remark 4.14. The statement of Theorem 4.13 remains valid in the setting of Theorem 4.4 if one replaces a and s by a^* and s^*, respectively.

Lower bounds for the statistical convergence rates can be concluded from results in [12]. Instead of the continuous Gaussian white noise model, they consider an n-dimensional *normal means model*. However, as their results in [12, Thms. 7 and 9] do not depend on the dimension n, one can send n to infinity so that the Le Cam distance of the two models goes to zero (compare [18, sect. 1]) and thus conclude for general estimators $S = S(g^\dagger + \varepsilon W) \in B^{s^*}_{p,q}$.

COROLLARY 4.15. *We have*

$$\inf_{S} \sup_{\|g^\dagger \,|\, B^{s^{**}}_{p,\infty}\| \le \varrho} \mathbb{E}\left(\left\| g^\dagger - S(g^\dagger + \varepsilon W) \,\middle|\, B^{s^*}_{p,q} \right\|\right) \ge c\varrho^{\frac{s^*+d/2}{s^{**}+d/2}} \varepsilon^{\frac{s^{**}-s^*}{s+d/2}}$$

with c depending on s^*, s^{**}, p, q.

Assume additionally on F_a that $F_a : B^s_{p,q} \to B^{s+a}_{p,q}$ is surjective and $\|f \,|\, B^s_{p,\infty}\| \le L\|F_a f \,|\, B^{s+a}_{p,\infty}\|$ for all $f \in B^s_{p,\infty}$. By setting $s^* = a$ and $s^{**} = s + a$, we find for F_a that

$$\inf_{S} \sup_{\|F_a(f^\dagger) \,|\, B^{s+a}_{p,\infty}\| \le \varrho} \mathbb{E}\left(\left\| F_a(f^\dagger) - S(F_a(f^\dagger) + \varepsilon W) \,\middle|\, B^{a}_{p,q} \right\|\right) \ge C\varrho^{\frac{a+d/2}{s+a+d/2}} \varepsilon^{\frac{s}{s+a+d/2}}.$$

Now by (4.1b) we have for reconstruction methods R that

$$\left\| f^\dagger - R(F_a(f^\dagger) + \varepsilon W) \,\middle|\, B^0_{p,q} \right\| \ge \frac{1}{L}\left\| F_a(f^\dagger) - F_a R(F_a(f^\dagger) + \varepsilon W) \,\middle|\, B^a_{p,q} \right\|.$$

Thus, we get a lower bound coinciding with the upper bound in Theorem 4.12 for $q \ge 2$:

$$\inf_{R} \sup_{\|f^\dagger \,|\, B^{s}_{p,\infty}\| \le L\varrho} \mathbb{E}\left(\left\| f^\dagger - R(F_a(f^\dagger) + \varepsilon W) \,\middle|\, B^{0}_{p,q} \right\|\right) \ge C\varrho^{\frac{a+d/2}{s+a+d/2}} \varepsilon^{\frac{s}{s+a+d/2}}.$$

4.5. Numerical validation. We are considering a problem of the type (4.1), where $F_a : B^0_{p,q}(\mathbb{T}) \to L^2(\mathbb{T})$ is given by $F_a := (I - \partial_x^2)^{-1}$; that is, we have $a = 2$ with a deterministic error model. The true solution $f^\dagger$ is given by a continuous, piecewise linear function; therefore, $f^\dagger \in B^s_{p,\infty}$ for $s \le 1 + 1/p$. As for $q = 2$ the obtained convergence rates are of optimal order, we test for different values of p if they are also achieved numerically using the sequential discrepancy principle on the grid $\alpha_j = 2^{-j}$ with parameter $\tau = 2$; see [1] for details.

Numerical computations are carried out in MATLAB. To obtain an efficient implementation of the operator F_a, we use the FFT on a grid with 2^{10} nodes. For the Besov norm we use the wavelet decomposition of the Wavelet toolbox with periodic db7-wavelets. An inverse crime is avoided via generating data on a finer grid and undersampling. In order to obtain the minimizer of the Tikhonov functional, we use the extension of the Chambolle–Pock algorithm to Banach spaces with a constant parameter choice rule (see [24, Thm. 6]), where the iterations are stopped when the current step gets small compared to the first. Note that the steps of this algorithm become especially simple since the considered spaces are 2-convex. The duality mappings are evaluated with the help of Theorem 3.6. For further details see Appendix B.

