Electronic Transactions on Numerical Analysis.
Volume 57, pp. 101–126, 2022.

ISSN 1068–9613.
DOI: 10.1553/etna_vol57s101

ETNA
Kent State University and
Johann Radon Institute (RICAM)

CONVERGENCE RATES FOR OVERSMOOTHING BANACH SPACE REGULARIZATION*

PHILIP MILLER† AND THORSTEN HOHAGE†

Abstract. This paper studies Tikhonov regularization for finitely smoothing operators in Banach spaces when the penalization enforces too much smoothness in the sense that the penalty term is not finite at the true solution. In a Hilbert space setting, Natterer [Applicable Anal., 18 (1984), pp. 29–37] showed with the help of spectral theory that optimal rates can be achieved in this situation. ("Oversmoothing does not harm.") For oversmoothing variational regularization in Banach spaces, only very recently has progress been achieved in several papers in different settings, all of which construct families of smooth approximations to the true solution. In this paper we propose to construct such a family of smooth approximations based on K-interpolation theory. We demonstrate that this leads to simple, self-contained proofs and to rather general results. In particular, we obtain optimal convergence rates for bounded variation regularization, general Besov penalty terms, and ℓ^p wavelet penalization with $p < 1$, which cannot be treated by previous approaches. We also derive minimax optimal rates for white noise models. Our theoretical results are confirmed in numerical experiments.

Key words. regularization, convergence rates, oversmoothing, BV-regularization, sparsity-promoting wavelet regularization, statistical inverse problems

AMS subject classifications. 65J22, 65N21, 35R20

1. Introduction.

Inverse problems occur in many areas of science and engineering when a quantity of interest f is not directly accessible, and only indirect effects g^{obs} can be observed under noise. Very often such inverse problems are formulated in the form of operator equations,

$$F(f) = g,$$

with some injective, but possibly nonlinear, forward operator $F : D_F \to \mathbb{Y}$ mapping a subset D_F of some Banach space to another Banach space $\mathbb{Y}$. Typically these operator equations are ill-posed in the sense that the inverse of F fails to be continuous with respect to useful Banach norms. Such problems have been studied in numerous papers and monographs; we only refer to [11, 23, 24].

Probably the most common and well-known method to deal with ill-posedness for inexact observed data g^{obs} and to compute stable reconstructions of f is Tikhonov regularization. If g^{obs} belongs to $\mathbb{Y}$ with deterministic error bound

$$\|g^{\mathrm{obs}} - F(f)\|_{\mathbb{Y}} \leq \delta, \tag{1.1}$$

we consider Tikhonov regularization in the form

$$S_\alpha(g^{\mathrm{obs}}) := \underset{h \in D_F \cap \mathbb{X}_{\mathcal{R}}}{\operatorname{argmin}} \left[\frac{1}{2\alpha} \|g^{\mathrm{obs}} - F(h)\|_{\mathbb{Y}}^2 + \frac{1}{u} \|h\|_{\mathbb{X}_{\mathcal{R}}}^u \right], \tag{1.2}$$

with a penalty term $(1/u)\|h\|_{\mathbb{X}_{\mathcal{R}}}^u$ given by the norm of another Banach space $\mathbb{X}_{\mathcal{R}}$, a regularization parameter $\alpha > 0$, and an exponent $u \in (0, \infty)$. Later in Section 5 we will also consider a variant of (1.2) for a white noise model.

Oversmoothing refers to the situation that the true solution f does not belong to the space $\mathbb{X}_{\mathcal{R}}$. This situation is likely to occur if the norm of $\mathbb{X}_{\mathcal{R}}$ contains derivatives. The use of such

*Received March 7, 2022. Accepted March 24, 2022. Published online on June 24, 2022. Recommended by Elena Resmerita. This work has been supported by the Deutsche Forschungsgemeinschaft (German Research Foundation, DFG) through Grant RTG 2088, project B01.
†Institute for Numerical and Applied Mathematics, University of Göttingen, Germany ({p.miller,hohage}@math.uni-goettingen.de).

ETNA
Kent State University and
Johann Radon Institute (RICAM)

penalty terms is common practice and was already proposed in the original paper by Tikhonov [26]. As usual in regularization theory, we aim to bound the reconstruction error in terms of the noise level δ. To give a specific example, we may be interested to bound the L^p-error for image deblurring with bounded variation regularization in the presence of texture.

As the Tikhonov reconstructions in (1.2) belong to $\mathbb{X}_{\mathcal{R}}$, but $f \notin \mathbb{X}_{\mathcal{R}}$, one cannot expect them to converge to f in $\mathbb{X}_{\mathcal{R}}$. Instead, the reconstruction error will be measured in a weaker norm $\|\cdot\|_{\mathbb{X}_{\mathrm{L}}}$ indexed by the subscript L for "loss function." We will further assume that F is finitely smoothing in the sense that it satisfies a two-sided Lipschitz condition with respect to the norm of an even larger space $\mathbb{X}_{-}$ (typically with negative smoothness index), and $\mathbb{X}_{\mathrm{L}}$ contains or coincides with some real interpolation space between $\mathbb{X}_{-}$ and $\mathbb{X}_{\mathcal{R}}$.

Let us now briefly sketch the literature on oversmoothing regularization. In a first seminal paper [21] inspiring numerous follow-up works, Natterer analyzed the case that $\mathbb{Y}$ is a Hilbert space, F is linear, $u = 2$, and $\mathbb{X}_{\mathcal{R}}$, $\mathbb{X}_{\mathrm{L}}$, and $\mathbb{X}_{-}$ all belong to a Hilbert scale, using the Heinz inequality for self-adjoint operators in Hilbert spaces as a main tool. In Banach space settings, only variational techniques are available, and usually a first step is to derive an inequality for the Tikhonov estimator by plugging the true solution f into the Tikhonov functional. In the oversmoothing case, this is not possible, and only recently has progress been achieved for this situation by several constructions of sequences of smooth elements approximating the true solution f. Hofmann and Mathé [15] (see also [16]) consider nonlinear operators, still in Hilbert spaces, but their approach for constructing smooth approximations to f (auxiliary elements in their terminology) is already essentially a special case of our approach. In [13] the case of ℓ^1-regularization with ℓ^2-loss and a diagonal operator was studied using truncation of the sequence f. In a previous work [20] the authors analyzed oversmoothing in sparsity-promoting wavelet regularization using hard thresholding to approximate f by smooth elements. The most general results so far have been obtained by Chen, Hofmann, and Yousept [5], who use functional calculus of sectorial operators to construct smooth approximating sequences.

In this paper we propose to construct a sequence of smooth approximations to f based on K-interpolation theory. We believe that our analysis is significantly simpler than the one in [5]. Moreover, we can derive optimal rates for some interesting cases, such as BV-regularization and Besov space regularization with $p = 1$, which do not seem to be covered by the analysis in [5].

We also derive convergence rates for oversmoothing regularization with statistical noise models covering both Besov space and BV-regularization. It seems that oversmoothing for statistical inverse problems has not received much attention in the literature so far; we are only aware of the preprint [22].

The remainder of this paper is organized as follows. In the following Section 2 we introduce our setting and prove our main result (Theorem 2.4) for the deterministic noise model (1.1). In Section 3 we formulate and discuss a convergence rate theorem for general oversmoothing Besov space regularization as a corollary to Theorem 2.4. In the following Section 4, we show L^p-error bounds for oversmoothing bounded variation regularization in a further corollary to Theorem 2.4. Oversmoothing regularization for statistical inverse problems is treated in Section 5 by adapting the proof of Theorem 2.4. We also discuss a parameter identification problem for an elliptic differential equation as a specific example and confirm the predicted convergence rates for this example in numerical experiments. The paper finishes with some conclusions and three appendices collecting results on interpolation theory, Besov spaces, and functions of bounded variation.

2. Deterministic analysis. In this section we present our main result. We will assume that $\mathbb{X}_{\mathcal{R}}$ is a quasi-Banach space. Recall that a quasi-Banach space $\mathbb{X}$ with norm $\|\cdot\|_{\mathbb{X}}$ satisfies

ETNA
Kent State University and
Johann Radon Institute (RICAM)

all axioms of a Banach space except for the triangle inequality, which only holds true in the weaker form $\|x+y\|_{\mathbb{X}} \leq c_{\mathbb{X}}(\|x\|_{\mathbb{X}} + \|y\|_{\mathbb{X}})$ with some constant $c_{\mathbb{X}} \geq 1$ independent of $x, y \in \mathbb{X}$. The most prominent examples of quasi-Banach spaces that are not Banach spaces are L^p and ℓ^p spaces with $p \in (0,1)$. A number of authors (see, e.g., [4, 25, 33]) have proposed ℓ^p penalty terms with $p \in (0,1)$ with the aim to enforce more sparsity of the regularizers, and this is our reason for not confining ourselves to Banach space penalties. Quasi-Banach space penalties do not cause any additional complications in our analysis and may thus be considered the natural setting for our approach.

2.1. Real interpolation of quasi-Banach spaces. Our analysis is based on real interpolation theory of quasi-Banach spaces via the K-method, which we will recall in the following.

Let $\mathbb{X}$ and $\mathbb{X}_-$ be quasi-Banach spaces with a continuous embedding $\mathbb{X} \subset \mathbb{X}_-$. The K-functional is given by

$$K(t,f) = \inf_{h \in \mathbb{X}} [\|f-h\|_{\mathbb{X}_-} + t\|h\|_{\mathbb{X}}] \quad \text{for } t>0 \text{ and } f \in \mathbb{X}_-. \tag{2.1}$$

With this, a scale of quasi-norms is defined by

$$\|f\|_{(\mathbb{X}_-,\mathbb{X})_{\theta,q}} = \left(\int_0^\infty (t^{-\theta} K(t,f))^q \, \frac{\mathrm{d}t}{t} \right)^{1/q}$$

for $0 < \theta < 1$ and $q \in [1,\infty)$, and

$$\|f\|_{(\mathbb{X}_-,\mathbb{X})_{\theta,\infty}} = \sup_{t>0} t^{-\theta} K(t,f)$$

for $0 \leq \theta \leq 1$. We obtain quasi-Banach spaces $(\mathbb{X}_-,\mathbb{X})_{\theta,q}$ consisting of all $f \in \mathbb{X}_-$ with $\|f\|_{(\mathbb{X}_-,\mathbb{X})_{\theta,q}} < \infty$; see, e.g., [3, Section 3.11].

2.2. Assumptions and preliminaries. Our basic assumption on the forward operator F is a two-sided Lipschitz condition with respect to the norm in $\mathbb{X}_-$. Similar conditions have been imposed in all previous papers on oversmoothing Tikhonov regularization that we are aware of. We start with F defined on $\tilde{D}_F := D_F \cap \mathbb{X}_{\mathcal{R}}$.

ASSUMPTION 2.1. *Suppose $\mathbb{X}_{\mathcal{R}}$ is a quasi-Banach space and $\mathbb{Y}$ is a Banach space, $\tilde{D}_F \subset \mathbb{X}_{\mathcal{R}}$, and $F\colon \tilde{D}_F \to \mathbb{Y}$ a map. Moreover, we assume that $\mathbb{X}_{\mathcal{R}}$ continuously embeds into a Banach space $\mathbb{X}_-$ with*

$$\frac{1}{M_1}\|f_1 - f_2\|_{\mathbb{X}_-} \leq \|F(f_1) - F(f_2)\|_{\mathbb{Y}} \leq M_2 \|f_1 - f_2\|_{\mathbb{X}_-} \quad \text{for all } f_1, f_2 \in \tilde{D}_F$$

for some constants $M_1, M_2 > 0$. Finally, let $\xi \in [0,1)$. If $\xi \in (0,1)$, let $\mathbb{X}_{\mathrm{L}}$ be a Banach space and suppose that there exists a continuous embedding

$$(\mathbb{X}_-, \mathbb{X}_{\mathcal{R}})_{\xi,1} \subset \mathbb{X}_{\mathrm{L}}.$$

If $\xi = 0$, we set $\mathbb{X}_{\mathrm{L}} := \mathbb{X}_-$.

Note that, under this assumption, F has a unique continuous extension, again denoted by F, to the norm closure D_F of $\tilde{D}_F$ in $\mathbb{X}_-$.

We start with a lemma that introduces smooth approximations to f based on real interpolation theory and provides estimates of their approximation rates in $\mathbb{X}_-$ and $\mathbb{X}_{\mathrm{L}}$ and their growth rate in $\mathbb{X}_{\mathcal{R}}$.

LEMMA 2.2 (Smooth approximations). *Suppose Assumption 2.1 holds true. Let $\theta \in (\xi, 1]$ and $\varrho > 0$. Suppose $f \in (\mathbb{X}_-, \mathbb{X}_{\mathcal{R}})_{\theta,\infty}$ with $\|f\|_{(\mathbb{X}_-,\mathbb{X}_{\mathcal{R}})_{\theta,\infty}} \leq \varrho$. Then there exists*

ETNA
Kent State University and
Johann Radon Institute (RICAM)

a net $(f_t)_{t>0} \subset \mathbb{X}_{\mathcal{R}}$ *such that the following bounds hold true:*

$$\|f - f_t\|_{\mathbb{X}_-} \leq 2\varrho t^{\theta}, \tag{2.2a}$$

$$\|f - f_t\|_{\mathbb{X}_{\mathrm{L}}} \leq C_{\mathrm{L}} \varrho t^{\theta-\xi}, \tag{2.2b}$$

$$\|f_t\|_{\mathbb{X}_{\mathcal{R}}} \leq 2\varrho t^{\theta-1}. \tag{2.2c}$$

Here $C_{\mathrm{L}} > 0$ *denotes a constant that is independent of* ϱ*,* t*, and* f*.*

Proof. Recall that $\|f\|_{(\mathbb{X}_-,\mathbb{X}_{\mathcal{R}})_{\theta,\infty}} \leq \varrho$ implies $K(t,f) \leq \varrho t^{\theta}$ for $t > 0$ with the K-functional from (2.1). Hence, for every $t > 0$ there exists $f_t \in \mathbb{X}_{\mathcal{R}}$ such that

$$\|f - f_t\|_{\mathbb{X}_-} + t\|f_t\|_{\mathbb{X}_{\mathcal{R}}} \leq 2K(t,f) \leq 2\varrho t^{\theta}.$$

We neglect the first summand on the left-hand side to see (2.2c), and the second to obtain (2.2a). This finishes the proof for $\xi = 0$, and we now turn to the case $\xi \in (0,1)$.

As an intermediate step to (2.2b), we first prove that $\|f - f_t\|_{(\mathbb{X}_-,\mathbb{X}_{\mathcal{R}})_{\theta,\infty}} \leq 3\varrho$ for all $t > 0$. To this end we first consider $s \geq t$ and insert $h = 0$ into the K-functional to wind up with

$$K(s, f - f_t) = \inf_{h\in\mathbb{X}_{\mathcal{R}}} [\|f - f_t - h\|_{\mathbb{X}_-} + s\|h\|_{\mathbb{X}_{\mathcal{R}}}] \leq \|f - f_t\|_{\mathbb{X}_-} \leq 2\varrho t^{\theta} \leq 2\varrho s^{\theta}.$$

For $s \leq t$ we substitute $h = h' - f_t$ and use the triangle inequality in $\mathbb{X}_{\mathcal{R}}$ to estimate

$$\begin{aligned} K(s, f - f_t) &= \inf_{h'\in\mathbb{X}_{\mathcal{R}}} [\|f - h'\|_{\mathbb{X}_-} + s\|h' - f_t\|_{\mathbb{X}_{\mathcal{R}}}] \\ &\leq K(s,f) + s\|f_t\|_{\mathbb{X}_{\mathcal{R}}} \leq \varrho s^{\theta} + 2\varrho s t^{\theta-1} \leq 3\varrho s^{\theta}. \end{aligned}$$

From the last two inequalities we conclude that

$$\|f - f_t\|_{(\mathbb{X}_-,\mathbb{X}_{\mathcal{R}})_{\theta,\infty}} = \sup_{s>0} s^{-\theta} K(f - f_t, s) \leq 3\varrho.$$

By the reiteration theorem (see Proposition A.3) we have

$$\mathbb{X}_{\mathrm{L}} \supset (\mathbb{X}_-, \mathbb{X}_{\mathcal{R}})_{\xi,1} = (\mathbb{X}_-, (\mathbb{X}_-, \mathbb{X}_{\mathcal{R}})_{\theta,\infty})_{\xi/\theta,1},$$

with equivalent norms of the latter two spaces. Hence, Lemma A.1 provides an interpolation inequality $\|\cdot\|_{\mathbb{X}_{\mathrm{L}}} \leq c\|\cdot\|_{\mathbb{X}_-}^{1-\xi/\theta} \cdot \|\cdot\|_{(\mathbb{X}_-,\mathbb{X}_{\mathcal{R}})_{\theta,\infty}}^{\xi/\theta}$. Inserting $f - f_t$ we finally get

$$\|f - f_t\|_{\mathbb{X}_{\mathrm{L}}} \leq c(2\varrho t^{\theta})^{1-\xi/\theta}(3\varrho)^{\xi/\theta} \leq 3c\varrho t^{\theta-\xi}. \qquad \square$$

REMARK 2.3. From the existence of approximations as in Lemma 2.2 one can reclaim the regularity assumption as follows. Let $f \in \mathbb{X}_-$ and suppose that there exists a net $(f_t)_{t>0} \subset \mathbb{X}_{\mathcal{R}}$ such that the bounds (2.2a) and (2.2c) hold true. Inserting f_t for h in the K-functional yields $K(t,f) \leq 4\varrho t^{\theta}$. Hence $f \in (\mathbb{X}_-, \mathbb{X}_{\mathcal{R}})_{\theta,\infty}$ with $\|f\|_{(\mathbb{X}_-,\mathbb{X}_{\mathcal{R}})_{\theta,\infty}} \leq 4\varrho$.

