Electronic Transactions on Numerical Analysis.
Volume 57, pp. 127–152, 2022.

ISSN 1068–9613.
DOI: 10.1553/etna_vol57s127

ETNA
Kent State University and
Johann Radon Institute (RICAM)

ERROR ESTIMATES FOR VARIATIONAL REGULARIZATION OF INVERSE PROBLEMS WITH GENERAL NOISE MODELS FOR DATA AND OPERATOR*

THORSTEN HOHAGE† AND FRANK WERNER‡

Abstract. This paper is concerned with variational regularization of inverse problems where both the data and the forward operator are given only approximately. We propose a general approach to derive error estimates which separates the analysis of smoothness of the exact solution from the analysis of the effect of errors in the data and the operator. Our abstract error bounds are applied to both discrete and continuous data, random and deterministic types of error, as well as Huber data fidelity terms for impulsive noise.

Key words. inverse problem, variational regularization, error bounds, operator noise, random noise

AMS subject classifications. 65J20, 60H30, 60H40

1. Introduction. The general setting of this paper (to be specified later) is as follows: Let $\mathcal{X}$ and $\mathcal{Y}$ be normed spaces and $F : D(F) \subset \mathcal{X} \to \mathcal{Y}$ be a continuous forward operator. We consider the inverse problem to reconstruct some unknown $u^\dagger \in D(F)$ from observations of

$$g^\dagger := F(u^\dagger).$$

We assume that only approximations

$$g^{\mathrm{obs}} \approx g^\dagger \quad \text{and} \quad F_h \approx F$$

of the right-hand side $g^\dagger$ and the operator F are at our disposal, where the precise meaning of $\approx$ depends on the specific model considered. Here

$$F_h : D(F) \to \mathcal{Y}_h$$

may be a numerical approximation of F in finite-dimensional spaces, or it may contain modeling errors of the forward problem. We do not assume that $\mathcal{Y}_h \subset \mathcal{Y}$ or that $g^{\mathrm{obs}} \in \mathcal{Y}$ or $g^{\mathrm{obs}} \in \mathcal{Y}_h$.

Famous examples of data noise models are deterministic ones where $g^{\mathrm{obs}} \in \mathcal{Y}$ satisfies $\|g^{\mathrm{obs}} - g^\dagger\|_{\mathcal{Y}} \leq \delta$ with some (small) noise parameter δ, or continuous random noise, where in general $g^{\mathrm{obs}} \notin \mathcal{Y}$ a.s., but $\mathbb{E}[g^{\mathrm{obs}}] = g$ in a suitable sense.

For the solution of this problem, we focus on variational regularization methods

$$\hat{u}_\alpha \in \operatorname*{argmin}_{u \in D(F)} \left[\frac{1}{\alpha} \mathcal{S}(F_h(u); g^{\mathrm{obs}}) + \mathcal{R}(u) \right] \tag{1.1}$$

where $\mathcal{S}(\cdot; g^{\mathrm{obs}}) : \mathcal{Y}_h \to (-\infty, \infty]$ is a so-called data fidelity term, $\mathcal{R} : \mathcal{X} \to (-\infty, \infty]$ is a penalty term, and $\alpha > 0$ is a regularization parameter (see, e.g., [31, 32] and references therein). The precise choice of $\mathcal{S}$ and $\mathcal{R}$ depends solely on the specific problem instance. Typically, $\mathcal{S}$ will be chosen as the negative log-likelihood of the data distribution, or an

*Received March 14, 2022. Accepted June 20, 2022. Published online on August 9, 2022. Recommended by Elena Resmerita.

†Institute for Numerical and Applied Mathematics, University of Göttingen, 37083 Göttingen, Germany (hohage@math.uni-goettingen.de).

‡Institute of Mathematics, University of Würzburg, 97074 Würzburg, Germany (frank.werner@mathematik.uni-wuerzburg.de).

ETNA
Kent State University and
Johann Radon Institute (RICAM)

approximation thereof (maybe including discretization effects; see Sections 4–6 for some examples), and $\mathcal{R}$ incorporates a priori information on $u^\dagger$ in the sense that $\mathcal{R}(u)$ is small if and only if u is compatible with this a priori information.

The existence of (global) minimizers of (1.1) can – at least in principle – be ensured by standard arguments under reasonable assumptions on $\mathcal{S}$ and $\mathcal{R}$; see, e.g., [10] or [31].

In this paper, we will just assume the existence of global minimizers of (1.1) and aim to provide a general framework for the derivation of error bounds for the reconstructions $\hat{u}_\alpha$ in (1.1). As usual in inverse problems, these error bounds will depend on both the ill-posedness of the inverse problem (in terms of smoothness of $u^\dagger$ and smoothing properties of F) and the specific noise model. However, we avoid focusing on one specific noise model, but provide abstract conditions under which error bounds can be obtained. Later on, we will discuss several examples in detail, such as classical deterministic noise, Gaussian white noise, impulsive noise, and the setting of discretized observations.

The abstract conditions formulated in this paper are threefold and can roughly be described as follows.

(A) Smoothness of $u^\dagger$: We assume that $u^\dagger$ satisfies a variational source condition (see Assumption A below). This way of measuring the smoothness of $u^\dagger$, first formulated in [15] (see also [11, 13, 39]) has turned out to be very useful, and in many situations variational source conditions are necessary and sufficient for convergence rates [18]. As typical for source conditions in general, the smoothness of $u^\dagger$ is measured relative to the smoothing properties of F.

(B) Data and operator error: As the datum g^{obs} as well as the approximation F_h of the forward operator F occur in (1.1) only via the data fidelity functional $u \mapsto \mathcal{S}(F_h(u); g^{\mathrm{obs}})$, we introduce a variational noise functional err, which measures the influence of g^{obs} and F_h on $\mathcal{S}$. Such a functional was first introduced in [43], and since then has been proven to be useful in many models [19, 41]. Our only assumption on the specific noise model considered for g^{obs} and F_h is then a suitable (but not uniform) upper bound for err which is compatible with the smoothing properties of the true F assumed in condition (C); cf. Assumption B below.

(C) Smoothing properties of F: We assume an interpolation-type inequality which allows us to interpolate data error bounds between the images of F and the loss functional ℓ to be introduced below; see Assumptions C1 and C2.

Under these assumptions, we are able to provide abstract error bounds for $\hat{u}_\alpha$ as in (1.1); see Theorems 3.1 and 3.4 below. Although rather technical, these error bounds are applicable to a wide range of situations. We will discuss their implications for several data and operator noise models, including Gaussian white noise, discrete data, and impulsive noise with a Huber-type data fidelity term.

Let us discuss some results in the literature on the treatment of different types of noise in the data and the operator in variational regularization: Neubauer and Scherzer [26, 27] considered nonlinear Tikhonov regularization in Hilbert spaces with finite-dimensional projections using spectral methods; see also [25] for results on discrete data and operators in variable Hilbert scales.

Concerning variational regularization in Banach spaces, consistency results for perturbed operators were derived in [29]. In [23] an error bound for perturbed operators is derived under the classical source condition $\partial\mathcal{R}(u^\dagger) \in \mathrm{range}(F'[u^\dagger]^*)$. Under the same source condition, error bounds were also shown in [6, 7, 33] for more general data fidelity terms, in the first two references for perturbed operators satisfying bounds in Banach lattices. We also refer to [14] for convergence rates of wavelet thresholding methods for perturbed operators.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Concerning convergence rates of variational regularization in Banach spaces with statistical errors in the data, we refer to [17, 41, 45], where the noise has been described by stochastic processes in a continuous setting. Here we will also analyze more realistic discrete settings.