We tested which convergence rate we observe if we choose $\mathcal{R}(f) = \frac{1}{2}\|f \,|\, B^0_{p,2}\|^2$ for different values of p. The results of this test are shown in Figure 4.1. It can be seen that for the tested values of p, the observed rates coincide quite well with the predicted optimal rates. The staircase behavior of the reconstruction error plots—best visible for $p = 2$—is due to the sequential discrepancy principle; if for different noise levels the relative error is almost constant, then the same regularization parameter α was chosen. For $p = 1.25$ the last three points are omitted in the plot since our code did not produce solutions satisfying the discrepancy principle.

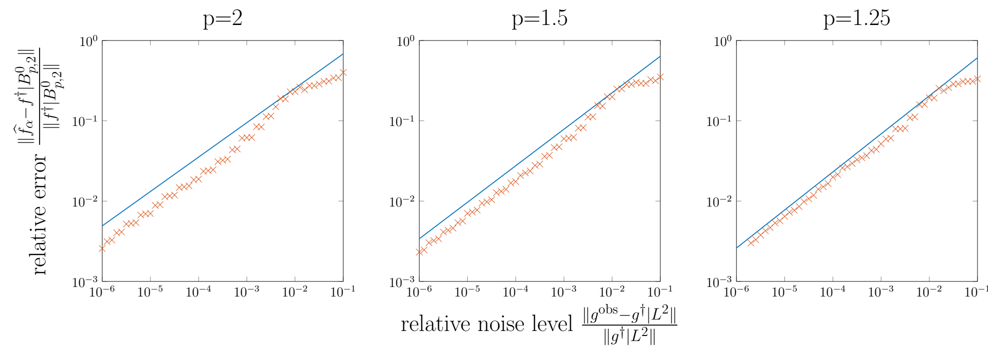

FIG. 4.1. *Convergence rates in $B^0_{p,2}$ for different values of p. The crosses indicate reconstruction errors and the lines convergence rates predicted by Theorem* 4.3.

5. Backwards heat equation.

As a second problem we consider the backwards heat equation on $\mathbb{T}^d$, that is, $(T_{\mathrm{BH}}f)(x) = u(x,\bar{t})$ for some fixed $\bar{t} > 0$, where u solves

$$\begin{aligned} \partial_t u &= \Delta u && \text{in } \mathbb{T}^d \times (0,\bar{t}), \\ u(\cdot,0) &= f && \text{on } \mathbb{T}^d. \end{aligned}$$

Note that T_{BH} can be conveniently expressed via the Fourier series transform $\mathcal{F}$ as

$$T_{\mathrm{BH}} f = \mathcal{F}^* \exp\left(-\bar{t}|\cdot|^2\right)\mathcal{F} f.$$

5.1. Deterministic convergence rates.

THEOREM 5.1. *Let $p \in (1,2]$, $q \in (1,\infty)$, and $s > 0$, and suppose that* (1.1) *and* (4.2) *hold true for $f^\dagger \in B^s_{p,\infty}$ with $\|f^\dagger \,|\, B^s_{p,\infty}\| \le \varrho$. Then $f^\dagger$ satisfies a variational source condition with*

$$\psi(\tau) = c\varrho^r \left[\frac{\sqrt{\tau}}{\varrho}\left(3 + \frac{\varrho}{\sqrt{\tau}}\right)^{1/2} + \left(\ln\left(\left(3 + \frac{\varrho}{\sqrt{\tau}}\right)^{1/2}\right)\right)^{-s(q-1)r'/2} \right].$$

Moreover, $\widehat{f}_\alpha$ in (1.4) *with $F = T_{\mathrm{BH}}$ exists and is unique for any $\alpha > 0$. For the parameter choice rule* (2.6a) *and $q \ge 2$, it satisfies the error bound*

$$\left\| \widehat{f}_\alpha - f^\dagger \,\middle|\, B^0_{p,q} \right\| \le c\varrho\left(\ln\left(\frac{\varrho}{\delta}\right)\right)^{-s/2} \qquad \text{as } \delta \to 0 \tag{5.1}$$

with a constant $c > 0$ independent of $f^\dagger$, δ, and ϱ.

Proof. As in the proof of Theorem 4.3, we apply Theorem 3.3 but this time with the choice P_j as in (3.7b). Again we obtain by Corollary 3.9 and our assumptions that we can choose $\kappa(j) = c\varrho^{r-1}2^{-js(q-1)}$.