2.3. Abstract convergence rate result. With Lemma 2.2 at hand, we are in position to prove the following convergence estimates as the main result of this paper.

THEOREM 2.4 (Error bounds). *Suppose Assumption 2.1 holds true. Let* $\theta \in (\xi, 1]$ *and* $\varrho > 0$ *. Assume that* $f \in (\mathbb{X}_-, \mathbb{X}_{\mathcal{R}})_{\theta,\infty}$ *with* $\|f\|_{(\mathbb{X}_-,\mathbb{X}_{\mathcal{R}})_{\theta,\infty}} \leq \varrho$*, and moreover that* D_F *contains an* $\mathbb{X}_{\mathrm{L}}$*-ball with radius* $\tau > 0$ *around* f*.*

ETNA
Kent State University and
Johann Radon Institute (RICAM)

1. *(Bias bounds) There exists a constant C_b independent of f, ϱ, and τ such that*

$$\|f - f_\alpha\|_{\mathbb{X}_-} \leq C_b \varrho^{u/((1-\theta)u+2\theta)} \alpha^{\theta/((1-\theta)u+2\theta)}, \tag{2.3a}$$

$$\|f - f_\alpha\|_{\mathbb{X}_\mathrm{L}} \leq C_b \varrho^{((1-\xi)u+2\xi)/((1-\theta)u+2\theta)} \alpha^{(\theta-\xi)/((1-\theta)u+2\theta)},$$

$$\|f_\alpha\|_{\mathbb{X}_\mathcal{R}} \leq C_b \varrho^{2/((1-\theta)u+2\theta)} \alpha^{(\theta-1)/((1-\theta)u+2\theta)} \tag{2.3b}$$

holds true for all $0 < \alpha < \varrho^{-((1-\theta)u+2\xi)/(\theta-\xi)} \tau^{((1-\theta)u+2\theta)/(\theta-\xi)}$ and $f_\alpha \in S_\alpha(F(f))$; see (1.2).

2. *(Rates with* a priori *choice of α) Let $0 < c_l \leq c_r$. Suppose $g^{\mathrm{obs}} \in \mathbb{Y}$ satisfies (1.1) with $0 < \delta < \varrho^{-\xi/(\theta-\xi)} \tau^{\theta/(\theta-\xi)}$. Let $\alpha > 0$ and $\hat{f}_\alpha \in S_\alpha(g^{\mathrm{obs}})$. There exists a constant C_c independent of f, g^{obs}, ϱ, τ, and δ such that*

$$c_l \varrho^{-u/\theta} \delta^{((1-\theta)u+2\theta)/\theta} \leq \alpha \leq c_r \varrho^{-u/\theta} \delta^{((1-\theta)u+2\theta)/\theta}$$

implies the bounds

$$\|f - \hat{f}_\alpha\|_{\mathbb{X}_-} \leq C_c \delta,$$
$$\|f - \hat{f}_\alpha\|_{\mathbb{X}_\mathrm{L}} \leq C_c \varrho^{\xi/\theta} \delta^{(\theta-\xi)/\theta},$$
$$\|\hat{f}_\alpha\|_{\mathbb{X}_\mathcal{R}} \leq C_c \varrho^{1/\theta} \delta^{(\theta-1)/\theta}.$$

3. *(Rates with discrepancy principle) Let $1 < c_\mathrm{D} \leq C_\mathrm{D}$. Suppose that $0 < \delta < \varrho^{-\xi/(\theta-\xi)} \tau^{\theta/(\theta-\xi)}$, and $g^{\mathrm{obs}} \in \mathbb{Y}$ with $\|g^{\mathrm{obs}} - F(f)\|_\mathbb{Y} \leq \delta$. Let $\alpha > 0$ and $\hat{f}_\alpha \in S_\alpha(g^{\mathrm{obs}})$. There exists a constant C_d independent of f, g^{obs}, ϱ, and δ such that*

$$c_\mathrm{D} \delta \leq \|g^{\mathrm{obs}} - F(\hat{f}_\alpha)\|_\mathbb{Y} \leq C_\mathrm{D} \delta$$

implies the bounds

$$\|f - \hat{f}_\alpha\|_{\mathbb{X}_\mathrm{L}} \leq C_d \varrho^{\xi/\theta} \delta^{(\theta-\xi)/\theta},$$
$$\|\hat{f}_\alpha\|_{\mathbb{X}_\mathcal{R}} \leq C_d \varrho^{1/\theta} \delta^{(\theta-1)/\theta}.$$

Proof. Let $(f_t)_{t>0}$ be as in Lemma 2.2.

1. We choose

$$t = C_\mathrm{L}{}^{-1/(\theta-\xi)} \varrho^{(u-2)/((1-\theta)u+2\theta)} \alpha^{1/((1-\theta)u+2\theta)} \tag{2.5}$$

with C_L from Lemma 2.2. Inequality (2.2b) yields

$$\|f - f_t\|_{\mathbb{X}_\mathrm{L}} \leq C_\mathrm{L} \varrho t^{\theta-\xi} = \varrho^{((1-\xi)u+2\xi)/((1-\theta)u+2\theta)} \alpha^{(\theta-\xi)/((1-\theta)u+2\theta)} < \tau. \tag{2.6}$$

Hence $f_t \in D_F$, i.e., we may insert f_t into the Tikhonov functional and use the Lipschitz condition of F, (2.2a), and (2.2c) to wind up with

$$\begin{aligned}
\frac{1}{2\alpha}\|F(f) - F(f_\alpha)\|_\mathbb{Y}^2 + \frac{1}{u}\|f_\alpha\|_{\mathbb{X}_\mathcal{R}}^u &\leq \frac{1}{2\alpha}\|F(f) - F(f_t)\|_\mathbb{Y}^2 + \frac{1}{u}\|f_t\|_{\mathbb{X}_\mathcal{R}}^u \\
&\leq \frac{M_2^2}{2\alpha}\|f - f_t\|_{\mathbb{X}_-}^2 + \frac{1}{u}\|f_t\|_{\mathbb{X}_\mathcal{R}}^u \\
&\leq \frac{2M_2^2}{\alpha}\varrho^2 t^{2\theta} + \frac{2^u}{u}\varrho^u t^{(\theta-1)u} \\
&= c_1 \varrho^{2u/((1-\theta)u+2\theta)} \alpha^{(\theta-1)u/((1-\theta)u+2\theta)},
\end{aligned}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

with c_1 depending on M_2, C_{L}, u, θ, and ξ. We neglect the penalty term and use the Lipschitz condition of the inverse of F to obtain the first bound

$$\|f-f_\alpha\|_{\mathbb{X}_-} \leq M_1\|F(f)-F(f_\alpha)\|_{\mathbb{Y}} \leq (2c_1)^{1/2} M_1 \varrho^{u/((1-\theta)u+2\theta)}\alpha^{\theta/((1-\theta)u+2\theta)}.$$

Together with (2.2a) we record

$$\|f_t-f_\alpha\|_{\mathbb{X}_-} \leq \|f-f_t\|_{\mathbb{X}_-} + \|f-f_\alpha\|_{\mathbb{X}_-} \leq c_2 \varrho^{u/((1-\theta)u+2\theta)}\alpha^{\theta/((1-\theta)u+2\theta)},$$

with c_2 depending on C_{L}, c_1, M_1, θ, and ξ.

Neglecting the data fidelity term in the above estimation of the Tikhonov functional provides

$$\|f_\alpha\|_{\mathbb{X}_{\mathcal{R}}} \leq (c_1 u)^{1/u} \varrho^{2/((1-\theta)u+2\theta)}\alpha^{(\theta-1)/((1-\theta)u+2\theta)}.$$

Furthermore, we see that $\|f_t\|_{\mathbb{X}_{\mathcal{R}}}$ satisfies the same upper bound. With the triangle inequality in $\mathbb{X}_{\mathcal{R}}$ we combine

$$\begin{aligned}\|f_t-f_\alpha\|_{\mathbb{X}_{\mathcal{R}}} &\leq c_{\mathbb{X}_{\mathcal{R}}}(\|f_t\|_{\mathbb{X}_{\mathcal{R}}} + \|f_\alpha\|_{\mathbb{X}_{\mathcal{R}}})\\ &\leq 2c_{\mathbb{X}_{\mathcal{R}}}(c_1 u)^{1/u} \varrho^{2/((1-\theta)u+2\theta)}\alpha^{(\theta-1)/((1-\theta)u+2\theta)}.\end{aligned}$$

Next, the interpolation inequality $\|\cdot\|_{\mathbb{X}_{\mathrm{L}}} \leq c_3\|\cdot\|_{\mathbb{X}_-}^{1-\xi}\cdot\|\cdot\|_{\mathbb{X}_{\mathcal{R}}}^{\xi}$ (see Lemma A.1) furnishes

$$\|f_t-f_\alpha\|_{\mathbb{X}_{\mathrm{L}}} \leq c_4 \varrho^{((1-\xi)u+2\xi)/((1-\theta)u+2\theta)}\alpha^{(\theta-\xi)/((1-\theta)u+2\theta)},$$

with c_4 depending on c_1, c_2, c_3, $c_{\mathbb{X}_{\mathcal{R}}}$, u, and ξ. Together with (2.6) we finally obtain

$$\begin{aligned}\|f-f_\alpha\|_{\mathbb{X}_{\mathrm{L}}} &\leq \|f-f_t\|_{\mathbb{X}_{\mathrm{L}}} + \|f_t-f_\alpha\|_{\mathbb{X}_{\mathrm{L}}}\\ &\leq (1+c_4)\varrho^{((1-\xi)u+2\xi)/((1-\theta)u+2\theta)}\alpha^{(\theta-\xi)/((1-\theta)u+2\theta)}.\end{aligned}$$

2. Taking $t = {C_{\mathrm{L}}}^{-1/(\theta-\xi)}\varrho^{-1/\theta}\delta^{1/\theta}$ we have

$$\|f-f_t\|_{\mathbb{X}_{\mathrm{L}}} \leq C_{\mathrm{L}}\varrho t^{\theta-\xi} = \varrho^{\xi/\theta}\delta^{(\theta-\xi)/\theta} < \tau. \tag{2.7}$$

This ensures that $f_t \in D_F$. We insert into the Tikhonov functional, use the elementary inequality $(a+b)^2 \leq 2a^2+2b^2$ for $a,b \geq 0$, (2.2a), (2.2c), the Lipschitz condition of F, and the choice of α to estimate

$$\begin{aligned}&\frac{1}{2\alpha}\|g^{\mathrm{obs}}-F(\hat{f}_\alpha)\|_{\mathbb{Y}}^2 + \frac{1}{u}\|\hat{f}_\alpha\|_{\mathbb{X}_{\mathcal{R}}}^u\\ &\leq \frac{1}{2\alpha}\|g^{\mathrm{obs}}-F(f)+F(f)-F(f_t)\|_{\mathbb{Y}}^2 + \frac{1}{u}\|f_t\|_{\mathbb{X}_{\mathcal{R}}}^u\\ &\leq \frac{\delta^2}{\alpha} + \frac{M_2^2}{\alpha}\|f-f_t\|_{\mathbb{X}_-}^2 + \frac{2^u}{u}\varrho^u t^{(\theta-1)u}\\ &\leq (1+4M_2^2{C_{\mathrm{L}}}^{-2\theta/(\theta-\xi)})\frac{\delta^2}{\alpha} + \frac{2^u}{u}{C_{\mathrm{L}}}^{(1-\theta)u/(\theta-\xi)}\varrho^{u/\theta}\delta^{(\theta-1)u/\theta}\\ &\leq c_5\varrho^{u/\theta}\delta^{(\theta-1)u/\theta},\end{aligned}$$

with c_5 depending on c_l, M_2, C_{L}, u, θ, and ξ.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Now we follow the argument in part 1. From the last inequality and the triangle inequality in $\mathbb{Y}$ we get

$$\begin{aligned}\|f-\hat{f}_\alpha\|_{\mathbb{X}_-} &\le M_1\|F(f)-g^{\rm obs}+g^{\rm obs}-F(\hat{f}_\alpha)\|_{\mathbb{Y}}\\ &\le M_1\delta + M_1(2c_5)^{1/2}\alpha^{1/2}\varrho^{u/(2\theta)}\delta^{((\theta-1)u)/(2\theta)}\\ &\le M_1(1+(2c_5c_r)^{1/2})\delta,\end{aligned}$$

which, together with (2.2a), implies $\|f_t-\hat{f}_\alpha\|_{\mathbb{X}_-}\le c_6\delta$, with c_6 depending on $C_{\rm L}$, M_1, c_5, c_r, θ, and ξ. Moreover, $\|\hat{f}_\alpha\|_{\mathbb{X}_\mathcal{R}}, \|f_t\|_{\mathbb{X}_\mathcal{R}}\le (c_5u)^{1/u}\varrho^{1/\theta}\delta^{(\theta-1)/\theta}$. Hence

$$\|f_t-\hat{f}_\alpha\|_{\mathbb{X}_\mathcal{R}}\le 2c_{\mathbb{X}_\mathcal{R}}(c_5u)^{1/u}\varrho^{1/\theta}\delta^{(\theta-1)/\theta}$$

by the triangle inequality in $\mathbb{X}_\mathcal{R}$.

We use the above interpolation inequality to combine the last two inequalities into $\|f_t-\hat{f}_\alpha\|_{\mathbb{X}_{\rm L}}\le c_7\varrho^{\xi/\theta}\delta^{(\theta-\xi)/\theta}$, with c_7 depending on c_3, c_6, c_5, $c_{\mathbb{X}_\mathcal{R}}$, u, and ξ. With (2.7) we conclude that

$$\|f-\hat{f}_\alpha\|_{\mathbb{X}_{\rm L}}\le (1+c_7)\varrho^{\xi/\theta}\delta^{(\theta-\xi)/\theta}.$$

3. We set $\varepsilon := \min\{(c_{\rm D}^2-1)/2, 4M_2C_{\rm L}^{-2\theta/(\theta-\xi)}\}$. Then $\varepsilon>0$. Furthermore, we take

$$t=\left(\frac{(4M_2)^{-1}\varepsilon}{1+\varepsilon^{-1}}\right)^{1/(2\theta)}\varrho^{-1/\theta}\delta^{1/\theta}.$$

Then (2.2a) reads as

$$\|f-f_t\|_{\mathbb{X}_-}\le 2\varrho t^\theta = \left(\frac{\varepsilon}{1+\varepsilon^{-1}}\right)^{1/2}M_2^{-1/2}\delta. \tag{2.8}$$

Due to (2.2b) we obtain

$$\begin{aligned}\|f-f_t\|_{\mathbb{X}_{\rm L}} &\le C_{\rm L}\varrho t^{\theta-\xi}\le C_{\rm L}((4M_2)^{-1}\varepsilon)^{(\theta-\xi)/(2\theta)}\varrho^{\xi/\theta}\delta^{(\theta-\xi)/\theta}\\ &\le \varrho^{\xi/\theta}\delta^{(\theta-\xi)/\theta}<\tau,\end{aligned} \tag{2.9}$$

which provides $f_t\in D_F$.