The proposed framework allows us to analyze for the first time convergence rates with respect to perturbations for non-quadratic variational regularization if the solution is not smooth enough to satisfy $\partial\mathcal{R}(u^\dagger) \in \mathrm{range}(F'[u^\dagger]^*)$. Moreover, we will derive new error bounds for perturbed operators and discrete statistical noise models as well as discrete impulsive noise.

The outline of this paper is as follows. In Section 2 we present and discuss the assumptions mentioned before in detail. Section 3 is then devoted to the main general results on error bounds, including some discussions. We first apply these to perturbed data and operators in a continuous setting in Section 4, and then to discrete data in Section 5. Finally, in Section 6 we study impulsive noise with a Huber-type data fidelity term, before we end this paper with some conclusions in Section 7.

2. Assumptions and preparations.

2.1. Bounds on the approximation error. In this subsection we report on recent progress in deriving sharp and computable bounds on the approximation error (or bias) of variational estimators. The focus of this paper is on bounds of the propagated data error rather than the bias, but it is convenient to start with the terminology needed for the bias bounds.

We study an estimator in which the data fidelity term $\mathcal{S}(F_h(u); g^{\mathrm{obs}})$ in (1.1) is replaced by an ideal noise-free data fidelity term involving the exact operator F and data fidelity functional $\mathcal{T} : \mathcal{Y} \times \mathcal{Y} \to [0, \infty]$:

$$u_\alpha \in \operatorname*{argmin}_{u \in D(F)} \left[\frac{1}{\alpha} \mathcal{T}(F(u); g^\dagger) + \mathcal{R}(u) \right]. \tag{2.1}$$

We will assume that

$$\mathcal{T}(g; g^\dagger) = 0 \quad \Longleftrightarrow \quad g = g^\dagger. \tag{2.2}$$

The approximation error will be measured in terms of a **loss function** $\ell : \mathcal{X} \times \mathcal{X} \to [0, \infty]$, i.e., we will consider $\ell(u_\alpha, u^\dagger)$ as the distance between u_α and $u^\dagger$. Note that ℓ does not have to be a distance in the sense of a metric, as we do not assume definiteness, symmetry, or a triangle inequality. The most common examples for ℓ are norm powers

$$\ell(u, u^\dagger) = \|u - u^\dagger\|_{\mathcal{X}}^r$$

with $r \geq 1$ and the Bregman divergence

$$\ell(u, u^\dagger) = \mathcal{D}_{\mathcal{R}}^{u^*}(u; u^\dagger) = \mathcal{R}(u) - \mathcal{R}(u^\dagger) - \langle u^*, u - u^\dagger \rangle,$$

where $u^* \in \partial\mathcal{R}(u^\dagger)$ is a subgradient of the penalty term $\mathcal{R}$ at $u^\dagger$. Often Bregman divergences are an analytically convenient distance measure for the study of variational regularization methods. In an r-convex Banach space with $r \in [2, \infty)$ and $\mathcal{R}(u) = (1/r)\|u\|_{\mathcal{X}}^r$ (e.g., L^r or $W^{k,r}$; see [46, Section 1] for a definition and discussion), the Bregman divergence is lower-bounded by a norm power (see [34, 46]), and in Hilbert spaces one even has equality. For ℓ^1-type regularization, error bounds with respect to Bregman divergences are less informative, and bounds with respect to the norm ($r = 1$) have been derived directly [17].

ASSUMPTION A. *Suppose that* $u^\dagger \in D(F)$ *satisfies a* ***variational source condition***

$$\ell(u,u^\dagger) \le \mathcal{R}(u) - \mathcal{R}(u^\dagger) + \varphi(\mathcal{T}(F(u);F(u^\dagger))) \quad \text{for all } u \in D(F), \tag{2.3}$$

with an index function $\varphi : [0,\infty) \to [0,\infty)$*; i.e.,* φ *is continuous, increasing, and* $\varphi(0) = 0$.

If ℓ and $\mathcal{T}$ are symmetric and Assumption A is satisfied for all $u^\dagger$ in some subset $\tilde{D} \subset D(F)$, then the global conditional stability estimate $\ell(u_1,u_2) \le \varphi(\mathcal{T}(F(u_1);F(u_2)))$ holds true for all $u_1,u_2 \in \tilde{D}$. In particular, we have global uniqueness in $\tilde{D}$ if ℓ is positive definite.

Comparing u_α with $u^\dagger$ in the minimality condition from (2.1) and using (2.2), this implies

$$\begin{aligned}\ell(u_\alpha,u^\dagger) &\le \mathcal{R}(u_\alpha) - \mathcal{R}(u^\dagger) + \varphi(\mathcal{T}(F(u_\alpha);F(u^\dagger)))\\ &\le -\frac{1}{\alpha}\mathcal{T}(F(u_\alpha);F(u^\dagger)) + \varphi(\mathcal{T}(F(u_\alpha);F(u^\dagger))),\end{aligned}$$

and using an argument due to Grasmair [13], Assumption A consequently yields the bias bound $\ell(u_\alpha,u^\dagger) \le \varphi_{\mathrm{app}}(\alpha)$ with

$$\varphi_{\mathrm{app}}(\alpha) := \sup_{\tau \ge 0}\left[-\frac{1}{\alpha}\tau + \varphi(\tau)\right]. \tag{2.4}$$

Note that

$$\varphi_{\mathrm{app}}(\alpha) = (-\varphi)^*\left(-\frac{1}{\alpha}\right), \quad \alpha > 0,$$

if $\psi^*(t) := \sup_{s\in\mathbb{R}}[st - \psi(s)]$ denotes the Fenchel conjugate of a function $\psi : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ and if we set $\varphi(t) := -\infty$ for $t < 0$.

REMARK 2.1. The function φ_{app} has the following properties (see, e.g., [44]):

(a) $\varphi_{\mathrm{app}}(\alpha) \ge 0$ for all $\alpha > 0$;

(b) if $\varphi^{1+\varepsilon}$ is concave for some $\varepsilon > 0$, then φ_{app} satisfies

$$\varphi_{\mathrm{app}}(C\alpha) \le C^{1/\varepsilon}\varphi_{\mathrm{app}}(\alpha) \quad \text{for all } C \ge 1,\ \alpha > 0;$$

(c) φ_{app} is monotonically increasing; and

(d) if φ^2 is concave, then $\varphi_{\mathrm{app}}(\alpha) \searrow 0$ as $\alpha \searrow 0$.

EXAMPLE 2.2. Let $\Omega \subset \mathbb{R}^n$ and $\mathbb{M} \subset \mathbb{R}^d$ be smooth bounded domains or smooth bounded manifolds. Assume that $F : D(F) \subset \mathbf{L}^2(\Omega) \to \mathbf{L}^2(\mathbb{M})$ is at most a-times smoothing in the sense that there exists some C_F such that

$$\|u - v\|_{H^{-a}(\Omega)} \le C_F\|F(u) - F(v)\|_{\mathbf{L}^2(\mathbb{M})} \quad \text{for all } u,v \in D(F), \tag{2.5}$$

and that $u^\dagger \in H_0^s(\Omega)$ with $0 < s < a$ and $\|u^\dagger\|_{H^s} \le \rho$.