In order to verify (3.2c) note that

$$\begin{aligned} \left\|P_j^*(f^\dagger - f) \,\middle|\, L^2\right\|^2 &= \sum_{k=0}^{j}\sum_{l\in I_k}\left|\widehat{(f^\dagger - f)}(l)\right|^2 \le \left(\mathrm{e}^{2^{2j}\bar{t}}\right)^2 \sum_{k=0}^{j}\sum_{l\in I_k}\left|\mathrm{e}^{-|l|^2\bar{t}}\widehat{(f^\dagger - f)}(l)\right|^2 \\ &\le \left(\mathrm{e}^{2^{2j}\bar{t}}\right)^2 \left\|T(f^\dagger - f) \,\middle|\, L^2\right\|^2. \end{aligned}$$

Therefore, we can estimate

$$\begin{aligned}\langle P_j f^*, f^\dagger - f\rangle &\le \left\|f^* \,\middle|\, L^{p'}\right\| \left\|1 \,\middle|\, L^{\frac{2p}{2-p}}\right\| \left\|P_j^*(f^\dagger - f) \,\middle|\, L^2\right\| \\ &\le c\left\|f^* \,\middle|\, B^{s(q-1)}_{p',\infty}\right\| \mathrm{e}^{2^{2j}\bar{t}} \left\|T(f^\dagger - f) \,\middle|\, L^2\right\| \\ &\le c\varrho^{r-1}\mathrm{e}^{2^{2j}\bar{t}} \left\|T(f^\dagger - f) \,\middle|\, L^2\right\|\end{aligned}$$

and hence choose $\sigma(j) = c\varrho^{r-1}\mathrm{e}^{2^{2j}\bar{t}}$ and $\gamma = 0$.

This implies by Theorem 3.3 that a variational source condition with

$$\psi_{\mathrm{vsc}}(\tau) = \inf_{j\in\mathbb{N}_0} c\left[\varrho^{r-1}\mathrm{e}^{2^{2j}\bar{t}}\sqrt{\tau} + \varrho^r 2^{-js(q-1)r'}\right]$$

holds true. Now choosing j such that $2^{2j} \sim \frac{1}{\bar{t}} \ln\sqrt{3 + \frac{\varrho}{\sqrt{\tau}}}$, we obtain that

$$\begin{aligned}\psi_{\mathrm{vsc}}(\tau) &\le c\varrho^r \left[\frac{\sqrt{\tau}}{\varrho}\left(3 + \frac{\varrho}{\sqrt{\tau}}\right)^{1/2} + \left(\ln\left(\left(3 + \frac{\varrho}{\sqrt{\tau}}\right)^{1/2}\right)\right)^{-s(q-1)r'/2}\right] \\ &\le c\varrho^r\left(\ln\left(\frac{\varrho}{\delta}\right)\right)^{-s(q-1)r'/2}[1 + o(1)] \qquad \text{as } \delta \to 0.\end{aligned}$$

Existence of $\widehat{f}_\alpha$ and (5.1) follows along the lines of the proof of Theorem 4.3. ∎

Remark 5.2. Note that the verification of (3.2c) is very different from Theorem 4.3. For the latter the choice of P_j is not important as long as the Bernstein and Jackson inequalities of Assumption 3.7 hold. Here, however, the forward operator F is too smoothing in order to get Lipschitz stability estimates in the scale of Besov spaces. Therefore, we use an inequality of the form $\|P_j(f_1 - f_2)\| \le \sigma_j\|F(f_1) - F(f_2)\|$ for some sequence $\sigma_j > 0$ quite similar to Example 3.4. Such an inequality can in general only be verified for appropriately chosen P_j. We refer to [26] for another example with $\gamma \neq 0$.

THEOREM 5.3. *Let $1 \le p \le 2$, $1 \le q \le \infty$, $s, \varrho > 0$. Then there cannot exist a reconstruction method $R : L^2 \to B^0_{p,q}$ for the backwards heat equation such that*

$$\sup\left\{\left\|f - R(g^{\mathrm{obs}}) \,\middle|\, B^0_{p,q}\right\| : \left\|f \,\middle|\, B^s_{p,\infty}\right\| \le \varrho, \left\|T_{\mathrm{BH}}f - g^{\mathrm{obs}} \,\middle|\, L^2\right\| \le \delta\right\} = o\left(\varrho\left(\ln\frac{\varrho}{\delta}\right)^{-\frac{s}{2}}\right)$$

as $\delta \to 0$. Hence, the convergence rate of Theorem 5.1 *is optimal for $q \ge 2$ as $\delta \to 0$ up to the value of the constant.*

Proof. Choosing $b_{j,l} = \exp(-2\bar{t}|l|^2)$ the assumptions of Proposition 3.12 are fulfilled as shown in Example 3.13(ii). Hence, we obtain

$$\omega(\delta, \mathcal{K}) \ge \max_{k\in\mathbb{N}_0}\left\{\min\left\{2^{-ks}\varrho, \mathrm{e}^{\bar{t}2^{2k}}\delta\right\}\right\} \ge \min\left\{\varrho\left(\frac{1}{\bar{t}}\ln\frac{\varrho}{\delta}\right)^{-\frac{s}{2}}, \varrho\right\},$$

where we have chosen $k \in \mathbb{N}_0$ such that $2^{2k} \sim \frac{1}{\bar{t}}\ln\frac{\varrho}{\delta}$. Hence, the claim follows from Lemma 3.11. ∎

5.2. Statistical convergence rates. For exponentially ill-posed problems a rather coarse interpolation bound is sufficient.