In the following we use the elementary inequality $(a+b)^2\le(1+\varepsilon)a^2+(1+\varepsilon^{-1})b^2$ for all $a,b\ge 0$ (which is proven by expanding the square and applying Young's inequality on the mixed term) and (2.8) to estimate

$$\begin{aligned}\|g^{\rm obs}-F(f_t)\|_{\mathbb{Y}}^2 &\le (1+\varepsilon)\delta^2+(1+\varepsilon^{-1})\|F(f)-F(f_t)\|_{\mathbb{Y}}^2\\ &\le (1+\varepsilon)\delta^2+(1+\varepsilon^{-1})M_2\|f-f_t\|_{\mathbb{X}_-}^2\\ &\le (1+2\varepsilon)\delta^2\le c_{\rm D}^2\delta^2\le\|g^{\rm obs}-F(\hat{f}_\alpha)\|_{\mathbb{Y}}^2.\end{aligned}$$

Therefore, a comparison of the Tikhonov functional taken at $\hat{f}_\alpha$ and f_t, and (2.2c) yields

$$\|\hat{f}_\alpha\|_{\mathbb{X}_\mathcal{R}}\le\|f_t\|_{\mathbb{X}_\mathcal{R}}\le 2\varrho t^{\theta-1}=c_8\varrho^{1/\theta}\delta^{(\theta-1)/\theta},$$

with c_8 depending on M_2, ε, and θ. Hence $\|f_t-\hat{f}_\alpha\|_{\mathbb{X}_\mathcal{R}}\le 2c_{\mathbb{X}_\mathcal{R}}c_8\varrho^{1/\theta}\delta^{(\theta-1)/\theta}$. Moreover,

$$\|g_t-F(\hat{f}_\alpha)\|_{\mathbb{Y}}\le\|g^{\rm obs}-g_t\|_{\mathbb{Y}}+\|g^{\rm obs}-F(\hat{f}_\alpha)\|_{\mathbb{Y}}\le 2C_{\rm D}\delta.$$

Therefore, $\|f_t - \hat{f}_\alpha\|_{\mathbb{X}_-} \leq M_1 C_\mathrm{D}\delta$ by the Lipschitz condition. As above, we conclude that $\|f_t - \hat{f}_\alpha\|_{\mathbb{X}_\mathrm{L}} \leq c_9 \varrho^{\xi/\theta}\delta^{(\theta-\xi)/\theta}$, with c_9 depending on c_3, C_D, c_8, $c_{\mathbb{X}_\mathcal{R}}$, and ξ, and we use (2.9) to finish up with

$$\|f - \hat{f}_\alpha\|_{\mathbb{X}_\mathrm{L}} \leq (1 + c_9)\varrho^{\xi/\theta}\delta^{(\theta-\xi)/\theta}. \qquad \square$$

We discuss our results in a series of remarks.

REMARK 2.5 (Interior point). The requirement that f be an interior point of the domain in $\mathbb{X}_\mathrm{L}$ may be weakened to the requirement that elements f_t satisfying the bounds given in Lemma 2.2 belong to D_F for t small enough.

REMARK 2.6 (Influence of the exponent u). A strength of the above theorem is that it provides convergence rates for all exponents u. Note that the choice of u does not influence the rate while it does influence the bias bounds and the parameter choice rule. An inspection of the *a priori* rule shows that a larger u allows for a larger choice of the parameter α. The flexibility in the choice of u in our theory is a remarkable difference to many other variational convergence theories where one has to pick a specific exponent; see, e.g., [15, 32]. The authors also do not expect any difficulty in generalizing this result to exponents other than 2 in the data fidelity.

REMARK 2.7 (Equivalent norms). The presented theory relies on a purely quasi-Banach space theoretic framework. That is, as we do not appeal to any metric or convex notions like subdifferentials or convexity, the result in Theorem 2.4 stays the same up to a change of the constants if we change the norm on any of the occurring spaces up to equivalence. This has an important impact on regularization with wavelet penalties that we will discuss in the next section.

Once again this is a major difference to classical variational regularization theory. For example, it is not clear how the subdifferential of a norm involved in the source condition $A^*\omega \in \partial\mathcal{R}(f)$ for a linear operator A changes if the norm is replaced by an equivalent one. Also classical variational source conditions are characterized by the smoothness of $\partial\mathcal{R}$ rather than by the smoothness of f (see [32]), and the former may change if the norm in the penalty term is replaced by an equivalent norm.

REMARK 2.8 (Converse result). Suppose minimizers in (1.2) exist for all $g \in \mathbb{Y}$ and $\alpha > 0$, and $D_F = \mathbb{X}_-$. In view of Remark 2.3, one can reclaim $f \in (\mathbb{X}_-, \mathbb{X}_\mathcal{R})_{\theta,\infty}$ from the bias bound (2.3a) together with (2.3b), as, with $\alpha(t) = c\varrho^{2-u}t^{(1-\theta)u+2\theta}$ for a suitably chosen constant c depending only on C_b, a net $(f_{\alpha(t)})_{t>0}$ with $f_{\alpha(t)} \in S_{\alpha(t)}$ satisfies the bounds (2.2a) and (2.2c) in Lemma 2.2.

REMARK 2.9 (Limiting case $\theta = 1$). In the case $\theta = 1$ the parameter choice rule in Theorem 2.4 becomes $\alpha \sim \delta^2$. Here the results provide boundedness of the estimators f_α and $\hat{f}_\alpha$ in $\mathbb{X}_\mathcal{R}$. Due to Proposition A.2, we have $\mathbb{X}_\mathcal{R} \subset (\mathbb{X}_-, \mathbb{X}_\mathcal{R})_{1,\infty}$. The latter two spaces agree if $\mathbb{X}_\mathcal{R}$ is reflexive; see [27, 1.3.2 Remark 2]).

Before we illustrate our theorem by simple sequence space models, let us point out that, in contrast to [5, 15], we do not need to require that $c_\mathrm{D} = C_\mathrm{D}$ in the discrepancy principle. As also mentioned in [5], this is desirable in view of practical implementations.

EXAMPLE 2.10 (Embedding operators in sequence spaces).

- Let $p \in (0, 2)$ and $u \in (0, \infty)$. We consider $\mathbb{X}_\mathcal{R} = \ell^p$, $\mathbb{Y} = \mathbb{X}_\mathrm{L} = \mathbb{X}_- = \ell^2$, and $F\colon \mathbb{X}_\mathcal{R} \to \mathbb{Y}$ with $x \mapsto x$ the embedding operator. Then Assumption 2.1 holds true with $\xi = 0$.

  Let $v \in (p, 2)$; then we obtain (see, e.g., [12])

$$(\ell^2, \ell^p)_{\theta_v,\infty} = \omega\ell^v \quad \text{with} \quad \theta_v = \frac{p(2-v)}{v(2-p)}.$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Here $\omega\ell^v$ stands for the weak ℓ^v-space given by the quasi-norm

$$\|x\|_{\omega\ell^v}^v = \sup_{\alpha>0} \alpha^v \#\{|x_k| > \alpha\}.$$

Theorem 2.4 yields that $x \in \omega\ell^v$ implies

$$\|x - \hat{x}_\alpha\|_2 = \mathcal{O}(\delta) \quad \text{and} \quad \|\hat{x}_\alpha\|_p = \mathcal{O}(\delta^{2(p-v)/(p(2-v))}),$$

with $\hat{x}_\alpha \in \operatorname{argmin}_{z\in\ell^p}[(1/2\alpha)\|x^{\text{obs}} - z\|_2^2 + (1/u)\|z\|_p^u]$ and $\|x - x^{\text{obs}}\|_2 \leq \delta$, and either of the parameter choice rules specified in Theorem 2.4.

- Once again let $p \in (0, 2)$ and $u \in (0, \infty)$. Now we consider $\mathbb{X}_\mathcal{R} = \ell^p$, $\mathbb{Y} = \mathbb{X}_- = \ell^\infty$, $\mathbb{X}_\mathrm{L} = \ell^2$, and again $F\colon \mathbb{X}_\mathcal{R} \to \mathbb{Y}$ the embedding operator. With $\xi = p/2$, the continuous embedding

$$(\ell^\infty, \ell^p)_{\xi,1} \subset (\ell^\infty, \ell^p)_{\xi,2} = \ell^2$$

yields Assumption 2.1.

For $v \in (p, 2)$ we have $(\ell^\infty, \ell^p)_{p/v,\infty} = \omega\ell^v$. Hence for $x \in \omega\ell^v$ we obtain

$$\|x - \hat{x}_\alpha\|_2 = \mathcal{O}(\delta^{(2-v)/2}) \quad \text{and} \quad \|\hat{x}_\alpha\|_p = \mathcal{O}(\delta^{(p-v)/p}),$$

with $\hat{x}_\alpha \in \operatorname{argmin}_{z\in\ell^p}[(1/2\alpha)\|x^{\text{obs}} - z\|_\infty^2 + (1/u)\|z\|_p^u]$, $\|x - x^{\text{obs}}\|_\infty \leq \delta$, and either of the parameter choice rules specified in Theorem 2.4.

3. Besov space regularization. In this section we apply Theorem 2.4 to regularization of finitely smoothing operators with Besov space penalty term. For a comprehensive treatment of Besov spaces, we refer the reader to [28, 29, 30], and also to [14, Chapter 4] for a self-contained introduction and applications in statistics. The Besov space $B^s_{p,q}(\mathbb{R}^d)$ for a smoothness index $s \in \mathbb{R}$, an integrability index $p \in (0, \infty]$, and a fine index $q \in (0, \infty]$ with quasi-norms $\|\cdot\|_{B^s_{p,q}(\mathbb{R}^d)}$ can be defined in several equivalent ways, among others via a dyadic partition of unity in Fourier space, via the modulus of continuity, or via wavelet decompositions. In contrast to the analysis of non-oversmoothing Besov regularization in [17, 19, 20, 32], it will not matter here which of these equivalent norms is used in the following.

In the following, let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. Then $B^s_{p,q}(\Omega) := \{f|_\Omega : f \in B^s_{p,q}(\mathbb{R}^d)\}$ with $\|g\|_{B^s_{p,q}} := \inf\{\|f\|_{B^s_{p,q}(\mathbb{R}^d)} : f|_\Omega = g\}$ is a quasi-Banach space, and even a Banach space if $p, q \geq 1$; see [29]. Some properties of these spaces and relations to other function spaces are summarized in Appendix B.

Throughout this section we use $\mathbb{X}_\mathcal{R} := B^r_{p,q}(\Omega)$ for fixed $r > 0$ and $p, q \in (0, \infty]$, and consider the regularization scheme

$$S_\alpha(g) = \operatorname*{argmin}_{h\in\tilde{D}_F}\left[\frac{1}{2\alpha}\|g - F(h)\|_\mathbb{Y}^2 + \frac{1}{u}\|h\|_{B^r_{p,q}}^u\right], \quad g \in \mathbb{Y}, \tag{3.1}$$

for a fixed exponent $u \in (0, \infty)$. A natural choice is $u = q$.

3.1. Convergence rate result. We first formulate our assumptions on the forward operator. Recall that $B^s_{2,2}(\Omega) = W^s_2(\Omega)$ with equivalent norms for all $s \in \mathbb{R}$; see Proposition B.1.

ASSUMPTION 3.1. *Suppose that $a \geq 0$ and $B^r_{p,q}(\Omega) \subset B^{-a}_{2,2}(\Omega)$ with continuous embedding. Let $\tilde{D}_F \subset B^r_{p,q}(\Omega)$, $\mathbb{Y}$ be a Banach space, and $F\colon \tilde{D}_F \to \mathbb{Y}$ be a map satisfying*

$$\frac{1}{M_1}\|f_1 - f_2\|_{B^{-a}_{2,2}} \leq \|F(f_1) - F(f_2)\|_\mathbb{Y} \leq M_2\|f_1 - f_2\|_{B^{-a}_{2,2}} \quad \text{for all } f_1, f_2 \in \tilde{D}_F$$

for constants $M_1, M_2 > 0$.

The assumption of a continuous embedding $B^r_{p,q}(\Omega) \subset B^{-a}_{2,2}(\Omega)$ is satisfied if $a + r > d(1/p - 1/2)$; see (B.3). For $q = 2$, even the condition $a + r \geq d(1/p - 1/2)$ suffices; see (B.2).

Now we state and prove the convergence rate result for oversmoothing Besov space regularization. We first state our theorem under the abstract smoothness condition given by the maximal real interpolation space in Theorem 2.4, and discuss how to find more handy smoothness conditions in terms of Besov spaces afterwards. For the sake of brevity, we do not state the bounds on the bias.

COROLLARY 3.2 (Rates for oversmoothing Besov space regularization). *Consider the regularization scheme (3.1) for some $p, q \in (0, \infty]$ with $q \leq p$, $r > 0$, and $u \in (0, \infty)$, such that $\overline{p} := 2p(a + r)/(2a + pr) \geq 1$ (i.e., $p \geq 2a/(2a + r)$), and suppose Assumption 3.1 holds true. Assume the true solution f has smoothness index $s \in (0, r]$ in the sense that*

(3.2)
$$f \in (B^{-a}_{2,2}(\Omega), B^r_{p,q}(\Omega))_{\theta_s,\infty} \quad \textit{for} \quad \theta_s := \frac{s+a}{a+r} \quad \textit{and} \quad \|f\|_{(B^{-a}_{2,2}(\Omega), B^r_{p,q}(\Omega))_{\theta_s,\infty}} \leq \varrho,$$

for some $\varrho > 0$; see also Remark 3.4. Suppose that the closure D_F of $\tilde{D}_F$ in $B^{-a}_{2,2}(\Omega)$ contains a $B^0_{\overline{p},\overline{p}}(\Omega)$-ball with radius $\tau > 0$ around f. Suppose also that $g^{\mathrm{obs}} \in \mathbb{Y}$ satisfies (1.1) for $0 < \delta < \varrho^{-a/s}\tau^{(s+a)/s}$ and $\hat{f}_\alpha \in S_\alpha(g^{\mathrm{obs}})$ defined in (3.1) for some $\alpha > 0$. Let $0 < c_l \leq c_r$ and $1 < c_{\mathrm{D}} \leq C_{\mathrm{D}}$. Then there is a constant C_r independent of f, g^{obs}, ϱ, τ, and δ such that either of the conditions

$$c_l\varrho^{-u(a+r)/(s+a)}\delta^{((2-u)s+2a+ur)/(s+a)} \leq \alpha \leq c_r\varrho^{-u(a+r)/(s+a)}\delta^{((2-u)s+2a+ur)/(s+a)}$$

and

$$c_{\mathrm{D}}\delta \leq \|g^{\mathrm{obs}} - F(\hat{f}_\alpha)\|_{\mathbb{Y}} \leq C_{\mathrm{D}}\delta$$

on the choice of α implies the following bounds:

$$\begin{aligned} \|f - \hat{f}_\alpha\|_{B^{-a}_{2,2}} &\leq C_r\delta, \\ \|f - \hat{f}_\alpha\|_{B^0_{\overline{p},\overline{p}}} &\leq C_r\varrho^{a/(s+a)}\delta^{s/(s+a)}, \\ \|\hat{f}_\alpha\|_{B^r_{p,q}} &\leq C_r\varrho^{(a+r)/(s+a)}\delta^{(s-r)/(s+a)}. \end{aligned}$$

Proof. We set $\xi := a/(a+r)$, $\mathbb{X}_{\mathcal{R}} = B^r_{p,q}(\Omega)$, $\mathbb{X}_- := B^{-a}_{2,2}(\Omega)$, and $\mathbb{X}_{\mathrm{L}} := B^0_{\overline{p},\overline{p}}(\Omega)$, and verify Assumption 2.1. The two-sided Lipschitz condition holds true due to Assumption 3.1. If $a = 0$, then $\xi = 0$ and we have $\overline{p} = 2$. Therefore, $\mathbb{X}_- = \mathbb{X}_{\mathrm{L}} = B^0_{2,2}(\Omega) = L^2(\Omega)$. If $a > 0$ we use $q \leq p$ and $1 \leq \overline{p}$ to obtain the following chain of continuous embeddings:

(3.3)
$$(B^{-a}_{2,2}(\Omega), B^r_{p,q}(\Omega))_{\xi,1} \subset (B^{-a}_{2,2}(\Omega), B^r_{p,q}(\Omega))_{\xi,\overline{p}} \subset (B^{-a}_{2,2}(\Omega), B^r_{p,p}(\Omega))_{\xi,\overline{p}} = B^0_{\overline{p},\overline{p}}(\Omega);$$

see [3, Theorem 3.4.1(b)] for the first embedding, (B.4) for the second, and (B.6) for the interpolation identity. This shows Assumption 2.1, i.e., $(\mathbb{X}_-, \mathbb{X}_{\mathcal{R}})_{\xi,1} \subset \mathbb{X}_{\mathrm{L}}$, and the result follows from Theorem 2.4. ◻

In contrast to the analysis in [17, 19, 20, 32], which is restricted to certain choices of r, p, q, and u, our only restrictions on the parameters are $r > 0$, $q \leq p$, and $\overline{p} \geq 1$. We will see that the assumption $q \leq p$ can be dropped by some refined argument using a complex interpolation identity.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

We further discuss our results in the following remarks.