Interpolation in Sobolev spaces (see, e.g., [36, Section 4.3.1]) and Young's inequality yield

$$\begin{aligned}\langle u^\dagger, u^\dagger - u\rangle &\le \|u^\dagger\|_{H^s}\|u^\dagger - u\|_{H^{-s}}\\ &\le \rho\|u^\dagger - u\|_{\mathbf{L}^2}^{(a-s)/a}\|u^\dagger - u\|_{H^{-a}}^{s/a}\\ &\le \frac{1}{4}\|u^\dagger - u\|_{\mathbf{L}^2}^2 + \frac{a+s}{2a}\left(\frac{a}{2a-2s}\right)^{-(a-s)/(a+s)}\rho^{2a/(a+s)}\|u^\dagger - u\|_{H^{-a}}^{2s/(a+s)}.\end{aligned}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

If we furthermore plug in the identity $\frac{1}{4}\|u-u^\dagger\|^2_{\mathbf{L}^2} = \frac{1}{2}\|u\|^2_{\mathbf{L}^2} - \frac{1}{2}\|u^\dagger\|^2_{\mathbf{L}^2} + \langle u^\dagger, u^\dagger - u\rangle - \frac{1}{4}\|u-u^\dagger\|^2_{\mathbf{L}^2}$ and (2.5), then we find that $u^\dagger$ satisfies the variational source condition

$$\frac{1}{4}\|u-u^\dagger\|^2_{\mathbf{L}^2} \le \frac{1}{2}\|u\|^2_{\mathbf{L}^2} - \frac{1}{2}\|u^\dagger\|^2_{\mathbf{L}^2} + C\rho^{2a/(a+s)}\|F(u^\dagger)-F(u)\|^{2s/(a+s)}_{\mathbf{L}^2} \quad \text{for all } u \in D(F), \tag{2.6}$$

with

$$C := \frac{a+s}{2a}\left(\frac{a}{2a-2s}\right)^{-(a-s)/(a+s)} C_F^{2s/(a+s)}.$$

This is a specific instance of (2.3) with $\ell(u,v) := \frac{1}{4}\|u-v\|^2_{\mathbf{L}^2}$, $\mathcal{R}(u) := \frac{1}{2}\|u\|^2_{\mathbf{L}^2}$, $\mathcal{T}(u,u^\dagger) = \|u-u^\dagger\|^2_{\mathbf{L}^2}$, and $\varphi(t) = C\rho^{2a/(a+s)}t^{s/(a+s)}$. For the function $\varphi_{\text{app}}(\alpha)$, one readily computes by differentiation that

$$\varphi_{\text{app}}(\alpha) = \tilde{C}\rho^2\alpha^{s/a}.$$

Assumption (2.5) has been verified for the Radon transform and parameter identification problems in partial differential equations (PDEs) with distributed measurements [17], and it holds true for injective, elliptic pseudo-differential operators of order $-a$ since the existence of a parametrix [35, Section 7.4] and mapping properties [35, Section 7.5] imply that such operators are Fredholm with index 0 from H^s to H^{s-a} for all s.

A more systematic approach to the characterization of variational source conditions which is applicable to a larger class of inverse problems has been developed in the Ph.D. thesis of Weidling [39]. It is based on describing the smoothness of the solution and the local degree of ill-posedness in terms of a family of finite-dimensional subspaces and has been successfully applied to inverse medium scattering problems [40] and electrical impedance tomography.

REMARK 2.3. Using the approach in [39], the smoothness assumption $u^\dagger \in H^s(\Omega)$ in Example 2.2 can be relaxed to $u^\dagger \in B^s_{2,\infty}(\Omega)$ under further assumptions on Ω, and $s < a$ can be improved to $s \le a$ (with $u^\dagger \in H^a(\Omega)$ for $s = a$), also in Corollaries 4.1, 4.3, and 4.5.

2.2. Bounds on data noise and errors in operator. The main subject of this paper is to derive bounds on the influence of the noise contained in the data g^{obs} and the operator F_h. Since g^{obs} (and also F_h) is only accessed via its associated data fidelity functional $\mathcal{S}(\cdot; g^{\text{obs}}) : \mathcal{Y}_h \to (\infty, \infty]$, a straightforward idea may be to compare the data fidelity term $u \mapsto \mathcal{S}(F_h(u); g^{\text{obs}})$ in (1.1) to the ideal data fidelity functional $u \mapsto \mathcal{T}(F(u); g^\dagger)$ in (2.1). However, $\mathcal{S}$ has a free additive constant, i.e., nothing changes if $\mathcal{S}$ is replaced by $\mathcal{S} + c$ for any $c \in \mathbb{R}$. We do not assume positive definiteness as in (2.2), and for many examples discussed below $\mathcal{S}$ may take negative values. Therefore, we compare $\mathcal{T}(F(\cdot); g^\dagger)$ to $\mathcal{S}(F_h(\cdot); g^{\text{obs}}) - \mathcal{S}(F_h(u^\dagger); g^{\text{obs}})$:

DEFINITION 2.4. *We call the functional* $\text{err} : D(F) \to [-\infty, \infty]$ *given by*

$$\text{err}(u) := \mathcal{T}(F(u); F(u^\dagger)) - (\mathcal{S}(F_h(u); g^{\text{obs}}) - \mathcal{S}(F_h(u^\dagger); g^{\text{obs}}))$$

the ***effective noise level functional*** *on* $D(F) \subset \mathcal{X}$.

If $\mathcal{Y}_h = \mathcal{Y}$, *we also introduce an effective noise level functional on* $\mathcal{Y}$ *by*

$$\text{err}_{\mathcal{Y}}(g) := \mathcal{T}(g; g^\dagger) - (\mathcal{S}(g; g^{\text{obs}}) - \mathcal{S}(g^\dagger; g^{\text{obs}})), \quad g \in \mathcal{Y}. \tag{2.7}$$

Note that $\text{err} = \text{err}_{\mathcal{Y}} \circ F$ if $F_h = F$. Further note that the effective noise level functionals may take negative values. If $\text{err}(\hat{u}_\alpha)$ happens to be negative, this is a favorable event, as it

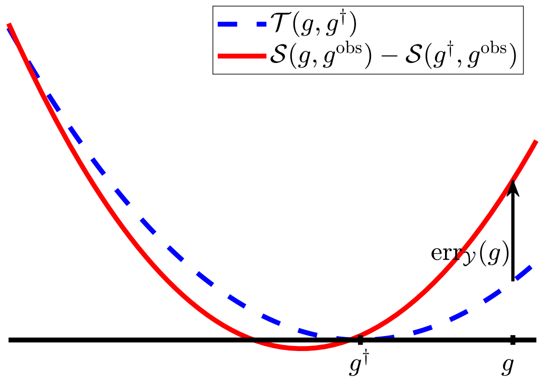

FIG. 2.1. *Illustration of the definition of the effective noise level functional in $\mathcal{Y}$.*

will allow for better bounds on the reconstruction error. However, in all our examples, we can only derive positive bounds on err and $\mathrm{err}_{\mathcal{Y}}$.