LEMMA 5.4. *The operator* T_{BH} *fulfills Assumption* 2.5 *with* $\beta = \frac{1}{2}, \gamma = \frac{1}{2r}$.

Proof. By K-interpolation there exists a constant $c > 0$ such that

$$\left\|T_{\mathrm{BH}}(f_1 - f_2)\,\middle|\, B^{d/2}_{p,1}\right\| \le c\big\|T_{\mathrm{BH}}(f_1-f_2)\,\big|\, B^0_{p,2}\big\|^{\frac{1}{2}} \big\|T_{\mathrm{BH}}(f_1-f_2)\,\big|\, B^d_{p,q}\big\|^{\frac{1}{2}}.$$

As $p \le 2$, the first factor can be bounded by $\big\|T_{\mathrm{BH}}(f_1-f_2)\,\big|\, L^2\big\|^{\frac{1}{2}}$. To control the second factor we again use $p \le 2$ to obtain

$$\begin{aligned}\big\|T_{\mathrm{BH}}(f_1-f_2)\,\big|\, B^d_{p,q}\big\|^q &\le \sum_{j\in\mathbb{N}_0} 2^{jdq}\left[\sum_{l\in I_j}\exp(-2\bar{t}|l|^2)\left|\widehat{(f_1-f_2)}(l)\right|^2\right]^{\frac{q}{2}}\\ &\le \sum_{j\in\mathbb{N}_0} 2^{jdq}\exp\left(-\frac{\bar{t}q}{16\pi^2}2^{2j}\right)\left[\sum_{l\in I_j}\left|\widehat{(f_1-f_2)}(l)\right|^2\right]^{\frac{q}{2}}.\end{aligned}$$

As there exists a constant $c > 0$ such that

$$2^{jdq}2^{-jd(\frac{1}{2}-\frac{1}{p})}\exp\left(-\tfrac{\bar{t}q}{16\pi^2}2^{2j}\right) \le c \qquad \text{for all } j \in \mathbb{N},$$

one obtains that

$$\big\|T_{\mathrm{BH}}(f_1-f_2)\,\big|\, B^d_{p,q}\big\| \le c\big\|f_1-f_2\,\big|\, B^{\frac{d}{2}-\frac{d}{p}}_{2,q}\big\| \le c\big\|f_1-f_2\,\big|\, B^0_{p,q}\big\| \le c\big(C_\Delta^{-1}\Delta_{\mathcal{R}}(f_1,f_2)\big)^{\frac{1}{r}}$$

by the embedding properties of Besov spaces (A.2) and Assumption 3.1. This completes the proof. ∎

THEOREM 5.5. *Assume that* (1.2) *holds true with* $F = T_{\mathrm{BH}}$ *and* Z *satisfying Assumption* 2.4, *and consider the setting* (4.2) *with* $p \in (1,2]$ *and* $q \in [2,\infty)$. *Moreover, suppose that* $f^\dagger \in B^s_{p,\infty}$ *for some* $s > 0$ *with* $\|f^\dagger \,|\, B^s_{p,\infty}\| \le \varrho$. *Then* $\widehat{f}_\alpha$ *in* (1.4) *is well-defined almost surely, and for* $\alpha = \varepsilon/4$ *it satisfies the error bounds*

$$\forall t > 0:\quad \mathbb{P}\left[\left\|\widehat{f}_\alpha - f^\dagger\,\middle|\, B^0_{p,q}\right\| \ge c\left(\varrho\left(\ln\left(\frac{\varrho}{\varepsilon}\right)\right)^{-s/2} + t\right)\right] \le \exp\left(-C_Z\varepsilon^{-\frac{\mu}{4}}t^{\frac{3q-2}{4}\mu}\right)$$

with a constant $c > 0$ *independent of* $f^\dagger$, ϱ, *and* ε.