REMARK 3.3 ($L^{\overline{p}}$ loss). Suppose $p \le 2$. Then $\overline{p} \le 2$. Hence the continuous embedding $B^0_{\overline{p},\overline{p}}(\Omega) \subset L^{\overline{p}}(\Omega)$ (see (B.1)) together with (3.3) yields $(B^{-a}_{2,2}(\Omega), B^r_{p,q}(\Omega))_{\xi,1} \subset L^{\overline{p}}(\Omega)$. Therefore, Corollary 3.2 remains valid word for word if one replaces $B^0_{\overline{p},\overline{p}}(\Omega)$ by $L^{\overline{p}}(\Omega)$ in this case.

REMARK 3.4 (Smoothness condition). Suppose that $p, q \ge 1$. By the complex interpolation (B.7) we have

$$\begin{aligned} {}_{\theta_s} &= B^s_{p_s,q_s}(\Omega), \\ p_s = \frac{2p(a+r)}{s(2-p)+2a+pr}&, \quad q_s = \frac{2q(a+r)}{s(2-q)+2a+qr}. \end{aligned} \tag{3.4}$$

With this we obtain a continuous embedding $B^s_{p_s,q_s}(\Omega) \subset (B^{-a}_{2,2}(\Omega), B^r_{p,q}(\Omega))_{\theta_s,\infty}$, as, for Banach spaces, the complex interpolation space $[\cdot,\cdot]_\theta$ is always continuously embedded in the real interpolation space $(\cdot,\cdot)_{\theta,\infty}$; see [3, Theorem 4.7.1]. Hence the statements in Corollary 3.2 remain true if the smoothness assumption on f formulated in terms of $(B^{-a}_{2,2}(\Omega), B^r_{p,q}(\Omega))_{\theta_s,\infty}$ is replaced by

$$f \in B^s_{p_s,q_s}(\Omega) \quad \text{with} \quad \|f\|_{B^s_{p_s,q_s}} \le \varrho. \tag{3.5}$$

REMARK 3.5 (Assumption $q \le p$). We also comment on the assumption $q \le p$, again for the Banach space case $p, q \ge 1$. Using complex interpolation, this restriction can be dropped as follows. Since the real interpolation space $(\cdot,\cdot)_{\theta,1}$ is always continuously embedded in the complex interpolation space (see [3, Theorem 4.7.1.]), identity (3.4) yields a continuous embedding

$$(B^{-a}_{2,2}(\Omega), B^r_{p,q}(\Omega))_{\xi,1} \subset B^0_{\overline{p},\overline{q}}(\Omega)$$

for $\overline{p}$ as in Corollary 3.2 and $\overline{q} := 2q(a+r)/(2a+qr)$. Hence the statements in Corollary 3.2 remain true in the case $q > p$ if one replaces $B^0_{\overline{p},\overline{p}}(\Omega)$ by $B^0_{\overline{p},\overline{q}}(\Omega)$.

REMARK 3.6 (Other domains and boundary conditions). For the sake of clarity, we have confined ourselves to bounded Lipschitz domains $\Omega \subset \mathbb{R}^d$ and to the Besov spaces $B^s_{p,q}(\Omega)$. However, Corollary 3.2 relies on only the interpolation identity (B.6), the embedding (B.4), and the embedding stated in Assumption 3.1. These are also valid in many other situations (sometimes under additional assumptions), e.g., for certain unbounded domains (in particular, $\mathbb{R}^d$ and half-spaces; see [30]), certain Riemannian manifolds (see [28, Chapter 7]), as well as Besov spaces with other boundary conditions (see [27, Chapter 4]).

EXAMPLE 3.7 (Hilbert spaces). For $p = q = u = 2$, the regularization scheme (3.1) becomes classical Tikhonov regularization with $W^r_2(\Omega) = B^r_{2,2}(\Omega)$ penalty. Here we obtain

$$\|\hat{f}_\alpha - f\|_{L^2} = \mathcal{O}(\rho^{a/(s+a)}\delta^{s/(s+a)}) \quad \text{if} \quad \|f\|_{B^s_{2,\infty}} \le \rho, \quad s \in (0,r).$$

Due to $W^s_2(\Omega) = B^s_{2,2}(\Omega) \subset B^s_{2,\infty}(\Omega)$, this reproduces the results in [15] and [21].

3.2. Sparsity-promoting wavelet regularization. In the following we explain how regularization by wavelet penalization and in particular weighted ℓ^1-regularization of wavelet coefficients is contained in our setup. The latter is often used since it leads to sparse estimators in the sense that only a finite (and often small) number of wavelet coefficients of $\hat{f}_\alpha$ do not vanish.

We introduce the scale of Besov sequence spaces $b^s_{p,q}$ that allows us to characterize Besov function spaces $B^s_{p,q}(\Omega)$ by decay properties of coefficients in wavelet expansions; see also

[29, Definition 2.6]. Let $c_\Lambda, C_\lambda > 0$ and $(\Lambda_j)_{j\in\mathbb{N}_0}$ be a family of finite sets such that

$$c_\Lambda 2^{jd} \le |\Lambda_j| \le C_\Lambda 2^{jd} \quad \text{for all } j \in \mathbb{N}_0.$$

We consider the index set

$$\Lambda := \{(j,k)\colon j \in \mathbb{N}_0, k \in \Lambda_j\}.$$

For a sequence $x = (x_{j,k})_{(j,k)\in\Lambda}$ and a fixed $j \in \mathbb{N}_0$, we denote by $x_j := (x_{j,k})_{k\in\Lambda_j} \in \mathbb{R}^{\Lambda_j}$ the projection onto the jth level. For $s \in \mathbb{R}$ and $p, q \in [1,\infty]$ let us introduce

$$b^s_{p,q} := \{x \in \mathbb{R}^\Lambda\colon \|x\|_{s,p,q} < \infty\} \quad \text{with} \quad \|x\|_{s,p,q} := \left\|(2^{js}2^{jd(1/2-1/p)}\|x_j\|_p)_{j\in\mathbb{N}_0}\right\|_q.$$

Suppose $(\psi_\lambda)_{\lambda\in\Lambda}$ is a wavelet system on Ω such that the wavelet synthesis operator

$$\mathcal{S}\colon b^r_{p,q} \to B^r_{p,q}(\Omega) \quad \text{given by} \quad (\mathcal{S}x)(r) = \sum_{\lambda\in\Lambda} x_\lambda \psi_\lambda(r) \quad \text{for } r \in \Omega$$

is a norm isomorphism for r, p, and q the parameters involved in $\mathbb{X}_\mathcal{R} = B^r_{p,q}(\Omega)$. In this case we use the norm $\|f\|_{B^r_{p,q}} := \|\mathcal{S}^{-1}f\|_{r,p,q}$ in (3.1). By transformation rules of argmin under composition with a bijective mapping, the estimators in (3.1) can then be rewritten in the form

$$S_\alpha(g) = \mathcal{S} \operatorname*{argmin}_{x\in\mathcal{S}^{-1}(\tilde{D}_F)} \left[\frac{1}{2\alpha}\|g - F(\mathcal{S}x)\|^2_\mathbb{Y} + \frac{1}{u}\|x\|^u_{r,p,q}\right].$$

This is the more common implementation of wavelet penalization methods. If there exists a wavelet analysis operator

$$\mathcal{A}\colon B^r_{p,q}(\Omega) \to b^r_{p,q} \quad \text{satisfying} \quad \|\mathcal{A}\cdot\|_{r,p,q} \sim \|\cdot\|_{B^r_{p,q}},$$

then $\|f\|_{\mathbb{X}_\mathcal{R}} := \|\mathcal{A}\cdot\|_{r,p,q}$ is equivalent to $\|\cdot\|_{B^r_{p,q}}$, and may be used as penalty term in the framework of Corollary 3.2.

EXAMPLE 3.8 ($p = 2$). In the case $p = 2$ we have $(B^{-a}_{2,2}(\Omega), B^r_{2,q}(\Omega))_{\theta_s,\infty} = B^s_{2,\infty}(\Omega)$ (see (B.5)), and Corollary 3.2 shows that

$$\|\hat{f}_\alpha - f\|_{L^2} = \mathcal{O}(\rho^{a/(s+a)}\delta^{s/(s+a)}) \quad \text{if} \quad \|f\|_{B^s_{2,\infty}} \le \rho, \quad s \in (0,r). \tag{3.6}$$

The same convergence rate has been obtained for non-oversmoothing Besov wavelet penalization in [32] for $q = u \ge 2$, and $s \in (0, a/(q-1)]$, and in [17] for $q = u = 1$, and $s \in (0,\infty)$, for infinitely smooth wavelets. In [32] it was shown that this rate is of optimal order.

As a reference example, we discuss rates for piecewise smooth univariate functions with jumps. As shown in [20, Example 30], such functions belong to $B^s_{p,\infty}$ if and only if $s \le 1/p$, and to $B^s_{p,q}$ with $q < \infty$ if and only if $s < 1/p$. Hence, in our setting we have $s = 1/2$ in (3.6).

EXAMPLE 3.9 ($p = q = 1$). Note that for $u = p = q = 1$ we obtain a weighted ℓ^1-penalty. The largest smoothness class $(B^{-a}_{2,2}(\Omega), B^r_{1,1}(\Omega))_{\theta_s,\infty} = \mathcal{S}(k_s)$ was characterized in [20] as the image of a weighted Lorentz sequence space k_s, and a converse result was derived for this class. As this is not a Besov space, we will work with the slightly smaller space $(B^{-a}_{2,2}(\Omega), B^r_{1,1}(\Omega))_{\theta_s,p_s} = B^s_{p_s,p_s}(\Omega) = W^s_{p_s}(\Omega)$ with $p_s = (2a+2r)/(2a+r+s) \in (1, (2a+2r)/(2a+r))$ for simplicity; see (B.6), Proposition B.1. Hence, Corollary 3.2 implies that

$$\|\hat{f}_\alpha - f\|_{L^{\overline{p}}} = \mathcal{O}(\varrho^{a/(s+a)}\delta^{s/(s+a)}) \quad \text{if} \quad \|f\|_{W^s_{p_s}} \sim \|f\|_{B^s_{p_s,p_s}} \le \rho, \quad s \in (0,r)$$

for $\overline{p} = (2a + 2r)/(2a + r)$. This re-proves results that were derived in [20] using hard thresholding approximations of the true solution.

For piecewise smooth functions with jumps, the condition $s < 1/p_s$ is equivalent to $s < (2a + r)/(2a + 2r - 1)$, and the right-hand side is always larger than $1/2$. Therefore, we obtain a faster rate for $p = 1$ than for $p = 2$ although only in the $L^{\overline{p}}$- rather than the L^2-norm.

EXAMPLE 3.10 ($p < 1$). For $p = q = u < 1$ we obtain a weighted ℓ^p-penalty. In analogy to Example 3.9 we use the smoothness class $(B^{-a}_{2,2}(\Omega), B^r_{p,p}(\Omega))_{\theta_s,\tilde{p}_s} = B^s_{\tilde{p}_s,\tilde{p}_s}(\Omega) = W^s_{\tilde{p}_s}(\Omega)$ with $\tilde{p}_s = 2p(a+r)/(2a+pr+(2-p)s) \in (p, 2p(a+r)/(2a+pr))$ and find that

$$\|\hat{f}_\alpha - f\|_{L^{\overline{p}}} = \mathcal{O}(\rho^{a/(s+a)}\delta^{s/(s+a)}) \quad \text{if} \quad \|f\|_{W^s_{\tilde{p}_s}} \sim \|f\|_{B^s_{\tilde{p}_s,\tilde{p}_s}} \leq \rho, \quad s \in (0, r)$$

for $\overline{p} = 2p(a + r)/(2a + pr)$.

For piecewise smooth functions with jumps, the condition $s < 1/\tilde{p}_s$ is equivalent to $s < (2a + r)/(p(2a + 2r + 1) - 2)$, where the denominator is positive due to the first part of Assumption 3.1. Hence choosing $p < 1$ rather than $p = 1$ pays off in the sense that we obtain an even higher rate of convergence, but also in an even weaker norm.

REMARK 3.11 (Does oversmoothing harm?). To conclude this section, we point out a difference in the previous three examples. For $p = 2$ our convergence rate analysis yields the same convergence rate $\mathcal{O}(\delta^{s/(s+a)})$ measured in the same norm, the L^2-norm, under the same smoothness condition given by $B^s_{2,\infty}(\Omega)$ as in the case $r = 0$. Hence the paradigm "Oversmoothing does not harm" known for Hilbert space regularization remains true for $B^r_{2,q}(\Omega)$ Banach space penalties with $q < 2$.

In contrast, in Examples 3.9 and 3.10, a higher value of r may cause an assignment of a lower smoothness s to a fixed true solution. On the other hand, the error is then measured in a stronger norm. This indicates that the integrability index in the loss function norm may have an influence on the convergence rate. It calls for the development of a convergence rate theory that is more flexible in the choice of the loss function and allows for norms that cannot be sharply bounded by powers of the norms of the spaces $\mathbb{X}_-$ and $\mathbb{X}_{\mathcal{R}}$ in Assumption 2.1 via interpolations.

4. Bounded variation regularization. This section contains an application of Theorem 2.4 to Tikhonov regularization with penalty term given by the BV-norm. Let $d \in \mathbb{N}$ and $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. A function $f \in L^1(\Omega)$ has *bounded variation* if

$$|f|_{\mathrm{BV}(\Omega)} := \sup\left\{\int_\Omega f(x)\,\mathrm{div}\,g(x)\,\mathrm{d}x \colon g \in C^1_c(\Omega, \mathbb{R}^d),\ \|g\|_{L^\infty(\Omega,\mathbb{R}^d)} \leq 1\right\} < \infty.$$

Here

$$\|g\|_{L^\infty(\Omega,\mathbb{R}^d)} := \left\|\left(\sum_{j=1}^d g_i^2\right)^{1/2}\right\|_{L^\infty(\Omega)} \quad \text{with} \quad g = (g_1, \ldots g_n).$$

Then

$$\mathrm{BV}(\Omega) := \{f \in L^1(\Omega) \colon |f|_{\mathrm{BV}(\Omega)} < \infty\}$$

is a Banach space equipped with $\|\cdot\|_{\mathrm{BV}(\Omega)} := \|\cdot\|_{L^1(\Omega)} + |\cdot|_{\mathrm{BV}(\Omega)}$. We refer to [2] for a detailed study of spaces of bounded variation.