The effective noise level functional yields the following estimate for the minimizers $\hat{u}_\alpha$ of generalized Tikhonov regularization (1.1):

$$\begin{aligned} \mathcal{R}(\hat{u}_\alpha) - \mathcal{R}(u^\dagger) &\leq \frac{1}{\alpha}[\mathcal{S}(F_h(u^\dagger); g^{\mathrm{obs}}) - \mathcal{S}(F_h(\hat{u}_\alpha); g^{\mathrm{obs}})] \\ &= \frac{1}{\alpha}[\mathrm{err}(\hat{u}_\alpha) - \mathcal{T}(F(\hat{u}_\alpha); g^\dagger)]. \end{aligned} \tag{2.8}$$

By our assumption that $\mathcal{T}(g^\dagger; g^\dagger) = 0$, this can be rewritten as

$$\frac{1}{\alpha}\mathcal{T}(F(\hat{u}_\alpha); g^\dagger) + \mathcal{R}(\hat{u}_\alpha) \leq \frac{1}{\alpha}\mathcal{T}(F(u^\dagger); g^\dagger) + \mathcal{R}(u^\dagger) + \frac{1}{\alpha}\mathrm{err}(\hat{u}_\alpha),$$

i.e., $\hat{u}_\alpha$ minimizes the ideal Tikhonov functional $u \mapsto (1/\alpha)\mathcal{T}(F(u); g^\dagger) + \mathcal{R}(u)$ up to $(1/\alpha)\,\mathrm{err}(\hat{u}_\alpha)$. As a consequence, wherever this term can be bounded uniformly (i.e., independent of $\hat{u}_\alpha$), then the convergence analysis can be carried out as in the deterministic case. However, in many interesting data models (especially random ones), this is not the case. We will therefore impose the following weaker assumption.

ASSUMPTION B. *Suppose that there exists a Banach space $\mathcal{Y}_{\mathrm{d}} \subseteq \mathcal{Y}$ with $R(F) \subset \mathcal{Y}_{\mathrm{d}}$ and random variables (or constants) $\eta_0, \eta_1 \geq 0$ and a parameter $\mu \geq 1$ such that*

$$\mathrm{err}(u) \leq \eta_0 + \eta_1 \|F(u) - F(u^\dagger)\|_{\mathcal{Y}_{\mathrm{d}}}^{\mu} \quad \textit{for all } u \in D(F). \tag{2.9}$$

Whenever the specific data model for g^{obs} satisfies Assumption B, we call it $\mathcal{Y}_{\mathrm{d}}$**-admissible**.

Note that $\mathcal{Y}_{\mathrm{d}}$-admissibility will allow us to treat deterministic and stochastic noise by the same techniques.

REMARK 2.5. In some situations, it turns out to be useful to separate the noise into more than the two terms η_0 and η_1 as in (2.9). In such situations, we assume the existence of M Banach spaces $\mathcal{Y}_{\mathrm{d}}^m$, $m = 1, \ldots, M$, with $R(F) \subseteq \mathcal{Y}_{\mathrm{d}}^m$ and random variables (or constants) $\eta_1, \ldots, \eta_m \geq 0$ such that

$$\mathrm{err}(u) \leq \eta_0 + \sum_{m=1}^{M} \eta_m \|F(u) - F(u^\dagger)\|_{\mathcal{Y}_{\mathrm{d}}^m}^{\mu_m}, \quad u \in D(F). \tag{2.10}$$

In this case we call the noise model $(\mathcal{Y}_{\mathrm{d}}^1, \ldots, \mathcal{Y}_{\mathrm{d}}^M)$-admissible.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

In this paper, bounds of the form (2.9) and (2.10) will arise from two sources. As approximation rates are determined by the smoothness of the function according to Jackson inequalities, inequalities of the form (2.9) or (2.10) appear in the analysis of numerical approximations of the forward operator F; see (4.2). Furthermore, stochastic noise often belongs to a larger space than $\mathcal{Y}$, and in this case $\mathcal{Y}_{\mathrm{d}}$ will be the dual of this space; see Section 4.2.

2.3. Interpolatory smoothing properties of F. In principle, $\mathcal{Y}_{\mathrm{d}}$-admissibility allows us to upper-bound the effective noise level functional for all $g = F(u),\ u \in D(F)$ by a function depending on the potentially stronger norm $\|F(u) - g^\dagger\|_{\mathcal{Y}_{\mathrm{d}}}$. The latter can then (hopefully) be controlled by interpolation, exploiting smoothing properties of the forward operator F.

This property will be encoded in the following assumption:

ASSUMPTION C1. *Suppose there exists a Banach space $\mathcal{Y}_{\mathrm{d}} \subseteq \mathcal{Y}$ with $R(F) \subset \mathcal{Y}_{\mathrm{d}}$, $\theta \in [0,1]$, and constants $C_\theta > 0$ and $r \geq 1$ such that*

$$\|F(u) - F(u^\dagger)\|_{\mathcal{Y}_{\mathrm{d}}} \leq C_\theta \|F(u) - F(u^\dagger)\|_{\mathcal{Y}}^{\theta}\, \ell(u, u^\dagger)^{(1-\theta)/r} \quad \textit{for all } u \in D(F). \tag{2.11}$$

Whenever the operator F satisfies Assumption C1, then we say that it is $(\mathcal{Y}_{\mathrm{d}}, \theta, \ell)$**-smoothing**.

This assumption requires some discussion. The additional exponent $r \geq 1$ is only to take care of potential exponents in ℓ in case of norm losses. Secondly, we argue that (2.11) is in fact a combination of a smoothing property of F together with intrinsic properties of the corresponding image space:

REMARK 2.6. Suppose that $\mathcal{Y}_{\mathrm{d}} \subseteq \mathcal{Y}$ is a Banach space and $F : D(F) \subset \mathcal{X} \to \mathcal{Y}$ is a continuous forward operator. If there exists a third Banach space $\mathcal{Y}_{F,\ell}$ such that $\mathcal{Y}_{F,\ell} \subset \mathcal{Y}_{\mathrm{d}} \subset \mathcal{Y}$ and F maps Lipschitz continuously to $\mathcal{Y}_{F,\ell}$ with respect to ℓ, i.e., there exist constants $C_L > 0$ and $r \geq 1$ such that

$$\|F(u) - F(u^\dagger)\|_{\mathcal{Y}_{F,\ell}} \leq C_L \ell(u, u^\dagger)^{1/r} \quad \text{for all } u \in D(F), \tag{2.12}$$

and if the spaces $\mathcal{Y}_{F,\ell} \subseteq \mathcal{Y}_{\mathrm{d}} \subseteq \mathcal{Y}$ satisfy a classical interpolation inequality

$$\|g\|_{\mathcal{Y}_{\mathrm{d}}} \leq C \|g\|_{\mathcal{Y}}^{\theta} \|g\|_{\mathcal{Y}_{F,\ell}}^{1-\theta} \quad \text{for all } g \in \mathcal{Y}_{F,\ell} \tag{2.13}$$

with $C > 0$, then F is $(\mathcal{Y}_{\mathrm{d}}, \theta, \ell)$-smoothing. In this case, $C_\theta = C C_L^{1-\theta}$.