Proof. Existence of $\widehat{f}_\alpha$ follows from Proposition 4.11. By Theorem 5.1 a logarithmic variational source condition of the form of Example 2.2 holds true. Hence, from (2.4b) one obtain that

$$\varphi_\psi(\tau) = C\varrho^r\left(\ln\left(\frac{\varrho}{\tau}\right)\right)^{-\frac{s(q-1)r'}{2}}(1+o(1)) \qquad \text{as } t \to 0.$$

Thus, Theorem 2.6(ii) together with Lemma 5.4 gives

$$\Delta_{\mathcal{R}}(\widehat{f}_\alpha, f^\dagger) \le C\alpha^{-\frac{3r}{3r-2}}\left\|\varepsilon Z\,\middle|\, B^{-d/2}_{p',\infty}\right\|^{\frac{4r}{3r-2}} + C\varphi_\psi(4\alpha).$$

Choosing $\alpha = \frac{\varepsilon}{4}$ and using (2.1) we get

$$\left\|\widehat{f}_\alpha - f^\dagger\,\middle|\, B^0_{p,q}\right\| \le C\varepsilon^{\frac{1}{3r-2}}\left\|Z\,\middle|\, B^{-d/2}_{p',\infty}\right\|^{\frac{4}{3r-2}} + C\varphi_\psi(\varepsilon)^{\frac{1}{r}}.$$

Finally, by the deviation inequality given by Assumption 2.4, we find that

$$\mathbb{P}\left[\left\|\widehat{f}_\alpha - f^\dagger \,\middle|\, B^0_{p,q}\right\| \geq C\left(\varrho\left(\ln\left(\frac{\varrho}{\varepsilon}\right)\right)^{-\frac{s(q-1)r'}{2r}} + \varepsilon^{\frac{1}{3r-2}} t'^{\frac{4}{3r-2}}\right)\right] \leq \exp(-C_Z t'^{\mu})$$

for all $t' > 0$. Substituting $t = \varepsilon^{\frac{1}{3r-2}} t'^{\frac{4}{3r-2}}$ shows the claim. ∎

For the optimality of this risk bound in the case $p = q = 2$, we refer the reader to [19].

6. Discussion. We have shown optimal rates of convergence for finitely smoothing operators for Tikhonov regularization with Besov norm penalties $B^0_{p,q}$ with $p \in (1, 2]$ and $q \geq 2$ if the true solution belongs to $B^s_{p,\infty}$. By Proposition 3.12 these rates cannot be improved in the deterministic case if we restrict to smaller spaces $B^s_{\tilde{p},\tilde{q}}$ with $\tilde{p} \geq p$ and $\tilde{q} \in [1, \infty]$. For $p = q = 2$ we have shown in [27] that $B^s_{2,\infty}$ is the largest space in which Tikhonov regularization achieves the rate $O(\delta^{s/(s+a)})$, and we conjecture that this is also the case for $p < 2$. Note that the smaller p is chosen, the larger becomes the smoothness class on which optimal rates are achieved.

For $q < 2$ our approach in its present form yields error bounds which are most likely suboptimal. This case requires further investigation. It would be desirable to get rid of the logarithmic factor in Corollary 4.7 and the following remark for L^p loss functions. Moreover, the case $p = q$ is more commonly used and a bit more convenient algorithmically.

Appendix A. Properties of Besov spaces. In this appendix we collect some properties of the Besov scale $B^s_{p,q}$.

Besov spaces are a quite general class of spaces. In order to provide some intuition on these spaces, we will list some properties and special cases here which can, e.g., be found in [39]. First of all, for $1 < p, q < \infty$ and $s \in \mathbb{R}$, the dual space is given via $(B^s_{p,q})' = B^{-s}_{p',q'}$. The spaces form scales with respect to the smoothness and summation index; for any $\varepsilon > 0$, $s \in \mathbb{R}$, $p \in [1, \infty]$, and $1 \leq q_1 \leq q_2 \leq \infty$, the embeddings

$$B^{s+\varepsilon}_{p,\infty} \subset B^s_{p,1} \subset B^s_{p,q_1} \subset B^s_{p,q_2} \subset B^s_{p,\infty} \subset B^{s-\varepsilon}_{p,1} \tag{A.1}$$

are continuous. Furthermore, one can give up smoothness to gain integrability. To be more precise, for $p_1 \leq p_2$ and $s_1 \geq s_2$ the embedding

$$B^{s_1}_{p_1,q} \subset B^{s_2}_{p_2,q} \tag{A.2}$$

is continuous if $s_1 - \frac{d}{p_1} \geq s_2 - \frac{d}{p_2}$ and compact if $s_1 > s_2$ (see [39, sect. 4.3.3, Rem. 1]).