For $a \geq 0$ with $a \geq d/2 - 1$ there is a continuous embedding $\mathrm{BV}(\Omega) \subset B^{-a}_{2,2}(\Omega)$; see Proposition C.1. In this section we will use the following assumption on the forward operator.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

ASSUMPTION 4.1. *Let $a \geq 0$ with $a \geq d/2-1$. Let $\tilde{D}_F \subset \mathrm{BV}(\Omega)$, $\mathbb{Y}$ be a Banach space, and $F\colon \tilde{D}_F \to \mathbb{Y}$ be a map satisfying*

$$\frac{1}{M_1}\|f_1-f_2\|_{B^{-a}_{2,2}} \leq \|F(f_1)-F(f_2)\|_{\mathbb{Y}} \leq M_2\|f_1-f_2\|_{B^{-a}_{2,2}} \quad \textit{for all } f_1,f_2 \in \tilde{D}_F$$

for constants $M_1, M_2 > 0$.

For $g \in \mathbb{Y}$ we consider

$$S_\alpha(g) = \operatorname*{argmin}_{h\in\tilde{D}_F}\left[\frac{1}{2\alpha}\|g-F(h)\|^2_{\mathbb{Y}} + \|h\|_{\mathrm{BV}}\right].$$

We refer to [1] for this kind of regularization scheme for linear operators, including a proof of existence of minimizers, and to [7] for a treatment of similar estimators in a statistical setting.

Let $a \geq 0$ and $s \in (-a,1)$. The following interpolation identity, based on the result by Cohen et al. in [6, Theorem 1.4], is a crucial ingredient for our convergence rates result,

$$B^s_{t_s,t_s}(\Omega) = (B^{-a}_{2,2}(\Omega), \mathrm{BV}(\Omega))_{\theta_s,t_s} \quad \text{with} \quad \theta_s := \frac{s+a}{a+1} \text{ and } t_s := \frac{2a+2}{s+2a+1} \tag{4.1}$$

with equivalent norms. In the latter reference, the authors show this identity for $\Omega = \mathbb{R}^d$ and from there we conclude the statement in Proposition C.2.

To avoid the abstract smoothness condition in Theorem 2.4, we state our theorem under a slightly stronger smoothness assumption and comment on the weaker condition in a remark afterwards. Again, we do not state bounds on the bias for the sake of brevity.

COROLLARY 4.2 (Convergence rates for BV-regularization). *Suppose Assumption 4.1 holds true, and the true solution f has smoothness*

$$f \in B^s_{t_s,t_s}(\Omega) \quad \textit{with } \|f\|_{B^s_{t_s,t_s}} < \varrho$$

for some $0 < s < 1$ and $\varrho > 0$, or

$$f \in \mathrm{BV}(\Omega) \quad \textit{with } \|f\|_{\mathrm{BV}} < \varrho.$$

In the latter case, we set $s=1$. Set $\overline{p} = (2a+2)/(2a+1)$ and suppose that the closure D_F of $\tilde{D}_F$ in $B^{-a}_{2,2}(\Omega)$ contains an $L^{\overline{p}}(\Omega)$-ball with radius τ around f. Suppose that $g^{\mathrm{obs}} \in \mathbb{Y}$ satisfies (1.1) for $0 < \delta < \varrho^{-a/s}\tau^{(s+a)/s}$, and let $\hat{f}_\alpha \in S_\alpha(g^{\mathrm{obs}})$ for some $\alpha > 0$. Let $0 < c_l \leq c_r$ and $1 < c_{\mathrm{D}} \leq C_{\mathrm{D}}$. Then there is a constant C_r independent of f, g^{obs}, ϱ, τ, and δ such that either of the conditions

$$c_l\varrho^{-(a+1)/(s+a)}\delta^{(s+2a+1)/(s+a)} \leq \alpha \leq c_r\varrho^{-(a+1)/(s+a)}\delta^{(s+2a+1)/(s+a)}$$

and

$$c_{\mathrm{D}}\delta \leq \|g^{\mathrm{obs}} - F(\hat{f}_\alpha)\|_{\mathbb{Y}} \leq C_{\mathrm{D}}\delta$$

on the choice of α implies the following bounds:

$$\begin{aligned}
\|f-\hat{f}_\alpha\|_{B^{-a}_{2,2}} &\leq C_r\delta,\\
\|f-\hat{f}_\alpha\|_{L^{\overline{p}}(\Omega)} &\leq C_r\varrho^{a/(s+a)}\delta^{s/(s+a)},\\
\|\hat{f}_\alpha\|_{\mathrm{BV}(\Omega)} &\leq C_r\varrho^{(a+1)/(s+a)}\delta^{(s-1)/(s+a)}.
\end{aligned}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Proof. We show that Assumption 2.1 is satisfied with $\mathbb{X}_{\mathcal{R}} = \mathrm{BV}(\Omega)$, $\mathbb{X}_{-} = B^{-a}_{2,2}(\Omega)$, $\mathbb{X}_{\mathrm{L}} = L^{\overline{p}}(\Omega)$, and $u = 1$. Due to Proposition C.1, we have a continuous embedding $\mathrm{BV}(\Omega) \subset B^{-a}_{2,2}(\Omega)$. If $a = 0$, then we have $\overline{p} = 2$ and $B^{0}_{2,2}(\Omega) = L^2(\Omega)$; see Proposition B.1. Hence we have Assumption 2.1 with $\xi = 1$ in this case and the result follows from Theorem 2.4.

If $a > 0$, we set $\xi := a/(a+1)$. Note that $1 < \overline{p} = t_0 < 2$. Hence [3, Theorem 3.4.1(b)], (4.1), and Proposition B.1 yield the following chain of continuous embeddings:

$$(B^{-a}_{2,2}(\Omega), \mathrm{BV}(\Omega))_{\xi,1} \subset (B^{-a}_{2,2}(\Omega), \mathrm{BV}(\Omega))_{\xi,\overline{p}} = B^{0}_{\overline{p},\overline{p}}(\Omega) \subset L^{\overline{p}}(\mathbb{R}^d).$$

Finally, [3, Theorem 3.4.1(b)] and (4.1) yield

$$B^{s}_{t_s,t_s}(\Omega) = (B^{-a}_{2,2}(\Omega), \mathrm{BV}(\Omega))_{\theta_s,t_s} \subset (B^{-a}_{2,2}(\Omega), \mathrm{BV}(\Omega))_{\theta_s,\infty}.$$

Hence the smoothness condition on f in the claim implies the smoothness condition in Theorem 2.4. Therefore, the stated result follows from Theorem 2.4. ∎

REMARK 4.3 (Weaker smoothness condition). The statements in Corollary 4.2 remain true if the smoothness assumption on f is replaced by $f \in (B^{-a}_{2,2}(\Omega), \mathrm{BV}(\Omega))_{\theta_s,\infty}$ with a bound by ϱ on the norm of f therein.

REMARK 4.4 (Similarity to $B^{1}_{1,1}(\Omega)$-regularization). We see that the convergence rates and also the smoothness condition for BV-regularization equal those for $B^{1}_{1,1}(\Omega)$-regularization in Corollary 3.2. The reason for this is that the interpolation identity in (4.1) holds true with $\mathrm{BV}(\Omega)$ replaced by $B^{1}_{1,1}(\Omega)$.

Whereas for $B^{1}_{1,1}(\Omega)$-regularization with a norm given by wavelet coefficients, we also have a convergence rate result in the non-oversmoothing case $s > r$ (see [20]), a similar result remains open for BV-regularization.

5. White noise.

In this section we extend the tools developed in the previous sections to derive convergence rates for oversmoothing regularization with stochastic noise models.

We will assume that $\Omega_{\mathbb{Y}} \subset \mathbb{R}^d$ is a bounded Lipschitz domain and $\mathbb{Y} = L^2(\Omega_{\mathbb{Y}})$. We consider noise models of the form

$$g^{\mathrm{obs}} = F(f) + \sigma Z, \quad Z \in B^{-d/2}_{p',\infty}(\Omega_{\mathbb{Y}}), \tag{5.1}$$

with a normalized noise process Z and a noise level $\sigma > 0$. Moreover, $p \in (1, \infty]$ is the same as in Section 3, and $1/p' + 1/p = 1$. The choice of the Besov space is motivated by the fact that Gaussian white noise belongs to $B^{-d/2}_{p',\infty}$ almost surely (see [31] for the d-dimensional torus), and to no smaller Besov spaces. Also point processes (i.e., random finite sums of delta-peaks) belong to $B^{-d/2}_{p',\infty}$ for $p \geq 2$ as well as local averages of noise processes over a finite number of detector areas. We will derive error bounds in terms of Besov norms of Z. The expectation of Z does not necessarily have to vanish, i.e., Z may also contain deterministic error components. However, to derive error bounds in expectation, we will have to assume that the norm of Z has finite moments:

$$\mathbb{E}\left[\|Z\|^{\kappa}_{B^{-d/2}_{p',\infty}}\right] < \infty \quad \text{for all } \kappa \in \mathbb{N}. \tag{5.2}$$

This easily follows from much stronger large-deviation inequalities (see, e.g., the proof of [17, Corollary 6.5]), which have been shown for Gaussian white noise in [31, Corollary 3.7] or [14, remark after Theorem 4.4.3]. For other noise processes, the verification of (5.2) may require further investigations.

Since the Tikhonov functional in (1.2) is not well defined in our setting, we formally subtract $\frac{1}{2}\|g^{\mathrm{obs}}\|^2_{\mathbb{Y}}$ from $\frac{1}{2}\|g^{\mathrm{obs}} - g\|^2_{\mathbb{Y}}$ to obtain the new data fidelity functional $\mathcal{S}_{g^{\mathrm{obs}}}(g) :=$

$\frac{1}{2}\|g\|_{\mathbb{Y}}^2 - \langle g^{\mathrm{obs}}, g\rangle$ and Tikhonov regularization of the form

$$T_\alpha(g^{\mathrm{obs}}) := \operatorname*{argmin}_{h\in\tilde{D}_F}\left[\frac{1}{\alpha}\mathcal{S}_{g^{\mathrm{obs}}}(F(h)) + \frac{1}{u}\|h\|_{\mathbb{X}_{\mathcal{R}}}^u\right], \tag{5.3}$$

with $u \in (0,\infty)$. Note that, for $g^{\mathrm{obs}} \in \mathbb{Y}$, we have $T_\alpha(g^{\mathrm{obs}}) = S_\alpha(g^{\mathrm{obs}})$, but $T_\alpha(g^{\mathrm{obs}})$ is also well defined for white noise. More precisely, in the setting of the following Theorem 5.1, the existence of minimizers in (5.3) can be shown by the same argument as in the non-oversmoothing case; see [17, Proposition 6.3].

5.1. Convergence rates. We first study Besov penalties with $p > 1$.

THEOREM 5.1 (Stochastic rates for oversmoothing Besov space regularization). *Let $1 < p \le 2$, $1 \le q \le p$, and $r > 0$. Let the data g^{obs} be described by (5.1), consider Tikhonov regularization in the form (5.3), and assume that $\|\cdot\|_{\mathbb{X}_{\mathcal{R}}}$ in (5.3) is equivalent to $\|\cdot\|_{B^r_{p,q}}$. Suppose the true solution $f \in \tilde{D}_F$ has regularity $s \in (0, r]$ with norm bound $\varrho > 0$ in the sense of (3.2) in Corollary 3.2 or (3.5) in Remark 3.4. In addition to Assumption 3.1, suppose that F satisfies the one-sided Lipschitz condition,*

$$\|F(f_1) - F(f_2)\|_{B^{a+r}_{p,q}(\Omega_{\mathbb{Y}})} \le \widetilde{M}_2\|f_1 - f_2\|_{B^r_{p,q}(\Omega)} \quad \text{for all } f_1, f_2 \in \tilde{D}_F, \tag{5.4}$$

and that $d/2 < a + r$. Let $\overline{p} := (2p(a + r))/(2a + pr)$ and assume that the closure D_F of $\tilde{D}_F$ in $B^{-a}_{2,2}(\Omega)$ contains a $B^0_{\overline{p},\overline{p}}(\Omega)$-ball with radius $\tau > 0$ around f.

Then there is an a priori *parameter choice rule $\alpha = \alpha(\sigma, \varrho)$ (specified in (5.12)) such that there exists a constant $C_r > 0$ such that the reconstruction error with $\hat{f}_\alpha \in T_\alpha(g^{\mathrm{obs}})$ satisfies the bounds*

$$\|f - \hat{f}_\alpha\|_{B^{-a}_{2,2}} \le C_r(1 + N^\eta)\rho^{(d/2)/(s+a+d/2)}\sigma^{(s+a)/(s+a+d/2)}, \tag{5.5a}$$

$$\|f - \hat{f}_\alpha\|_{L^{\overline{p}}} \le C_r(1 + N^\eta)\varrho^{(a+d/2)/(s+a+d/2)}\sigma^{(s)/(s+a+d/2)}, \tag{5.5b}$$

$$\|\hat{f}_\alpha\|_{\mathbb{X}_{\mathcal{R}}} \le C_r(1 + N^\eta)\varrho^{(r+a+d/2)/(s+a+d/2)}\sigma^{(s-r)/(s+a+d/2)}, \tag{5.5c}$$

for all $0 < \sigma < \varrho^{-(a+d/2)/s}\tau^{(s+a+d/2)/s}$ with $N := \|Z\|_{B^{-d/2}_{p',\infty}}$ and

$$\eta := \frac{(a + r)u}{(a + r + d/2)u/2 - d/2}.$$

In particular, if (5.2) holds true, then

$$\mathbb{E}[\|f - \hat{f}_\alpha\|^\kappa_{L^{\overline{p}}}]^{1/\kappa} = \mathcal{O}(\varrho^{(a+d/2)/(s+a+d/2)}\sigma^{s/(s+a+d/2)}) \quad \text{as } \sigma \to 0 \text{ for all } \kappa \ge 1. \tag{5.6}$$

Proof. As in Section 3 set $\mathbb{X}_- = B^{-a}_{2,2}(\Omega)$. If $a > 0$ then we have $(\mathbb{X}_-, \mathbb{X}_{\mathcal{R}})_{a/(a+r),\overline{p}} \subset B^0_{\overline{p},\overline{p}}(\Omega) \subset L^{\overline{p}}(\Omega) = \mathbb{X}_{\mathrm{L}}$ with continuous embeddings due to $p \le 2$ and $q \le p$; see (3.3) and Remark 3.3. If $a = 0$, then $\mathbb{X}_- = \mathbb{X}_{\mathrm{L}} = L^2(\Omega)$. We choose

$$t = C_{\mathrm{L}}{}^{-(a+r)/s}(\sigma/\varrho)^{(a+r)/(s+a+d/2)},$$

and from Lemma 2.2 with $\theta = (a + s)/(a + r)$ we obtain

$$\|f - f_t\|_{\mathbb{X}_{\mathrm{L}}} \le \varrho^{(a+d/2)/(s+a+d/2)}\sigma^{s/(s+a+d/2)} < \tau. \tag{5.7}$$

Hence $f_t \in D_F$, and by definition $\hat{f}_\alpha \in T_\alpha(g^{\mathrm{obs}})$ implies

$$\frac{1}{\alpha}\mathcal{S}_{g^{\mathrm{obs}}}(\hat{g}_\alpha) + \frac{1}{u}\|\hat{f}_\alpha\|^u_{\mathbb{X}_{\mathcal{R}}} \le \frac{1}{\alpha}\mathcal{S}_{g^{\mathrm{obs}}}(g_t) + \frac{1}{u}\|f_t\|^u_{\mathbb{X}_{\mathcal{R}}},$$

with $g_t := F(f_t)$ and $\hat{g}_\alpha := F(\hat{f}_\alpha)$. Adding $(1/(2\alpha))\|g_t - \hat{g}_\alpha\|_{\mathbb{Y}}^2 - (1/\alpha)\mathcal{S}_{g^{\mathrm{obs}}}(\hat{g}_\alpha)$ to this equation yields

$$\begin{aligned} &\frac{1}{2\alpha}\|g_t - \hat{g}_\alpha\|_{\mathbb{Y}}^2 + \frac{1}{u}\|\hat{f}_\alpha\|_{\mathbb{X}_{\mathcal{R}}}^u \\ &\le \frac{1}{2\alpha}\|g_t - \hat{g}_\alpha\|_{\mathbb{Y}}^2 + \frac{1}{\alpha}\mathcal{S}_{g^{\mathrm{obs}}}(g_t) - \frac{1}{\alpha}\mathcal{S}_{g^{\mathrm{obs}}}(\hat{g}_\alpha) + \frac{1}{u}\|f_t\|_{\mathbb{X}_{\mathcal{R}}}^u \\ &= \frac{1}{\alpha}\langle \sigma Z, \hat{g}_\alpha - g_t\rangle + \frac{1}{\alpha}\langle F(f) - g_t, \hat{g}_\alpha - g_t\rangle + \frac{1}{u}\|f_t\|_{\mathbb{X}_{\mathcal{R}}}^u. \end{aligned} \tag{5.8}$$