Proof. This follows immediately from combining (2.13) with (2.12) as

$$\begin{aligned}\|F(u) - F(u^\dagger)\|_{\mathcal{Y}_{\mathrm{d}}} &\leq C \|F(u) - F(u^\dagger)\|_{\mathcal{Y}}^{\theta} \|F(u) - F(u^\dagger)\|_{\mathcal{Y}_{F,\ell}}^{1-\theta} \\ &\leq C \|F(u) - F(u^\dagger)\|_{\mathcal{Y}}^{\theta} (C_L \ell(u, u^\dagger)^{1/r})^{1-\theta}\end{aligned}$$

for all $u \in D(F)$. ◻

Sometimes it turns out that an additive version of the smoothing property is easier to verify:

ASSUMPTION C2. *Suppose there exists a Banach space $\mathcal{Y}_{\mathrm{d}} \subseteq \mathcal{Y}$, a function $\gamma : [0, \delta_0] \to [0, \infty)$, and a constant $r \geq 1$ such that*

$$\|F(u) - F(u^\dagger)\|_{\mathcal{Y}_{\mathrm{d}}} \leq \gamma(\delta) l(u, u^\dagger)^{1/r} + \frac{1}{\delta} \|F(u) - F(u^\dagger)\|_{\mathcal{Y}} \quad \textit{for all } u \in D(F) \tag{2.14}$$

*and all $\delta \in (0, \delta_0)$. In this case we say that F is **additively** $(\mathcal{Y}_{\mathrm{d}}, \gamma, \ell)$**-smoothing**.*

The additive formulation of the smoothing property allows for functions γ which decay faster to 0 than power-type functions as in (2.11), as we will see at the end of Section 6.

REMARK 2.7. Suppose that $\mathcal{Y}_{\mathrm{d}} \subseteq \mathcal{Y}$ is a Banach space and $F : D(F) \subset \mathcal{X} \to \mathcal{Y}$ is a continuous forward operator. If there exists a third Banach space $\mathcal{Y}_{F,\ell}$ such that $\mathcal{Y}_{F,\ell} \subset \mathcal{Y}_{\mathrm{d}} \subset \mathcal{Y}$ and F maps in fact Lipschitz continuously to $\mathcal{Y}_{F,\ell}$ with respect to ℓ, i.e., there exist constants $C_L > 0$ and $r \geq 1$ such that (2.12) holds true, and the spaces $\mathcal{Y}_{F,\ell} \subset \mathcal{Y}_{\mathrm{d}} \subset \mathcal{Y}$ satisfy an additive interpolation inequality

$$\|g\|_{\mathcal{Y}_{\mathrm{d}}} \leq \tilde{\gamma}(\delta)\|g\|_{\mathcal{Y}_{F,\ell}} + \frac{1}{\delta}\|g\|_{\mathcal{Y}} \quad \text{for all } g \in \mathcal{Y}_{F,\ell} \tag{2.15}$$

and $\delta \in (0, \delta_0)$, then F is additively $(\mathcal{Y}_{\mathrm{d}}, C_L\tilde{\gamma}, \ell)$-smoothing.

Concerning (2.15), we already argued in [19] that, for power-type functions γ, this can be interpreted as an interpolation-type inequality:

- The classical interpolation inequality (2.13), which holds true, e.g., in Sobolev or Besov scales, implies (2.15) with
$$\gamma(\delta) = (1-\theta)C^{1/(1-\theta)}\theta^{\theta/(1-\theta)}\delta^{\theta/(1-\theta)}$$
by Hölder's inequality.
- Vice versa, if $\mathcal{Y}_{F,\ell}$ is continuously embedded in $\mathcal{Y}$ with embedding constant
$$A := \sup_{g \neq 0} \frac{\|g\|_{\mathcal{Y}}}{\|g\|_{\mathcal{Y}_{F,\ell}}},$$
the optimal value of δ in (2.15) for $\gamma(\delta) = \delta^{\theta/(1-\theta)}$ is attained either at
$$\bar{\delta} := \left(\frac{\|g\|_{\mathcal{Y}}}{\|g\|_{\mathcal{Y}_{F,\ell}}(1-\theta)/\theta}\right)^{1-\theta}$$
or at the boundary δ_0. However, the continuity of the embedding yields $\bar{\delta} \leq \delta_{\max} := (A(1-\theta)/\theta)^{1-\theta}$, and hence (2.15) with $\gamma(\delta) = \delta^{\theta/(1-\theta)}$ implies
$$\begin{aligned}
\|g\|_{\mathcal{Y}_{\mathrm{d}}} &\leq \inf_{\delta \in (0,\delta_0)} \left[\delta^{\theta/(1-\theta)}\|g\|_{\mathcal{Y}_{F,\ell}} + \frac{1}{\delta}\|g\|_{\mathcal{Y}}\right] \\
&\leq \frac{\delta_{\max}}{\delta_0} \inf_{\delta \in (0,\delta_{\max})} \left[\delta^{\theta/(1-\theta)}\|g\|_{\mathcal{Y}_{F,\ell}} + \frac{1}{\delta}\|g\|_{\mathcal{Y}}\right] \\
&= \frac{\delta_{\max}}{\delta_0} \left[\bar{\delta}^{\theta/(1-\theta)}\|g\|_{\mathcal{Y}_{F,\ell}} + \frac{1}{\bar{\delta}}\|g\|_{\mathcal{Y}}\right] \\
&= C\|g\|_{\mathcal{Y}}^{\theta}\|g\|_{\mathcal{Y}_{F,\ell}}^{1-\theta}
\end{aligned}$$
for some constant $C > 0$.

REMARK 2.8. In the setting of Remark 2.5, we have (in the multiplicative case) to assume that F is $(\mathcal{Y}_{\mathrm{d}}^m, \theta_m, \ell)$-smoothing for all $1 \leq m \leq M$, i.e., on each $\mathcal{Y}_{\mathrm{d}}^m$ there holds an interpolation inequality of the form

$$\|F(u_1) - F(u_2)\|_{\mathcal{Y}_{\mathrm{d}}^m} \leq C_\theta^{(m)}\|F(u_1) - F(u_2)\|_{\mathcal{Y}}^{\theta_m}\,\ell(u_1, u_2)^{(1-\theta_m)/r} \tag{2.16}$$

for all $u_1, u_2 \in D(F)$ with parameters $\theta_m, C_\theta^{(m)}$. Alternatively, we can also assume that F is additively $(\mathcal{Y}_{\mathrm{d}}^m, \gamma_m, \ell)$-smoothing for all $1 \leq m \leq M$.

3. Error bounds. Now we are in position to derive error bounds for the reconstructions $\hat{u}_\alpha$ in (1.1).

The following theorem is the first main result of this paper. It provides abstract, but also somewhat technical and lengthy, error bounds under the assumptions discussed in Section 2. We will later on discuss several examples in which the error bounds simplify. The proof makes use of Young's inequality with $\varepsilon > 0$, stating that

$$ab \le \varepsilon a^p + \frac{1}{p'}\left(\frac{1}{p\varepsilon}\right)^{p'/p} b^{p'} \tag{3.1}$$

for all $a, b \ge 0,\ p \ge 1$, where $1/p' + 1/p = 1$.