The classical Lebesgue spaces L^p for $p \neq 2$ are not Besov spaces, but for $1 < p < \infty$ the following inclusions hold true with continuous embeddings:

$$B^0_{p,\min\{2,p\}} \subset L^p \subset B^0_{p,\max\{2,p\}}. \tag{A.3}$$

However, if s is not an integer, then $B^s_{p,p} = W^{s,p}$, the Sobolev spaces with the norm given by (assume for simplicity that $0 < s < 1$)

$$\|f \,|\, W^{s,p}\| = \|f \,|\, L^p\| + \left(\int\int \frac{|f(x) - f(y)|^p}{|x - y|^{d+sp}} \,\mathrm{d}y \,\mathrm{d}x\right)^{\frac{1}{p}}.$$

An important class of solutions to inverse problems are functions f which are smooth up to jumps (or jumps in the kth-derivative). It is well known that such functions are in $W^{d/p-\varepsilon,p}$ (or in $W^{k+d/p-\varepsilon,p}$, respectively) for all $\varepsilon > 0$; however, in this scale of spaces such functions do not have a maximal smoothness index. Using the Nikol'skij representation of the Besov norm $B^s_{p,\infty}$ (see [39, sect. 2.5.12]), one can easily calculate that such functions belong to $B^s_{p,\infty}$ (or to $B^{k+s}_{p,\infty}$, respectively) for $s = d/p$.

For $p, q \in (1, \infty)$, Assumption 3.1 is fulfilled via Example 2.1(ii), as Besov spaces $B^s_{p,q}$ equipped with the wavelet norm are $\max\{2, p, q\}$-convex; see [29].

Appendix B. Details on numerical simulation. The true solution is given by the linear spline interpolating the points

$$\left\{(0,0), \left(\tfrac{1}{8},0\right), \left(\tfrac{3}{16},\tfrac{1}{16}\right), \left(\tfrac{1}{4},0\right), \left(\tfrac{5}{16},0\right), \left(\tfrac{7}{16},\tfrac{1}{8}\right), \left(\tfrac{1}{2},\tfrac{1}{16}\right), \left(\tfrac{5}{8},\tfrac{3}{16}\right), \left(\tfrac{11}{16},\tfrac{1}{16}\right), \left(\tfrac{3}{4},\tfrac{1}{8}\right), \left(\tfrac{7}{8},0\right)\right\}$$

with periodic boundary condition on $\mathbb{T}^1$.

In order to find a noise vector ξ in (1.1) which (at least approximately) maximizes the left-hand side of (4.3), we proceeded as follows: Let F be a compact operator with singular system $(f_j, g_j, \sigma_j)_{j\in\mathbb{N}}$, and denote by f_α the minimizer of the Tikhonov functional (1.3) for noise-free data $g^{\mathrm{obs}} = F(f^\dagger)$. One can expect that there exists $c_1 > 0$ such that

$$\sup_{\xi} \left\| f^\dagger - \widehat{f}_\alpha \,\middle|\, L^2 \right\| \geq c_1 \left\| f_\alpha - \widehat{f}_\alpha \,\middle|\, L^2 \right\|.$$

By first-order optimality conditions $\widehat{f}_\alpha - f_\alpha = (F^*F + \alpha I)^{-1} F^* \xi$, therefore, the right-hand side is of the form

$$\left\| f_\alpha - \widehat{f}_\alpha \,\middle|\, L^2 \right\|^2 = \left\| (F^*F + \alpha I)^{-1} F^* \xi \,\middle|\, L^2 \right\|^2 = \sum_{j\in\mathbb{N}} \left| \frac{\sigma_j \langle \xi, g_j \rangle}{\sigma_j^2 + \alpha} \right|^2 .$$

Since the function $\lambda/(\lambda^2 + \alpha)$ is maximized by $\lambda = \sqrt{\alpha}$, the right-hand side will be close to its maximum if for $0 < c_2 < c_3$ we choose $\xi = \sum_{j\in J} \delta_j g_j$ with $\sum_{j\in J} \delta_j^2 \leq \delta^2$, where J is such that $c_2\sqrt{\alpha} \leq \sigma_j \leq c_3\sqrt{\alpha}$ for all $j \in J$. This leads to the lower bound

$$\left\| f^\dagger - \widehat{f}_\alpha \,\middle|\, L^2 \right\|^2 \geq c_1^2 \frac{c_2^2}{1 + c_3^2} \frac{\delta^2}{\alpha}$$

on the reconstruction error. Recall that $\frac{\delta^2}{\alpha}$ also appears in the upper bound on the reconstruction error in Proposition 2.3; hence, for any parameter choice of α for which the upper bound is of optimal order, the lower bound will be of the same order; i.e., up to constants we will observe the worst-case error rate. Although for $p \neq 2$ the estimator $\widehat{f}_\alpha$ depends in a nonaffine way on ξ, we will assume that the same choice of the noise vector will lead to the worst-case error.