The first term on the left-hand side is estimated using the Besov space interpolation

$$(B_{p,2}^0(\Omega_{\mathbb{Y}}), B_{p,q}^{a+r}(\Omega_{\mathbb{Y}}))_{d/(2a+2r),1} = B_{p,1}^{d/2}(\Omega_{\mathbb{Y}}), \tag{5.9}$$

the Lipschitz condition (5.4), and the continuity of the embedding $B_{2,2}^0(\Omega_{\mathbb{Y}}) = \mathbb{Y} \hookrightarrow B_{p,2}^0(\Omega_{\mathbb{Y}})$:

$$\begin{aligned} \frac{1}{\alpha}\langle \sigma Z, \hat{g}_\alpha - g_t\rangle &\le \frac{1}{\alpha}\|\sigma Z\|_{B_{p',\infty}^{-d/2}}\|g_t - \hat{g}_\alpha\|_{B_{p,1}^{d/2}} \\ &\le c_1\frac{\sigma N}{\alpha}\|g_t - \hat{g}_\alpha\|_{B_{p,2}^0}^{1-d/(2a+2r)}\|g_t - \hat{g}_\alpha\|_{B_{p,q}^{a+r}}^{d/(2a+2r)} \\ &\le c_2\frac{\sigma N}{\alpha}\|g_t - \hat{g}_\alpha\|_{\mathbb{Y}}^{1-d/(2a+2r)}\|f_t - \hat{f}_\alpha\|_{\mathbb{X}_{\mathcal{R}}}^{d/(2a+2r)} \\ &= (c_2\sigma N\alpha^{-(a+r+d/2)/(2a+2r)})\left(\frac{1}{\alpha}\|g_t - \hat{g}_\alpha\|_{\mathbb{Y}}^2\right)^{(a+r-d/2)/(2a+2r)} \\ &\quad \times (\|f_t - \hat{f}_\alpha\|_{\mathbb{X}_{\mathcal{R}}}^u)^{d/(u(2a+2r))}, \end{aligned} \tag{5.10}$$

with c_2 depending on c_1, the embedding constant, $\widetilde{M}_2$, and the constant in the equivalence of $\|\cdot\|_{B_{p,q}^r}$ and $\|\cdot\|_{\mathbb{X}_{\mathcal{R}}}$. Now Young's inequality $xyz \le (1/\theta)x^\theta + (1/\mu)y^\mu + (1/\nu)z^\nu$ for $1/\theta + 1/\mu + 1/\nu = 1$ with

$$\eta = \frac{u(2a+2r)}{(a+r+d/2)u - d}, \quad \mu := \frac{2a+2r}{a+r-d/2}, \quad \text{and} \quad \nu := \frac{u(2a+2r)}{d},$$

and the elementary inequality $(x+y)^u \le 2^{u-1}(x^u + y^u)$ yield

$$\begin{aligned} &\frac{1}{\alpha}\langle \sigma Z, \hat{g}_\alpha - g_t\rangle \\ &\le c_3\left(\sigma N\alpha^{-(a+r+d/2)/(2a+2r)}\right)^\eta + \frac{1}{8\alpha}\|g_t - \hat{g}_\alpha\|_{\mathbb{Y}}^2 + \frac{1}{2u}2^{1-u}\|f_t - \hat{f}_\alpha\|_{\mathbb{X}_{\mathcal{R}}}^u \\ &\le c_3\left(\sigma N\alpha^{-(a+r+d/2)/(2a+2r)}\right)^\eta + \frac{1}{8\alpha}\|g_t - \hat{g}_\alpha\|_{\mathbb{Y}}^2 + \frac{1}{2u}\|\hat{f}_\alpha\|_{\mathbb{X}_{\mathcal{R}}}^u + \frac{1}{2u}\|f_t\|_{\mathbb{X}_{\mathcal{R}}}^u, \end{aligned}$$

with a constant c_3 that depends on c_2, u, η, μ, and ν. The second and third summands on the right-hand side can be absorbed in the left-hand side of (5.8). The second term on the right-hand side (5.8) is estimated by

$$\begin{aligned} \frac{1}{\alpha}\langle F(f) - g_t, \hat{g}_\alpha - g_t\rangle &\le \frac{1}{\alpha}\|F(f) - g_t\|_{\mathbb{Y}}\|\hat{g}_\alpha - g_t\|_{\mathbb{Y}} \\ &\le \frac{M_2}{\alpha}\|f - f_t\|_{\mathbb{X}_-}\|\hat{g}_\alpha - g_t\|_{\mathbb{Y}} \\ &\le \frac{4M_2^2}{\alpha}\|f - f_t\|_{\mathbb{X}_-}^2 + \frac{1}{8\alpha}\|\hat{g}_\alpha - g_t\|_{\mathbb{Y}}^2, \end{aligned}$$

and the second term can be absorbed in the left-hand side of (5.8). Altogether we have shown that

$$\begin{aligned}
&\frac{1}{4\alpha}\|g_t-\hat g_\alpha\|^2_{\mathbb{Y}}+\frac{1}{2u}\|\hat f_\alpha\|^u_{\mathbb{X}_\mathcal{R}}\\
&\quad\le c_3(\sigma N\alpha^{-(a+r+d/2)/(2a+2r)})^\eta+\frac{3}{2u}\|f_t\|^u_{\mathbb{X}_\mathcal{R}}+\frac{4M_2^2}{\alpha}\|f-f_t\|^2_{\mathbb{X}_-}\\
&\quad\le c_3(\sigma N\alpha^{-(a+r+d/2)/(2a+2r)})^\eta \qquad (5.11)\\
&\qquad+c_4\varrho^{2u(a+r)/((2-u)s+2a+ur)}\alpha^{u(s-r)/((2-u)s+2a+ur)}\\
&\quad\le(c_3+c_4)(1+N^\eta)\varrho^{u(a+r+d/2)/(s+a+d/2)}\sigma^{u(s-r)/(s+a+d/2)}
\end{aligned}$$

using Lemma 2.2, the choice of t, and the parameter choice rule

$$\alpha=c_\alpha\varrho^{-(u(a+r+d/2)-d)/(s+a+d/2)}\sigma^{((2-u)s+2a+ur)/(s+a+d/2)} \qquad (5.12)$$

for α with a constant c_α. Here the constant c_4 depends on M_2, C_L, u, a, s, r, and c_α.

This shows on the one hand that

$$\|f_t-\hat f_\alpha\|_{\mathbb{X}_-}\le M_1\|g_t-\hat g_\alpha\|_{\mathbb{Y}}\le c_5(1+N^\eta)\varrho^{(d/2)/(s+a+d/2)}\sigma^{(s+a)/(s+a+d/2)},$$

with c_5 depending on M_1, c_3, and c_4, where we use the choice of α once again. This finishes the proof if $a=0$. On the other hand, (5.11) and Lemma 2.2 imply that

$$\|f_t-\hat f_\alpha\|_{\mathbb{X}_R}\le\|f_t\|_{\mathbb{X}_R}+\|\hat f_\alpha\|_{\mathbb{X}_R}\le c_6(1+N^\eta)\varrho^{(a+r+d/2)/(s+a+d/2)}\sigma^{(s-r)/(s+a+d/2)},$$

with c_6 depending on C_L, u, c_3, and c_4. Putting both estimates together and using the interpolation and embedding results from the very beginning of this proof, we obtain

$$\begin{aligned}
\|f_t-\hat f_\alpha\|_{\mathbb{X}_\mathrm{L}}&\le\|f_t-\hat f_\alpha\|^{r/(a+r)}_{\mathbb{X}_-}\|f_t-\hat f_\alpha\|^{a/(a+r)}_{\mathbb{X}_R}\\
&\le c_7(1+N^\eta)\varrho^{(a+d/2)/(a+s+d/2)}\sigma^{s/(a+s+d/2)},
\end{aligned}$$

with c_7 depending on c_5 and c_6. Together with (5.7) we wind up with

$$\begin{aligned}
\|f-\hat f_\alpha\|_{\mathbb{X}_\mathrm{L}}&\le\|f-f_t\|_{\mathbb{X}_\mathrm{L}}+\|f_t-\hat f_\alpha\|_{\mathbb{X}_\mathrm{L}}\\
&\le(c_7+1)(1+N^\eta)\varrho^{(a+d/2)/(a+s+d/2)}\sigma^{s/(s+a+d/2)}. \qquad \square
\end{aligned}$$

In our noise model (5.1) we excluded the case $p=1$, i.e., $p'=\infty$, since $Z\notin B^{-d/2}_{\infty,\infty}$ almost surely; see [31]. However, the interesting case $p=q=1$ can be treated if we impose an additional one-sided Lipschitz condition on F, as follows.

THEOREM 5.2 (Stochastic rates for oversmoothing regularization with BV or $B^r_{1,1}$ penalties). *For the data g^{obs} described by (5.1) with p as defined below, consider Tikhonov regularization of the form (5.3) with either $\|\cdot\|_{\mathbb{X}_\mathcal{R}}=\|\cdot\|_{\mathrm{BV}}$ or $\|\cdot\|_{\mathbb{X}_\mathcal{R}}=\|\cdot\|_{B^r_{1,1}}$ for some $r>\max(0,d/2-a)$. For BV we set $r=1$ and assume that $a>d/2-1$. Suppose the true solution has regularity*

$$f\in B^s_{t_s,t_s}(\Omega)\quad\textit{with }\|f\|_{B^s_{t_s,t_s}}<\varrho$$

for $s\in(0,r]$ and $t_s=(2a+2r)/(s+2a+r)$. For $\|\cdot\|_{\mathbb{X}_\mathcal{R}}=\|\cdot\|_{\mathrm{BV}}$ and $s=1$ we assume $f\in\mathrm{BV}(\Omega)$ and $\|f\|_{\mathrm{BV}}\le\rho$. In addition to Assumption 4.1 or Assumption 3.1, respectively,

ETNA
Kent State University and
Johann Radon Institute (RICAM)

suppose that there exists $e \in (0, a + r - d/2)$ such that F satisfies the one-sided Lipschitz condition,

$$\|F(f_1) - F(f_2)\|_{B^{a+r-e}_{p,p}(\Omega_{\mathbb{Y}})} \le \widetilde{M}_2 \|f_1 - f_2\|_{B^{r-e}_{p,p}(\Omega)} \quad \text{for all } f_1, f_2 \in \widetilde{D}_F, \tag{5.13}$$

with $p := (a + r)/(a + r - e/2)$. Let $\overline{p} := (2a + 2r)/(2a + r)$ and assume that the closure D_F of $\tilde{D}_F$ in $B^{-a}_{2,2}(\Omega)$ contains a $B^0_{\overline{p},\overline{p}}(\Omega)$-ball with radius $\tau > 0$ around f.

Then there is an a priori *parameter choice rule $\alpha = \alpha(\sigma, \varrho)$ (specified in (5.12)) such that there exists a constant $C_r > 0$ such that the reconstruction error with $\hat{f}_\alpha \in T_\alpha(g^{\text{obs}})$ satisfies all three bounds in (5.5) for all σ as in Theorem 5.1 with $N := \|Z\|_{B^{-d/2}_{p',\infty}}$ and η as in Theorem 5.1. Under the assumption in (5.2) we also have (5.6).*

Proof. The proof follows along the lines of the proof of Theorem 5.1, we just have to replace (5.10) as follows. Note that $1 < p < 2$. The starting point is

$$\frac{1}{\alpha}\langle \sigma Z, \hat{g}_\alpha - g_t\rangle \le \frac{1}{\alpha}\|\sigma Z\|_{B^{-d/2}_{p',\infty}} \|g_t - \hat{g}_\alpha\|_{B^{d/2}_{p,1}} = \frac{\sigma N}{\alpha}\|g_t - \hat{g}_\alpha\|_{B^{d/2}_{p,1}}. \tag{5.14}$$

We replace (5.9) by

$$(B^0_{p,2}(\Omega_{\mathbb{Y}}), B^{a+r-e}_{p,p}(\Omega_{\mathbb{Y}}))_{d/(2a+2r-2e),1} = B^{d/2}_{p,1}(\Omega_{\mathbb{Y}}),$$

and use the continuity of the embedding $B^0_{2,2}(\Omega_{\mathbb{Y}}) = \mathbb{Y} \hookrightarrow B^0_{t,2}(\Omega_{\mathbb{Y}})$ and (5.13) to obtain

$$\begin{aligned} \|g_t - \hat{g}_\alpha\|_{B^{d/2}_{p,1}} &\le c_1 \|g_t - \hat{g}_\alpha\|^{1-d/(2a+2r-2e)}_{B^0_{p,2}} \|g_t - \hat{g}_\alpha\|^{d/(2a+2r-2e)}_{B^{a+r-e}_{p,p}} \\ &\le c_2 \|g_t - \hat{g}_\alpha\|^{1-d/(2a+2r-2e)}_{\mathbb{Y}} \|f_t - \hat{f}_\alpha\|^{d/(2a+2r-2e)}_{B^{r-e}_{p,p}}, \end{aligned} \tag{5.15}$$

with c_2 depending on c_1, the embedding constant, and $\tilde{M}_2$. To estimate the second factor on the right-hand side, we use the interpolation identity

$$B^{r-e}_{p,p}(\Omega) = (B^{-a}_{2,2}(\Omega), \mathbb{X}_{\mathcal{R}})_{(a+r-e)/(a+r),p}$$

(note that $-a < r - e$), which follows from Proposition C.2 or from [30, 2.4.3], respectively. Together with Assumption 4.1, respectively Assumption 3.1, we obtain

$$\begin{aligned} \|f_t - \hat{f}_\alpha\|_{B^{r-e}_{p,p}} &\le c_3 \|f_t - \hat{f}_\alpha\|^{e/(a+r)}_{B^{-a}_{2,2}} \|f_t - \hat{f}_\alpha\|^{(a+r-e)/(a+r)}_{\mathbb{X}_{\mathcal{R}}} \\ &\le c_4 \|g_t - \hat{g}_\alpha\|^{e/(a+r)}_{\mathbb{Y}} \|f_t - \hat{f}_\alpha\|^{(a+r-e)/(a+r)}_{\mathbb{X}_{\mathcal{R}}}, \end{aligned}$$

with c_4 depending on c_3 and M_1. Inserting into (5.15) and then into (5.14) yields the inequality

$$\frac{1}{\alpha}\langle \sigma Z, \hat{g}_\alpha - g_t\rangle \le c_5 \frac{\sigma N}{\alpha} \|g_t - \hat{g}_\alpha\|^{1-d/(2a+2r)}_{\mathbb{Y}} \|f_t - \hat{f}_\alpha\|^{d/(2a+2r)}_{\mathbb{X}_{\mathcal{R}}},$$

which replaces (5.10). Here c_5 depends on c_4 and c_2. The rest of the proof can be copied from the proof of Theorem 5.1. ∎

REMARK 5.3 (Minimax optimality). It can be shown as in [17, Proposition 6.6] that the error bound in (5.6) is optimal in a minimax sense.