THEOREM 3.1. *Suppose the data model for g^{obs} is $\mathcal{Y}_{\mathrm{d}}$-admissible with $\eta_0, \eta_1 \ge 0$ and parameter $\mu \ge 1$ (cf. Assumption B), and the variational source condition (2.3) is satisfied (cf. Assumption A). If $\eta_1 > 0$, let F additionally be $(\mathcal{Y}_{\mathrm{d}}, \theta, \ell)$-smoothing in the sense of Assumption C1 and suppose there exist constants $C_{\mathcal{T}} \ge 0,\ t \ge 1$, such that*

$$\|g - g^\dagger\|_{\mathcal{Y}} \le C_{\mathcal{T}} \mathcal{T}(g; g^\dagger)^{1/t} \quad \text{for all } g \in R(F). \tag{3.2}$$

Let $\theta\mu < t$.

(a) If $t(1-\theta)\mu < r(t-\mu\theta)$, then each global minimizer $\hat{u}_\alpha$ of (1.1) (if it exists) satisfies the error bound

$$\ell(\hat{u}_\alpha, u^\dagger) \le \frac{4\eta_0}{\alpha} + C\alpha^{-r(t-\mu\theta)/[r(t-\mu\theta)-t(1-\theta)\mu]}\eta_1^{rt/[r(t-\mu\theta)-t(1-\theta)\mu]} + 4\varphi_{\mathrm{app}}(2\alpha)$$

with φ_{app} as in (2.4).

(b) Suppose $t(1-\theta)\mu = r(t-\mu\theta)$. If $t - \mu\theta > 0$, then there exist constants $c, C > 0$ such that for $\eta_1^{t/(t-\mu\theta)} < c\alpha$ each global minimizer $\hat{u}_\alpha$ of (1.1) (if it exists) satisfies the error bound

$$\ell(\hat{u}_\alpha, u^\dagger) \le C\left[\frac{\eta_0}{\alpha} + \varphi_{\mathrm{app}}(2\alpha)\right]$$

with φ_{app} as in (2.4). The same holds true if $t - \mu\theta = 0$ and $C_{\mathcal{T}}\eta < 1$.

Proof. We estimate

$$\begin{aligned}
\ell(\hat{u}_\alpha, u^\dagger) &\overset{(2.3)}{\le} \mathcal{R}(\hat{u}_\alpha) - \mathcal{R}(u^\dagger) + \varphi(\mathcal{T}(F(\hat{u}_\alpha); g^\dagger)) \\
&\overset{(2.8)}{=} \frac{1}{\alpha}\operatorname{err}(\hat{u}_\alpha) - \frac{1}{\alpha}\mathcal{T}(F(\hat{u}_\alpha); g^\dagger) + \varphi(\mathcal{T}(F(\hat{u}_\alpha); g^\dagger)) \\
&\le \frac{1}{\alpha}\operatorname{err}(\hat{u}_\alpha) - \frac{q}{\alpha}\mathcal{T}(F(\hat{u}_\alpha); g^\dagger) + \sup_{\tau \ge 0}\left[\varphi(\tau) - \frac{(1-q)\tau}{\alpha}\right] \\
&= \frac{1}{\alpha}\operatorname{err}(\hat{u}_\alpha) - \frac{q}{\alpha}\mathcal{T}(F(\hat{u}_\alpha); g^\dagger) + \varphi_{\mathrm{app}}\left(\frac{\alpha}{1-q}\right)
\end{aligned}$$

with arbitrary $q \in [0, 1)$. Thus we especially have the inequalities

$$\ell(\hat{u}_\alpha, u^\dagger) \le \frac{1}{\alpha}\operatorname{err}(\hat{u}_\alpha) + \varphi_{\mathrm{app}}(\alpha), \tag{3.3a}$$

$$\mathcal{T}(F(\hat{u}_\alpha); g^\dagger) \le 2\operatorname{err}(\hat{u}_\alpha) + 2\alpha\varphi_{\mathrm{app}}(2\alpha). \tag{3.3b}$$

To treat the effective noise level functional, we combine $\mathcal{Y}_{\mathrm{d}}$-admissibility with the $(\mathcal{Y}_{\mathrm{d}}, \theta, \ell)$-smoothing property of F to obtain

$$\begin{aligned}
\mathrm{err}(\hat{u}_\alpha) &\overset{(2.9)}{\leq} \eta_0 + \eta_1 \|F(\hat{u}_\alpha) - F(u^\dagger)\|_{\mathcal{Y}_{\mathrm{d}}}^{\mu} \\
&\overset{(2.11)}{\leq} \eta_0 + \eta_1 C_\theta^\mu \ell(\hat{u}_\alpha, u^\dagger)^{(1-\theta)\mu/r} \|F(\hat{u}_\alpha) - F(u^\dagger)\|_{\mathcal{Y}}^{\mu\theta} \\
&\overset{(3.2)}{\leq} \eta_0 + \eta_1 C_\theta^\mu C_{\mathcal{T}}^{\mu\theta} \ell(\hat{u}_\alpha, u^\dagger)^{(1-\theta)\mu/r} \mathcal{T}(F(\hat{u}_\alpha); g^\dagger)^{\mu\theta/t} \\
&\overset{(3.3b)}{\leq} \eta_0 + \eta_1 C \ell(\hat{u}_\alpha, u^\dagger)^{(1-\theta)\mu/r} [\mathrm{err}(\hat{u}_\alpha) + \alpha 4 \varphi_{\mathrm{app}}(2\alpha)]^{\mu\theta/t} \\
&\overset{(3.1)}{\leq} \eta_0 + C \eta_1^{t/(t-\theta\mu)} \ell(\hat{u}_\alpha, u^\dagger)^{(1-\theta)\mu t/[r(t-\mu\theta)]} + \tfrac{1}{2} \mathrm{err}(\hat{u}_\alpha) + \tfrac{1}{2} \alpha \varphi_{\mathrm{app}}(2\alpha)
\end{aligned}$$

with some generic constant $C > 0$, where we have used $\varepsilon = \frac{1}{2}$ and $p = t/\theta\mu$ in the last line. Rearranging this implies

$$\mathrm{err}(\hat{u}_\alpha) \leq 2\eta_0 + C \eta_1^{t/(t-\mu\theta)} \ell(\hat{u}_\alpha, u^\dagger)^{t\mu(1-\theta)/(r(t-\mu\theta))} + \alpha \varphi_{\mathrm{app}}(2\alpha). \tag{3.4}$$

Combining this with (3.3a) yields by monotonicity of $\alpha \mapsto \varphi_{\mathrm{app}}(\alpha)$ that

$$\begin{aligned}
\ell(\hat{u}_\alpha, u^\dagger) &\overset{(3.3a)}{\leq} \frac{1}{\alpha} \mathrm{err}(\hat{u}_\alpha) + \varphi_{\mathrm{app}}(\alpha) \\
&\overset{(3.4)}{\leq} \frac{2\eta_0}{\alpha} + \frac{C}{\alpha} \eta_1^{t/(t-\mu\theta)} \ell(\hat{u}_\alpha, u^\dagger)^{t\mu(1-\theta)/(r(t-\mu\theta))} + 2\varphi_{\mathrm{app}}(2\alpha).
\end{aligned}$$

If $t(1-\theta)\mu = r(t - \mu\theta)$, then we can rearrange to

$$\left(1 - \frac{C\eta_1^{t/(t-\mu\theta)}}{\alpha}\right) \ell(\hat{u}_\alpha, u^\dagger) \leq \frac{2\eta_0}{\alpha} + 2\varphi_{\mathrm{app}}(2\alpha).$$

Otherwise, we apply Young's inequality (3.1) with $\varepsilon = \frac{1}{2}$ and $p = r(t - \mu\theta)/(t(1-\theta)\mu)$ to obtain the bound

$$\begin{aligned}
\ell(\hat{u}_\alpha, u^\dagger) \leq{}& \frac{2\eta_0}{\alpha} + \frac{1}{2} \ell(\hat{u}_\alpha, u^\dagger) \\
&+ C \alpha^{-r(t-\theta\mu)/[r(t-\theta\mu)-t(1-\theta)\mu]} \eta_1^{rt/[r(t-\theta\mu)-t(1-\theta)\mu]} + 2\varphi_{\mathrm{app}}(2\alpha),
\end{aligned}$$

which proves the claim.