The operator of the tested example $F = (I - \frac{1}{4\pi^2}\partial_x^2)^{-1}$ on $\mathbb{T}^1$ is compact with singular system

$$f_j(x) = g_j(x) = \begin{cases} \exp\left(2\pi\mathrm{i}\frac{j-1}{2}x\right) & j \text{ odd} \\ \exp\left(-2\pi\mathrm{i}\frac{j}{2}x\right) & j \text{ even} \end{cases} \quad \text{and} \quad \sigma_j = \begin{cases} \left(1 + \left(\frac{j-1}{2}\right)^2\right)^{-1} & j \text{ odd} \\ \left(1 + \left(\frac{j}{2}\right)^2\right)^{-1} & j \text{ even} \end{cases}$$

with $j \in \mathbb{N}$. Therefore, we use the approximation $\sigma_j \approx (\frac{j}{2})^{-2}$ and apply the above error model with $c_1 = 1/2$ and $c_2 = 2$. The values for δ_j are drawn from a normal distribution and normalized in order to fulfill $\sum_{j\in J} \delta_j^2 = \delta^2$.

Acknowledgment. We would like to thank two anonymous referees for their advice which helped to improve the paper.

REFERENCES

[1] S. W. Anzengruber, B. Hofmann, and P. Mathé, *Regularization properties of the sequential discrepancy principle for Tikhonov regularization in Banach spaces*, Appl. Anal., 93 (2014), pp. 1382–1400, https://doi.org/10.1080/00036811.2013.833326.

[2] P. Auscher, *Ondelettes à support compact et conditions aux limites*, J. Funct. Anal., 111 (1993), pp. 29–43, https://doi.org/10.1006/jfan.1993.1002.

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

[4] T. Bonesky, K. S. Kazimierski, P. Maass, F. Schöpfer, and T. Schuster, *Minimization of Tikhonov functionals in Banach spaces*, Abstr. Appl. Anal., 2008 (2008), pp. 1–19, https://doi.org/10.1155/2008/192679.

[5] M. Burger and S. Osher, *Convergence rates of convex variational regularization*, Inverse Problems, 20 (2004), pp. 1411–1421.

[6] A. K. Chakrabarty, P. Shunmugaraj, and C. Zălinescu, *Continuity properties for the subdifferential and ϵ-subdifferential of a convex function and its conjugate*, J. Convex Anal., 14 (2007), pp. 479–514.

[7] A. Cohen, *Numerical Analysis of Wavelet Methods*, Stud. Math. Appl. 32, North-Holland, Amsterdam, 2003.

[8] A. Cohen, I. Daubechies, and P. Vial, *Wavelets on the interval and fast wavelet transforms*, Appl. Comput. Harmon. Anal., 1 (1993), pp. 54–81, https://doi.org/10.1006/acha.1993.1005.

[9] A. Cohen, M. Hoffmann, and M. Reiss, *Adaptive wavelet Galerkin methods for linear inverse problems*, SIAM J. Numer. Anal., 42 (2004), pp. 1479–1501, https://doi.org/10.1137/S0036142902411793.

[10] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conf. Ser. Appl. Math. 61, SIAM, Philadelphia, 1992, https://doi.org/10.1137/1.9781611970104.

[11] I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., 57 (2004), pp. 1413–1457, https://doi.org/10.1002/cpa.20042.

[12] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, *Wavelet shrinkage: Asymptopia?*, J. Roy. Statist. Soc. Ser. B, 57 (1995), pp. 301–369.

[13] P. P. B. Eggermont, *Maximum entropy regularization for Fredholm integral equations of the first kind*, SIAM J. Math. Anal., 24 (1993), pp. 1557–1576.

[14] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, North-Holland, Amsterdam, 1976.

[15] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Math. Appl. (Berlin) 375, Kluwer Academic, Dordrecht, 1996, https://doi.org/10.1007/978-94-009-1740-8, 8.

[16] J. Flemming, *Generalized Tikhonov Regularization and Modern Convergence Rate Theory in Banach Spaces*, Shaker, Aachen, Germany, 2012.

[17] J. Flemming and B. Hofmann, *Convergence rates in constrained Tikhonov regularization: Equivalence of projected source conditions and variational inequalities*, Inverse Problems, 27 (2011), 085001, https://doi.org/10.1088/0266-5611/27/8/085001.

[18] E. Giné and R. Nickl, *Mathematical Foundations of Infinite-Dimensional Statistical Models*, Camb. Ser. Stat. Probab. Math., Cambridge University Press, Cambridge, 2016.

[19] G. K. Golubev and R. Z. Khasminski, *Statistical approach to some inverse boundary problems for partial differential equations*, Probl. Inf. Transm., 35 (1999), pp. 51–66.

[20] M. Grasmair, *Generalized Bregman distances and convergence rates for non-convex regularization methods*, Inverse Problems, 26 (2010), 115014, https://doi.org/10.1088/0266-5611/26/11/115014.