REMARK 5.4 (Duality). In view of the fact that the dual of Besov spaces $B^{-s}_{p',q'}(\Omega)$ for $s \in \mathbb{R}$ and $p, q \in (1, \infty)$ on smooth, bounded domains Ω is given by the spaces $\tilde{B}^s_{p,q}(\Omega) := \{f \in B^s_{p,q}(\mathbb{R}^d) : \operatorname{supp} f \subset \overline{\Omega}\}$ (see [27, Theorem 4.8.2]), it may appear more natural to impose the assumptions in (5.4) and (5.13) in these spaces. (Note that, if $\Omega_{\mathbb{Y}} = \Omega$ and F is linear and

ETNA
Kent State University and
Johann Radon Institute (RICAM)

self-adjoint on $L^2(\Omega)$ with a bijective continuous extension to $B^{-a}_{2,2}(\Omega) \to L^2(\Omega)$ such that Assumption 3.1 holds true, then $F : L^2(\Omega) \to \tilde{B}^a_{2,2}(\Omega)$ is also bijective by duality.) However, the spaces $\tilde{B}^s_{p,q}(\Omega)$ are closed subspaces of $B^s_{p,q}(\Omega)$, which can be written as nullspaces of certain trace operators, except for smoothness indices s with $s - 1/p \in \mathbb{N}_0$ at which the number of well-defined traces changes; see [27, Theorems 4.3.2/1 and 4.7.1]. Therefore, the given formulations of (5.4) and (5.13) are more general, and boundary conditions can be incorporated in the domain $\tilde{D}_F$ of F.

REMARK 5.5 (Special case $f \in \mathrm{BV}$). Theorem 5.2 for $\|\cdot\|_{\mathbb{X}_\mathcal{R}} = \|\cdot\|_{\mathrm{BV}}$ and $f \in \mathrm{BV}(\Omega)$ improves the rate in [9] obtained for a different estimator by eliminating logarithmic factors in the noise level. Furthermore, we do not need to assume the existence of a wavelet–vaguelette decomposition of the forward operator.

REMARK 5.6 (Implications for regression). Our setting includes the case $F = I$ corresponding to regression problems. We discuss two particular cases.

- If we choose Besov wavelet norms with $p = q = 1$ as in Section 3.2, then the minimization of the Tikhonov functional splits into a family of minimization problems for each wavelet coefficient, resulting in *soft thresholding* or *wavelet shrinkage* estimators with level-dependent threshold. Such estimators have been studied extensively in mathematical statistics; see, e.g., [10].
- For $\|\cdot\|_{\mathbb{X}_\mathcal{R}} = \|\cdot\|_{\mathrm{BV}}$ we obtain BV-denoising. Here our assumption $a \geq d/2 - 1$ is only satisfied for $d = 1$. In this case Theorem 5.2 shows optimal $L^{\overline{p}}$-convergence rates of this estimator for functions with Besov smoothness ≤ 1; see also [18]. In higher dimensions, convergence rates of a multiresolution estimator for BV functions were established in [8].

5.2. Numerical experiments for a parameter identification problem. We confirm the theoretical results in Theorems 5.1 and 5.2 by numerical experiments for the nonlinear identification of c in the elliptic boundary value problem

$$
\begin{aligned}
-u'' + cu &= \varphi \quad \text{in } (0,1), \\
u(0) = u(1) &= 1.
\end{aligned}
\tag{5.16}
$$

The forward operator in the function space setting is $F(c) := u$ for the fixed smooth right-hand side φ. For this problem, the verification of Assumption 3.1 with $a = 2$ is discussed in [17, Example 2.8 and Lemma 2.9]. The experiments are carried out in the same setup as in [20], where more details on the implementation can be found. We added independent $N(0, \tilde{\sigma}^2)$-distributed random variables to $n = 2^{10}$ equidistant measurement points as a discrete approximation of Gaussian white noise on $[0,1]$ with $\sigma = \tilde{\sigma}/\sqrt{n}$.

The true coefficient c^{jump} is given by a piecewise smooth function with finitely many jumps. For each noise level $\tilde{\sigma}$, we drew 10 data sets and took the average of the reconstruction errors.

The regularization parameter α was chosen according to the rule (5.12) with c_α chosen optimally for the median value of $\tilde{\sigma}$. Of course, in practice α would have to be chosen in a completely data-driven manner, e.g., by the Lepskiĭ balancing principle, but this is not within the scope of this paper.

EXAMPLE 5.7 ($p = 2, q = 1$). First, we use as penalty the norm (with power $u = 1$) on the Besov space $B^2_{2,1}((0,1))$ given by the $b^2_{2,1}$-norm of wavelet coefficients with respect to Daubechies wavelets of order 7. According to Remark 3.8, the smoothness of the solution c^{jump} is then measured in the scale $B^s_{2,\infty}((0,1))$, and in this scale the maximal smoothness index of c^{jump} is $s = 1/2$, i.e., $c^{\mathrm{jump}} \in B^{1/2}_{2,\infty}((0,1))$; see [17, Example 30]. In Figure 5.1

ETNA
Kent State University and
Johann Radon Institute (RICAM)

we see a good agreement of the reconstruction error in the numerical experiment with the predicted rate $\mathcal{O}(\sigma^{1/6})$ measured in the L^2-norm.

EXAMPLE 5.8 ($p = q = 1$). Now we use the $b^2_{1,1}$-norm on $db7$ wavelet coefficients norm as penalty term. As $a = r = 2$, we have $\overline{p} = 4/3$. As in [17] one shows that c^{jump} belongs to $B^s_{t_s,t_s}((0,1))$ for $s < 6/7$. Therefore, Corollary 3.2 and Remark 3.3 predict the rate $\mathcal{O}(\sigma^e)$ for all $e < 12/47$ measured in the $L^{4/3}$-norm. In Figure 5.1 we see a good agreement with the reconstruction error in the numerical experiment.

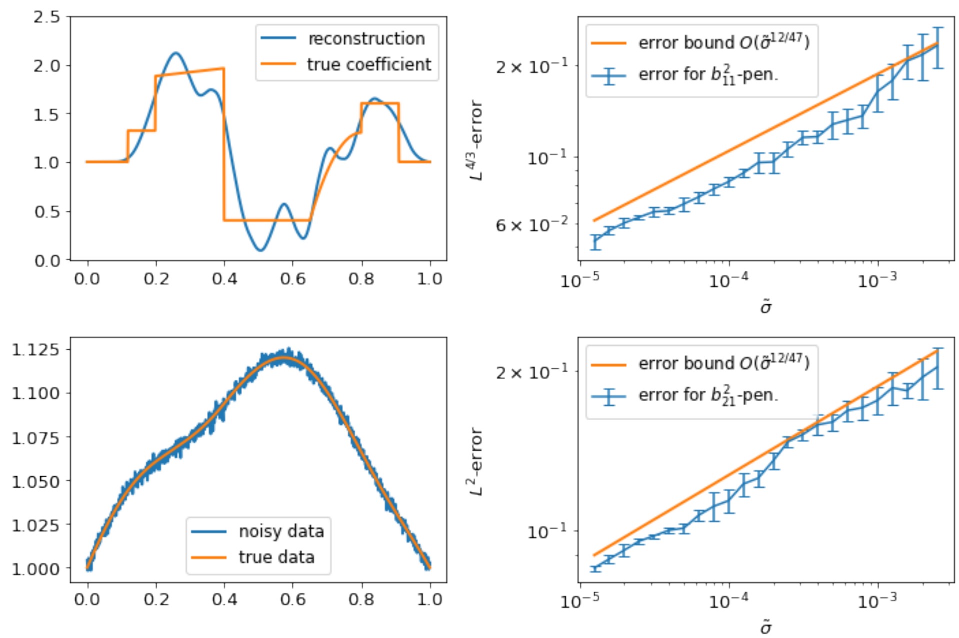

FIG. 5.1. *Top left: The true coefficient c^{jump} with jumps in the boundary value problem (5.16) together with the reconstruction for $b^2_{1,1}$-penalization at noise level $\tilde{\sigma} = 2.51 \times 10^{-3}$. Bottom left: Corresponding noisy data together with $F(c^{\mathrm{jump}})$. Top right: Averaged reconstruction error and its standard deviation using $b^2_{2,1}$-penalization, the rate $\mathcal{O}(\tilde{\sigma}^{1/6})$ in the L^2-norm predicted by Theorem 5.1 (see Example 5.7). Bottom right: Reconstruction error using $b^2_{1,1}$-penalization, the rate $\mathcal{O}(\tilde{\sigma}^{12/47})$ in the $L^{4/3}$-norm predicted by Theorem 5.2 (see Example 5.8).*

6. Discussion and conclusions. We end this paper with a summary of our results and a comparison to non-oversmoothing regularization theory. Until recently the oversmoothing case in variational regularization theory has been considered more difficult to analyze due to the failure of the tools developed for the non-oversmoothing case so far, which are usually based on some type of source condition. The analysis of this paper, inspired by a series of recent papers discussed in the Introduction, suggests that, on the contrary, oversmoothing may be considered the easier case. The theory is now more complete in many respects than the theory of non-oversmoothing Banach space regularization, as the following examples demonstrate:

- For oversmoothing Banach space regularization, in contrast to the non-oversmoothing case, convergence rate results always remain valid if the norm in the penalty term is replaced by an equivalent norm.

- So far the analysis of non-oversmoothing Besov space penalization (see [17, 19, 20, 32]) has been restricted to certain choices of the Besov norm indices r, p, and q and the norm power u, whereas Corollary 3.2 with the generalization in Remark 3.4 only assumes $r > 0$.
- We are not aware of a convergence rate analysis of BV-regularization for the case that the solution belongs to a smoothness class that is smaller than BV. (The case that the solution smoothness is exactly BV has been analyzed in [9] in a statistical setting.) In contrast, Corollary 4.2 provides optimal convergence rates for BV-regularization if the solution only belongs to smoothness classes larger than BV.

On the other hand, an analysis of exponentially smoothing forward operators and other operators not satisfying a two-sided Lipschitz condition is still missing so far for oversmoothing Banach space regularization. Moreover, more flexibility in the choice of the loss function would be desirable for both the oversmoothing and the non-oversmoothing cases, to allow for natural or desirable norms and for comparisons of different methods.

Appendix A. Tools from abstract interpolation theory. We first characterize the second part of Assumption 2.1.

PROPOSITION A.1 (Interpolation inequality (see [3, Section 3.5 and Theorem 3.11.4])). *Suppose that $\mathbb{X}_{\mathcal{R}}$, $\mathbb{X}_{\mathrm{L}}$, and $\mathbb{X}_-$ are quasi-Banach spaces with continuous embeddings $\mathbb{X}_{\mathcal{R}} \subset \mathbb{X}_{\mathrm{L}} \subset \mathbb{X}_-$ and $\xi \in (0,1)$. Then the following statements are equivalent:*

1. *$\mathbb{X}_{\mathrm{L}}$ continuously embeds into $(\mathbb{X}_-, \mathbb{X})_{\xi,1}$.*
2. *There exists a constant $c > 0$ such that*

$$\|f\|_{\mathbb{X}_{\mathrm{L}}} \leq c\,\|f\|_{\mathbb{X}_-}^{1-\xi} \cdot \|f\|_{\mathbb{X}}^{\xi} \quad \text{for all } f \in \mathbb{X}.$$

PROPOSITION A.2. *Let $\mathbb{X}$ and $\mathbb{X}_-$ be quasi-Banach spaces with a continuous embedding $\mathbb{X} \subset \mathbb{X}_-$. Then we have $\mathbb{X} \subset (\mathbb{X}_-, \mathbb{X})_{1,\infty}$ with embedding constant equal to 1.*

Proof. Let $f \in \mathbb{X}$. Then we insert $h = f$ in the K-functional (2.1) to obtain

$$K(t, f) \leq t\|f\|_{\mathbb{X}} \quad \text{for all } t > 0.$$

Hence $f \in (\mathbb{X}_-, \mathbb{X})_{1,\infty}$ with $\|f\|_{(\mathbb{X}_-,\mathbb{X})_{1,\infty}} \leq \|f\|_{\mathbb{X}}$. ∎

PROPOSITION A.3 (Reiteration). *Let $\mathbb{X}$ and $\mathbb{X}_-$ be quasi-Banach spaces with a continuous embedding $\mathbb{X} \subset \mathbb{X}_-$, and let $0 < \xi < \theta \leq 1$. Then*

$$(\mathbb{X}_-, \mathbb{X})_{\xi,1} = (\mathbb{X}_-, (\mathbb{X}_-, \mathbb{X})_{\theta,\infty})_{\xi/\theta,1}$$

with equivalent quasi-norms.

Proof. In the notation of [3, Definition 3.5.1] we have that $\mathbb{X}_-$ is of class $\mathcal{C}(0, (\mathbb{X}_-, \mathbb{X}))$. Moreover, $(\mathbb{X}_-, \mathbb{X})_{\theta,\infty}$ is of class $\mathcal{C}(\theta, (\mathbb{X}_-, \mathbb{X}))$. If $\theta < 1$ this is due to [3, Theorem 3.11.4]. For $\theta = 1$ the definition yields that $(\mathbb{X}_-, \mathbb{X})_{1,\infty}$ is of class $\mathcal{C}_K(1, (\mathbb{X}_-, \mathbb{X}))$; see [3, Definition 3.5.1]. Moreover, from Proposition A.2 we see that

$$\|f\|_{(\mathbb{X}_-,\mathbb{X})_{1,\infty}} \leq \|f\|_{\mathbb{X}} \leq t^{-1} \max\{\|f\|_{\mathbb{X}_-}, t\|f\|_{\mathbb{X}}\}.$$

Hence $(\mathbb{X}_-, \mathbb{X})_{1,\infty}$ is of class $\mathcal{C}_J(1, (\mathbb{X}_-, \mathbb{X}))$; see again [3, Definition 3.5.1]. Therefore, the result follows from the reiteration theorem [3, Theorem 3.11.5]. ∎

Appendix B. Properties of Besov spaces. As elsewhere, let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. We first review the relations of Besov spaces to L^p-spaces and to the Sobolev spaces $W_p^s(\Omega)$, with $s \geq 0$ and $p \in (1, \infty)$. Recall that, for $s \in \mathbb{N}_0$, Sobolev norms are given by $\|f\|_{W_p^s}^p = \sum_{\alpha \in \mathbb{N}_0^d, |\alpha| \leq s} \|D^\alpha f\|_{L^p}^p$. For non-integer s, these spaces are also called

Sobolev–Slobodeckij spaces, and for $p = 2$ they coincide with the $H_2^s(\Omega)$ spaces defined on $\mathbb{R}^d$ via Fourier transformation for all $s \in \mathbb{R}$.

PROPOSITION B.1 (Embeddings with L^p and Sobolev spaces).

1. *Let $p \in (1, \infty)$. Then we have continuous embeddings*

(B.1) $$B^0_{p,\min\{p,2\}}(\Omega) \subset L^p(\Omega) \subset B^0_{p,\max\{p,2\}}(\Omega).$$

For $p = 1$ the following continuous embeddings hold true:

$$B^0_{1,1}(\Omega) \subset L^1(\Omega) \subset B^0_{1,\infty}(\Omega).$$

2. $B^s_{p,p}(\Omega) = W^s_p(\Omega)$ *with equivalent norms for all $0 < s \notin \mathbb{N}$ and $p \in (1, \infty)$, and in the case of $p = 2$ for all $s \in \mathbb{R}$.*
3. *Let $p_0, p_1, q_0, q_1 \in (0, \infty]$ and $-\infty < s_1 < s_0 < \infty$. Then*

(B.2) $$B^{s_0}_{p_0,q_0}(\Omega) \subset B^{s_1}_{p_1,q_0}(\Omega) \quad \text{if } s_0 - d/p_0 = s_1 - d/p_1,$$

and

(B.3) $$B^{s_0}_{p_0,q_0}(\Omega) \subset B^{s_1}_{p_1,q_1}(\Omega) \quad \text{if } s_0 - d/p_0 > s_1 - d/p_1.$$

4. *Let $s \in \mathbb{R}$ and $p, q_0, q_1 \in (0, \infty]$. Then*

(B.4) $$B^s_{p,q_0}(\Omega) \subset B^s_{p,q_1}(\Omega) \quad \text{if } q_0 \le q_1.$$

Proof. First note that, as all occurring spaces on bounded Lipschitz domains in $\mathbb{R}^d$ are defined by restriction of the respective spaces on $\mathbb{R}^d$, it suffices to prove the assertions for $\Omega = \mathbb{R}^d$.