The assertion for the case $t - \mu\theta = 0$ follows directly from the first equation of this proof if we choose $q > C_{\mathcal{T}} \eta$. ◻

REMARK 3.2. The above theorem can be generalized to the situation of Remarks 2.5 and 2.8. In this case, suppose that $\theta_m \mu_m \leq t$ and $t(1-\theta_m)\mu_m \leq r(t - \mu_m\theta_m)$ for all $1 \leq m \leq M$ and denote

$$I = \{m \in \{1, \ldots, M\} \mid t(1-\theta_m)\mu_m = r(t - \mu_m\theta_m)\}.$$

If $\eta_m^{t/(t-\mu_m\theta_m)}/\alpha$ is sufficiently small for all $m \in I$ with $t - \mu_m\theta_m > 0$ and η_m is sufficiently small for all $m \in I$ with $t - \mu_m\theta_m = 0$, then we obtain the error bound

$$\begin{aligned}
\ell(\hat{u}_\alpha, u^\dagger) \leq{}& \frac{2(M+1)\eta_0}{\alpha} + 2(M+1)4\varphi_{\mathrm{app}}(2\alpha) \\
&+ C \sum_{m \in \{1,\ldots,M\} \setminus I} \alpha^{-r(t-\mu_m\theta_m)/[r(t-\mu_m\theta_m)-t(1-\theta_m)\mu_m]} \\
&\qquad \times \eta_m^{rt/[r(t-\mu_m\theta_m)-t(1-\theta_m)\mu_m]}.
\end{aligned}$$

REMARK 3.3. It is obvious from the proof that one can replace both the upper bound $\|F(u) - F(u^\dagger)\|_{\mathcal{Y}_\mathrm{d}^m}$ in the admissibility assumption (2.10) and the lower bound $\|F(u) - F(u^\dagger)\|_{\mathcal{Y}_\mathrm{d}^m}$ in the smoothingness assumption (2.16) by a common term $\|G_m(u) - G_m(u^\dagger)\|_{\mathcal{Y}_\mathrm{d}^m}$ for some $m \in \{1, \dots, M\}$ with any other operator G_m with domain $D(F)$, and the results of Theorem 3.1 and Remark 3.2 remain valid. This is of particular interest for the case $G = F_h$.

In the case of the additive smoothing property in Assumption C2, we have to restrict ourselves to the case that $t = \mu = 1$. The corresponding result, which is the second main theorem of this paper, is as follows.

THEOREM 3.4. *Suppose the data model for g^{obs} is $\mathcal{Y}_\mathrm{d}$-admissible with $\eta_0, \eta_1 \geq 0$ and parameter $\mu = 1$ (cf. Assumption B), and the variational source condition (2.3) is satisfied (cf. Assumption A). If $\eta_1 > 0$, let F additionally be additively $(\mathcal{Y}_\mathrm{d}, \gamma, \ell)$-smoothing (see Assumption C2) and suppose there exists a constant $C_\mathcal{T} \geq 0$ such that (3.2) holds true with $t = 1$. If $4C_\mathcal{T}\eta_1 < \delta_0$, then each global minimizer $\hat{u}_\alpha$ of (1.1) (if it exists) satisfies the error bound*

$$\ell(\hat{u}_\alpha, u^\dagger) \leq \frac{4\eta_0}{\alpha} + C\left(\frac{\eta_1}{\alpha}\right)^{r/(r-1)} \gamma(4C_\mathcal{T}\eta_1)^{r/(r-1)} + 4\varphi_{\mathrm{app}}(2\alpha).$$

Proof. The estimates (3.3) are still valid, as shown in the proof of Theorem 3.1. To bound the effective noise level functional, we combine $\mathcal{Y}_\mathrm{d}$-admissibility with the additive $(\mathcal{Y}_\mathrm{d}, \gamma, \ell)$-smoothing property of F to obtain

$$\begin{aligned}
\mathrm{err}(\hat{u}_\alpha) &\overset{(2.9)}{\leq} \eta_0 + \eta_1\|F(\hat{u}_\alpha) - F(u^\dagger)\|_{\mathcal{Y}_\mathrm{d}} \\
&\overset{(2.14)}{\leq} \eta_0 + \eta_1\gamma(\delta)\ell(\hat{u}_\alpha, u^\dagger)^{1/r} + \frac{\eta_1}{\delta}\|F(\hat{u}_\alpha) - F(u^\dagger)\|_\mathcal{Y} \\
&\overset{(3.2)}{\leq} \eta_0 + \eta_1\gamma(\delta)\ell(\hat{u}_\alpha, u^\dagger)^{1/r} + C_\mathcal{T}\frac{\eta_1}{\delta}\mathcal{T}(F(\hat{u}_\alpha); g^\dagger) \\
&\overset{(3.3b)}{\leq} \eta_0 + \eta_1\gamma(\delta)\ell(\hat{u}_\alpha, u^\dagger)^{1/r} + 2C_\mathcal{T}\frac{\eta_1}{\delta}[\mathrm{err}(\hat{u}_\alpha) + \alpha\varphi_{\mathrm{app}}(2\alpha)]
\end{aligned}$$

for all $0 < \delta \leq \delta_0$. Setting $\delta = 4C_\mathcal{T}\eta_1$ yields

$$\mathrm{err}(\hat{u}_\alpha) \leq 2\eta_0 + 2\eta_1\gamma(4C_\mathcal{T}\eta_1)\ell(\hat{u}_\alpha, u^\dagger)^{1/r} + \alpha\varphi_{\mathrm{app}}(2\alpha). \tag{3.5}$$

Combining this with (3.3a) we get

$$\begin{aligned}
\ell(\hat{u}_\alpha, u^\dagger) &\overset{(3.3a)}{\leq} \frac{1}{\alpha}\,\mathrm{err}(\hat{u}_\alpha) + \varphi_{\mathrm{app}}(2\alpha) \\
&\overset{(3.5)}{\leq} \frac{2\eta_0}{\alpha} + \frac{2\eta_1}{\alpha}\gamma(4C_\mathcal{T}\eta_1)\ell(\hat{u}_\alpha, u^\dagger)^{1/r} + 2\varphi_{\mathrm{app}}(2\alpha) \\
&\overset{(3.1)}{\leq} \frac{2\eta_0}{\alpha} + C\left(\frac{2\eta_1}{\alpha}\gamma(4C_\mathcal{T}\eta_1)\right)^{r/(r-1)} + \frac{1}{2}\ell(\hat{u}_\alpha, u^\dagger) + 2\varphi_{\mathrm{app}}(2\alpha),
\end{aligned}$$

where we have used $\varepsilon = \frac{1}{2}$ in the last line. Rearranging terms implies the claim. ∎