[21] M. Hanke, A. Neubauer, and O. Scherzer, *A convergence analysis of the Landweber iteration for nonlinear ill-posed problems*, Numer. Math., 72 (1995), pp. 21–37.

[22] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer, *A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators*, Inverse Problems, 23 (2007), pp. 987–1010, https://doi.org/10.1088/0266-5611/23/3/009.

[23] B. Hofmann and P. Mathé, *Parameter choice in Banach space regularization under variational inequalities*, Inverse Problems, 28 (2012), 104006, 17, https://doi.org/10.1088/0266-5611/28/10/104006.

[24] T. Hohage and C. Homann, *A Generalization of the Chambolle-Pock Algorithm to Banach Spaces with Applications to Inverse Problems*, https://arxiv.org/abs/1412.0126, 2014.

[25] T. Hohage and P. Miller, *Optimal convergence rates for sparsity promoting wavelet-regularization in Besov spaces*, Inverse Problems, (2019), https://doi.org/10.1088/1361-6420/ab0b15.

[26] T. Hohage and F. Weidling, *Verification of a variational source condition for acoustic inverse medium scattering problems*, Inverse Problems, 31 (2015), 075006, https://doi.org/10.1088/0266-5611/31/7/075006.

[27] T. Hohage and F. Weidling, *Characterizations of variational source conditions, converse results, and maxisets of spectral regularization methods*, SIAM J. Numer. Anal., 55 (2017), pp. 598–620, https://doi.org/10.1137/16M1067445.

[28] T. Hohage and F. Werner, *Convergence rates for inverse problems with impulsive noise*, SIAM J. Numer. Anal., 52 (2014), pp. 1203–1221, https://doi.org/10.1137/130932661.

[29] K. S. Kazimierski, *On the smoothness and convexity of Besov spaces*, J. Inverse Ill-Posed Probl., 21 (2013), pp. 411–429, https://doi.org/10.1515/jip-2013-0006.

[30] G. Kerkyacharian, P. Petrushev, D. Picard, and T. Willer, *Needlet algorithms for estimation in inverse problems*, Electron. J. Stat., 1 (2007), pp. 30–76, https://doi.org/10.1214/07-EJS014.

[31] E. Klann, P. Maass, and R. Ramlau, *Two-step regularization methods for linear inverse problems*, J. Inverse Ill-Posed Probl., 14 (2006), pp. 583–609.

[32] D. A. Lorenz and D. Trede, *Optimal convergence rates for Tikhonov regularization in Besov scales*, Inverse Problems, 24 (2008), 055010, 14, https://doi.org/10.1088/0266-5611/24/5/055010.

[33] R. Ramlau and E. Resmerita, *Convergence rates for regularization with sparsity constraints*, Electron. Trans. Numer. Anal., 37 (2010), pp. 87–104.

[34] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational Methods in Imaging*, Appl. Math. Sci. 167, Springer, New York, 2009.

[35] T. Schuster, B. Kaltenbacher, B. Hofmann, and K. Kazimierski, *Regularization Methods in Banach Spaces*, Radon Ser. Comput. Appl. Math., De Gruyter, Berlin, 2012.

[36] B. Sprung and T. Hohage, *Higher Order Convergence Rates for Bregman Iterated Variational Regularization of Inverse Problems*, https://arxiv.org/abs/1710.09244, 2017.

[37] H. Triebel, *Interpolation Theory, Function Spaces, Differential Operators*, North-Holland Mathematical Library 18, North-Holland, Amsterdam, 1978.

[38] H. Triebel, *Function Spaces and Wavelets on Domains*, EMS Tracts Math. 7, European Mathematical Society, Zürich, 2008.

[39] H. Triebel, *Theory of Function Spaces*, Modern Birkhäuser Classics, Springer, Basel, reprint ed., 2010.

[40] G. M. Vainikko, *On the optimality of methods for ill-posed problems*, Z. Anal. Anwend., 6 (1987), pp. 351–362.

[41] M. C. Veraar, *Regularity of Gaussian white noise on the d-dimensional torus*, Banach Center Publ., 95 (2011), pp. 385–398, https://doi.org/10.4064/bc95-0-24.

[42] F. Weidling and T. Hohage, *Variational source conditions and stability estimates for inverse electromagnetic medium scattering problems*, Inverse Probl. Imaging, 11 (2017), pp. 203–220, https://doi.org/10.3934/ipi.2017010.

[43] F. Werner and T. Hohage, *Convergence rates in expectation for Tikhonov-type regularization of inverse problems with Poisson data*, Inverse Problems, 28 (2012), 104004, https://doi.org/10.1088/0266-5611/28/10/104004.