1. Let $F^s_{p,q}(\Omega)$ be the function spaces defined in [30, 2.3.1 Definition 2(ii)] for $\Omega = \mathbb{R}^d$. By [30, 3.2.4(3)] we have continuous embeddings

$$B^s_{p,\min\{p,q\}}(\Omega) \subset F^s_{p,q}(\Omega) \subset B^s_{p,\max\{p,q\}}(\Omega) \quad \text{for all } p, q \in [1, \infty).$$

With this, the embeddings for $p > 1$ follow from the identity $F^0_{p,2}(\Omega) = L^p(\Omega)$ with equivalent norms. The latter identity can be found in [30, 2.5.6]. For the assertion in the case $p = 1$, we refer to [30, 2.5.7(2)].
2. See [27, Section 2.3 and Theorem 4.2.4].
3. See [30, Proposition 3.3.1].
4. The inclusion for Ω replaced by $\mathbb{R}^d$ can be found in [30, equation (2.3.2/5)]. Using the definition of the $B^s_{p,q}(\Omega)$ spaces, this easily implies the assertion.

The proof is complete. □

We now recall some well-known results on interpolation of Besov spaces. Besides K-interpolation reviewed in Section 2.1, we also refer to the complex interpolation method in some remarks. The latter only works for complex Banach spaces X_0, X_1 as well as some quasi-Banach spaces, and it is denoted by $[X_0, X_1]_\theta = X_\theta$ for $\theta \in (0, 1)$; see [3].

PROPOSITION B.2 (Interpolation of Besov spaces). *Let $s_0, s_1 \in \mathbb{R}$, $s_0 \neq s_1$, $\theta \in (0, 1)$, and $s_\theta := (1 - \theta)s_0 + \theta s_1$.*

1. *For $p, q_0, q_1 \in (0, \infty]$ and $q \in [1, \infty]$ we have*

(B.5) $$(B^{s_0}_{p,q_0}(\Omega), B^{s_1}_{p,q_1}(\Omega))_{\theta,q} = B^{s_\theta}_{p,q}(\Omega).$$

2. *If $p_0, p_1, p_\theta \in (0, \infty)$ with $1/p_\theta = (1 - \theta)/p_0 + \theta/p_1$, then*

(B.6) $$(B^{s_0}_{p_0,p_0}(\Omega), B^{s_1}_{p_1,p_1}(\Omega))_{\theta,p_\theta} = B^{s_\theta}_{p_\theta,p_\theta}(\Omega).$$

3. *If* $p_0, p_1, p_\theta \in [1,\infty)$ *with* $1/p_\theta = (1-\theta)/p_0 + \theta/p_1$, *and* $q_0, q_\theta, q_1 \in [1,\infty)$ *with* $1/q_\theta = (1-\theta)/q_0 + \theta/q_1$, *then*

$$[B^{s_0}_{p_0,q_0}(\Omega), B^{s_1}_{p_1,q_1}(\Omega)]_\theta = B^{s_\theta}_{p_\theta,q_\theta}(\Omega). \tag{B.7}$$

Proof. See [30, Theorem 3.3.6] for the first and last statements and [30, Theorem 2.4.3 and Remark 8 in Section 3.3.6] for the second statement. □

Appendix C. On spaces of functions of bounded variation. Finally, we also recall and generalize some results on functions of bounded variation.

PROPOSITION C.1 (Embedding). *Let* $a \geq 0$ *with* $a \geq d/2 - 1$. *Then there is a continuous embedding* $\mathrm{BV}(\Omega) \subset B^{-a}_{2,2}(\Omega)$.

Proof. For all embeddings involving the space $\mathrm{BV}(\Omega)$ in this proof, we refer to [2, Corollary 3.49 and Proposition 3.21]. For $d = 1$ there is a continuous embedding $\mathrm{BV}(\Omega) \subset L^2(\Omega)$. We have $L^2(\Omega) = B^0_{2,2}(\Omega) \subset B^{-a}_{2,2}(\Omega)$, which yields the claim in this case.

For $d > 1$ we set $p := d/(d-1)$. Then $p \in (1, 2]$ and there is a continuous embedding $\mathrm{BV}(\Omega) \subset L^p(\Omega)$. By Proposition B.1 we have a continuous embedding $L^p(\Omega) \subset B^0_{p,2}(\Omega)$. Furthermore, $a + d/2 \geq d - 1 = d/p$ yields a continuous embedding $B^0_{p,2}(\Omega) \subset B^{-a}_{2,2}(\Omega)$. Putting together the last three embeddings yields the claim. □

PROPOSITION C.2. *Let* $a \geq 0$, $s \in (-a, 1)$, *and* $\Omega \subset \mathbb{R}^d$ *a bounded Lipschitz domain. Then*

$$B^s_{t_s,t_s}(\Omega) = (B^{-a}_{2,2}(\Omega), \mathrm{BV}(\Omega))_{\theta_s,t_s} \quad \textit{with } \theta_s := \frac{s+a}{a+1} \textit{ and } t_s := \frac{2a+2}{s+2a+1},$$

with equivalent norms.

Proof. First note that if $f \in \mathrm{BV}(\mathbb{R}^d)$, then

$$f|_\Omega \in \mathrm{BV}(\Omega) \quad \text{with } \|f|_\Omega\|_{\mathrm{BV}(\Omega)} \leq \|f\|_{\mathrm{BV}(\mathbb{R}^d)}.$$

Due to [6, Theorem 1.4] the claim holds true for $\Omega = \mathbb{R}^d$. Note that here the condition $\gamma < 1 - 1/d$ from the latter reference on $\gamma := -(2a+2)/d + 1$ is satisfied. Let c_1 be a constant such that the norm in $B^s_{t_s,t_s}(\mathbb{R}^d)$ is bounded by c_1 times the norm in $(B^{-a}_{2,2}(\mathbb{R}^d), \mathrm{BV}(\mathbb{R}^d))_{\theta_s,t_s}$ and the other way around.

We transfer this result to bounded Lipschitz domains. To this end we separately prove both inclusions in the stated identity.

Let $f \in B^s_{t_s,t_s}(\Omega)$. Then there exists $\tilde{f} \in B^s_{t_s,t_s}(\mathbb{R}^d)$ with

$$\tilde{f}|_\Omega = f \quad \text{and} \quad \|\tilde{f}\|_{B^s_{t_s,t_s}(\mathbb{R}^d)} \leq 2\|f\|_{B^s_{t_s,t_s}(\Omega)}.$$

Let $t > 0$ and $\tilde{f} = \tilde{f}_1 + \tilde{f}_2$ with $\tilde{f}_1 \in B^{-a}_{2,2}(\mathbb{R}^d)$ and $\tilde{f}_2 \in \mathrm{BV}(\mathbb{R}^d)$ be a decomposition such that

$$\|\tilde{f}_1\|_{B^{-a}_{2,2}(\mathbb{R}^d)} + t\|\tilde{f}_2\|_{\mathrm{BV}(\mathbb{R}^d)} \leq 2K(t, \tilde{f})$$

with the K-functional from real interpolation of Banach spaces. Then $\tilde{f}_1|_\Omega \in B^{-a}_{2,2}(\Omega)$, $\tilde{f}_2|_\Omega \in \mathrm{BV}(\Omega)$, $f = f_1 + f_2$, and

$$K(t, f) \leq \|\tilde{f}_1|_\Omega\|_{B^{-a}_{2,2}(\Omega)} + t\|\tilde{f}_2|_\Omega\|_{\mathrm{BV}(\Omega)} \leq 2K(t, \tilde{f}).$$

Hence with the definition of the norm on real interpolation spaces, we obtain

$$\begin{aligned} \|f\|_{(B^{-a}_{2,2}(\Omega),\mathrm{BV}(\Omega))_{\theta_s,t_s}} &\leq 2\|\tilde{f}\|_{(B^{-a}_{2,2}(\mathbb{R}^d),\mathrm{BV}(\mathbb{R}^d))_{\theta_s,t_s}} \\ &\leq 2c_1\|\tilde{f}\|_{B^s_{t_s,t_s}(\mathbb{R}^d)} \leq 4c_1\|f\|_{B^s_{t_s,t_s}(\Omega)}. \end{aligned}$$

We turn to the other inclusion. There exists a constant $C_{\text{ext}} > 0$ such that for every $f \in B^{-a}_{2,2}(\Omega)$ there exists $\tilde{f} \in B^{-a}_{2,2}(\mathbb{R}^d)$ with

$$\tilde{f}|_\Omega = f \quad \text{and} \quad \|\tilde{f}\|_{B^{-a}_{2,2}(\mathbb{R}^d)} \le C_{\text{ext}} \|f\|_{B^{-a}_{2,2}(\Omega)},$$

and likewise for every $f \in \mathrm{BV}(\Omega)$ there exists $\tilde{f} \in \mathrm{BV}(\mathbb{R}^d)$ with

$$\tilde{f}|_\Omega = f \quad \text{and} \quad \|\tilde{f}\|_{\mathrm{BV}(\mathbb{R}^d)} \le C_{\text{ext}} \|f\|_{\mathrm{BV}(\Omega)}.$$

This holds true by the definition of $B^{-a}_{2,2}(\Omega)$ via restrictions and due to [2, Proposition 3.21] for bounded variation functions. Now suppose $f \in (B^{-a}_{2,2}(\Omega), \mathrm{BV}(\Omega))_{\theta_s, t_s}$. Let $f = f_1 + f_2$ with $f_1 \in B^{-a}_{2,2}(\Omega)$ and $f_2 \in \mathrm{BV}(\Omega)$ such that

$$\|f_1\|_{B^{-a}_{2,2}(\Omega)} + t\|f_2\|_{\mathrm{BV}(\Omega)} \le 2K(t, f).$$

Let $\tilde{f}_1 \in B^{-a}_{2,2}(\mathbb{R}^d)$ and $\tilde{f}_2 \in \mathrm{BV}(\mathbb{R}^d)$ be extensions as above. Then $\tilde{f} := \tilde{f}_1 + \tilde{f}_2$ satisfies $\tilde{f}|_\Omega = f$, and

$$K(t, \tilde{f}_1 + \tilde{f}_2) \le \|\tilde{f}_1\|_{B^{-a}_{2,2}(\mathbb{R}^d)} + t\|\tilde{f}_2\|_{\mathrm{BV}(\mathbb{R}^d)} \le 2C_e K(t, f).$$

We conclude that

$$\begin{aligned} \|f\|_{B^s_{t_s,t_s}(\Omega)} &\le \|\tilde{f}\|_{B^s_{t_s,t_s}(\mathbb{R}^d)} \le c_1 \|\tilde{f}\|_{(B^{-a}_{2,2}(\mathbb{R}^d), \mathrm{BV}(\mathbb{R}^d))_{\theta_s,t_s}} \\ &\le 2c_1 C_e \|f\|_{(B^{-a}_{2,2}(\Omega), \mathrm{BV}(\Omega))_{\theta_s,t_s}}. \qquad \square \end{aligned}$$

REFERENCES

[1] R. ACAR AND C. R. VOGEL, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems, 10 (1994), pp. 1217–1229.
[2] L. AMBROSIO, N. FUSCO, AND D. PALLARA, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford University Press, New York, 2000.
[3] J. BERGH AND J. LÖFSTRÖM, *Interpolation Spaces*, Springer, Berlin, 1976.
[4] K. BREDIES AND D. A. LORENZ, *Regularization with non-convex separable constraints*, Inverse Problems, 25 (2009), Art. 085011, 14 pages.
[5] D.-H. CHEN, B. HOFMANN, AND I. YOUSEPT, *Oversmoothing Tikhonov regularization in Banach spaces*, Inverse Problems, 37 (2021), Art. 085007, 28 pages.
[6] A. COHEN, W. DAHMEN, I. DAUBECHIES, AND R. DEVORE, *Harmonic analysis of the space BV*, Rev. Mat. Iberoamericana, 19 (2003), pp. 235–263.
[7] M. DEL ÁLAMO, *Multiscale Total Variation Estimators for Regression and Inverse Problems*, PhD. Thesis, PhD. School of Math. Sci., Georg-August University, Göttingen, 2019.
[8] M. DEL ÁLAMO, H. LI, AND A. MUNK, *Frame-constrained total variation regularization for white noise regression*, Ann. Statist., 49 (2021), pp. 1318–1346.
[9] M. DEL ÁLAMO AND A. MUNK, *Total variation multiscale estimators for linear inverse problems*, Inf. Inference, 9 (2020), pp. 961–986.
[10] D. L. DONOHO AND I. M. JOHNSTONE, *Minimax estimation via wavelet shrinkage*, Ann. Statist., 26 (1998), pp. 879–921.
[11] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.
[12] D. FREITAG, *Real interpolation of weighted L_p-spaces*, Math. Nachr., 86 (1978), pp. 15–18.
[13] D. GERTH AND B. HOFMANN, *Oversmoothing regularization with ℓ^1-penalty term*, AIMS Math., 4 (2019), pp. 1223–1247.
[14] E. GINÉ AND R. NICKL, *Mathematical Foundations of Infinite-Dimensional Statistical Models*, Cambridge University Press, New York, 2016.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

[15] B. HOFMANN AND P. MATHÉ, *Tikhonov regularization with oversmoothing penalty for non-linear ill-posed problems in Hilbert scales*, Inverse Problems, 34 (2018), Art. 015007, 14 pages.
[16] B. HOFMANN AND R. PLATO, *Convergence results and low-order rates for nonlinear Tikhonov regularization with oversmoothing penalty term*, Electron. Trans. Numer. Anal., 53 (2020), pp. 313–328. http://etna.ricam.oeaw.ac.at/vol.53.2020/pp313-328.dir/pp313-328.pdf
[17] T. HOHAGE AND P. MILLER, *Optimal convergence rates for sparsity promoting wavelet-regularization in Besov spaces*, Inverse Problems, 35 (2019), Art. 065005, 27 pages.
[18] E. MAMMEN AND S. VAN DE GEER, *Locally adaptive regression splines*, Ann. Statist., 25 (1997), pp. 387–413.
[19] P. MILLER, *Variational regularization theory based on image space approximation rates*, Inverse Problems, 37 (2021), Art. 065003, 32 pages.
[20] P. MILLER AND T. HOHAGE, *Maximal spaces for approximation rates in ℓ^1-regularization*, Numer. Math., 149 (2021), pp. 341–374.
[21] F. NATTERER, *Error bounds for Tikhonov regularization in Hilbert scales*, Applicable Anal., 18 (1984), pp. 29–37.
[22] A. RASTOGI, *Tikhonov regularization with oversmoothing penalty for nonlinear statistical inverse problems*, Commun. Pure Appl. Anal., 19 (2020), pp. 4111–4126.
[23] O. SCHERZER, M. GRASMAIR, H. GROSSAUER, M. HALTMEIER, AND F. LENZEN, *Variational Methods in Imaging*, Springer, New York, 2009.
[24] T. SCHUSTER, B. KALTENBACHER, B. HOFMANN, AND K. S. KAZIMIERSKI, *Regularization Methods in Banach Spaces*, de Gruyter, Berlin, 2012.
[25] I. W. SELESNICK AND I. BAYRAM, *Sparse signal estimation by maximally sparse convex optimization*, IEEE Trans. Signal Process., 62 (2014), pp. 1078–1092.
[26] A. N. TIKHONOV, *On the solution of incorrectly formulated problems and the regularization method*, Soviet Math. Doklady, 4 (1963), pp. 1035–1038.
[27] H. TRIEBEL, *Interpolation Theory, Function Spaces, Differential Operators*, VEB Deutscher Verlag der Wissenschaften, Berlin, 1978.
[28] ———, *Theory of Function Spaces II*, Birkhäuser, Basel, 1992.
[29] ———, *Function Spaces and Wavelets on Domains*, European Mathematical Society, Zürich, 2008.
[30] ———, *Theory of Function Spaces*, Springer, Basel, 2010.
[31] M. C. VERAAR, *Regularity of Gaussian white noise on the d-dimensional torus*, in Marcinkiewicz Centenary Volume, M. Nawrocki and W. Wnuk, eds., vol. 95 of Banach Center Publ., Polish Acad. Sci. Inst. Math., Warsaw, 2011, pp. 385–398.
[32] F. WEIDLING, B. SPRUNG, AND T. HOHAGE, *Optimal convergence rates for Tikhonov regularization in Besov spaces*, SIAM J. Numer. Anal., 58 (2020), pp. 21–47.
[33] C. A. ZARZER, *On Tikhonov regularization with non-convex sparsity constraints*, Inverse Problems, 25 (2009), Art. 025006, 13 pages.