REMARK 3.5. Also this result can be generalized to the situations of Remarks 2.5 and 2.8 with a result similar to the one described in Remark 3.2. Moreover, it is possible to replace $\|F(u) - F(u^\dagger)\|_{\mathcal{Y}_\mathrm{d}^m}$ in both the admissibility assumption and the smoothingness assumption by $\|G(u) - G(u^\dagger)\|_{\mathcal{Y}_\mathrm{d}}$ for another operator G, e.g., $G = F_h$ in analogy to Remark 3.3.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

4. Examples with noise models in continuous spaces. Let us now discuss different examples to interpret the results of Theorems 3.1 and 3.4 and get an impression of the range of applicability. We start with focusing on continuous data models.

4.1. Operator approximations in classical deterministic noise models. Consider the situation that the datum g^{obs} belongs to the Banach space $\mathcal{Y}$ and satisfies the classical deterministic noise assumption

$$\|g^\dagger - g^{\mathrm{obs}}\|_{\mathcal{Y}} \leq \delta. \tag{4.1}$$

Moreover, we assume that a family of numerical approximations $F_h : D(F) \to \mathcal{Y}$ of F is given satisfying an error bound

$$\|F(u) - F_h(u)\|_{\mathcal{Y}} \leq C_w h^w \|F(u)\|_{\mathcal{Y}_{\mathrm{d}}} \quad \text{for all } u \in D(F), \tag{4.2}$$

some $w > 0$, and a Banach space $\mathcal{Y}_{\mathrm{d}}$ continuously embedded in $\mathcal{Y}$. This type of assumption is satisfied for many commonly used numerical approximation schemes. For example, if $F_h = P_h F$ with some projection operator onto finite element, spline, or wavelet spaces, then (4.2) typically holds true for many pairs of Besov spaces $\mathcal{Y} = B^0_{p,q}(\mathbb{M})$ and $\mathcal{Y}_{\mathrm{d}} = B^s_{p,q}(\mathbb{M})$ with some bounded open domain $\mathbb{M} \subset \mathbb{R}^d$; see, e.g., [9, 12, 28] and the discussion in Section 4.2 below. Estimates of the form (4.2) can also be shown if $F(u^\dagger)$ is a solution to an elliptic PDE, and $F_h(u^\dagger)$ is a finite element solution of the PDE; see [5].

As we are working in a Banach space $\mathcal{Y}$, we follow a general paradigm (see, e.g., [4, 20, 21]) and choose

$$\mathcal{S}(g; g^{\mathrm{obs}}) = \frac{1}{t}\|g - g^{\mathrm{obs}}\|^t_{\mathcal{Y}}, \quad \mathcal{T}(g; g^\dagger) = \frac{2^{1-t}}{t}\|g - g^\dagger\|^t_{\mathcal{Y}}$$

with some $t \in [1, \infty)$. Note the different scaling factors in $\mathcal{S}$ and $\mathcal{T}$, which are needed to establish the following estimates. From the triangle inequality we first obtain

$$\begin{aligned} t\,\mathrm{err}(u) &= 2^{1-t}\|F(u) - F(u^\dagger)\|^t_{\mathcal{Y}} - \|F_h(u) - g^{\mathrm{obs}}\|^t_{\mathcal{Y}} + \|F_h(u^\dagger) - g^{\mathrm{obs}}\|^t_{\mathcal{Y}} \\ &\leq 2^{1-t}\|F(u) - F(u^\dagger)\|^t_{\mathcal{Y}} + \|F_h(u^\dagger) - g^{\mathrm{obs}}\|^t_{\mathcal{Y}} \\ &\quad - \left|\|F(u) - F(u^\dagger)\|_{\mathcal{Y}} - \|(F(u) - F_h(u)) - (F(u^\dagger) - g^{\mathrm{obs}})\|_{\mathcal{Y}}\right|^t. \end{aligned}$$

Now using $|a-b|^t \geq 2^{1-t}a^t - b^t$ and $|a+b|^t \leq 2^{t-1}(a^t + b^t)$ for $a, b \geq 0$, we find

$$t\,\mathrm{err}(u) \leq \left|\|F(u) - F_h(u)\|_{\mathcal{Y}} + \|F(u^\dagger) - g^{\mathrm{obs}}\|_{\mathcal{Y}}\right|^t + \|F_h(u^\dagger) - g^{\mathrm{obs}}\|^t_{\mathcal{Y}},$$

and thus by plugging (4.2) and (4.1) in we finally obtain

$$\begin{aligned} t\,\mathrm{err}(u) &\leq (2^{t-1}+1)\|F_h(u^\dagger) - g^{\mathrm{obs}}\|^t_{\mathcal{Y}} + 2^{t-1}C_w^t h^{wt}\|F(u)\|^t_{\mathcal{Y}_{\mathrm{d}}} \\ &\leq (2^{t-1}+1)(C_w h^w \|g^\dagger\|_{\mathcal{Y}_{\mathrm{d}}} + \delta)^t + 2^{2t-2}C_w^t h^{wt}(\|g^\dagger\|^t_{\mathcal{Y}_{\mathrm{d}}} + \|F(u) - g^\dagger\|^t_{\mathcal{Y}_{\mathrm{d}}}). \end{aligned}$$

Consequently, (2.9) in Assumption B holds true with

$$\eta_0 = c_1 h^{wt}\|g^\dagger\|^t_{\mathcal{Y}_{\mathrm{d}}} + c_2\delta^t, \quad \eta_1 = c_3 h^{wt}, \quad \mu = t,$$

and constants $c_1 = (1/t)C_w^t(2^{2t-1}+2^{t-1})$, $c_2 = (1/t)2^{2t-2}+2^{t-1}$, and $c_3 = (1/t)2^{2t-2}C_w^t$.

Let us now focus on the case $\mathcal{Y} = \mathbf{L}^2(\mathbb{M})$. In the setting of an a-times smoothing operator as in Example 2.2 with $a \geq w$, the variational source condition (2.6) corresponds to $\varphi(\lambda) =$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

$C\rho^{2a/(a+s)}\lambda^{(2/t)u/(a+s)}$ and $\ell(u_1,u_2)=\|u_1-u_2\|^2_{\mathbf{L}^2}$ in Assumption A, equation (2.3), for this choice of $\mathcal{T}$. Hence, $\varphi_{\mathrm{app}}(\alpha)=\tilde{C}\rho^{2ta/(ta-(2-t)s)}\alpha^{2s/(ta-(2-t)s)}$.

Complementing assumption (2.5) ("F is at most a-times smoothing"), we also assume that F is at least a-times smoothing in the sense that

$$\|F(u)-F(u^\dagger)\|_{H^a}\le C_F\|u-u^\dagger\|_{\mathbf{L}^2}\quad \text{for all } u\in D(F). \tag{4.3}$$

By Sobolev interpolation we have

$$\begin{aligned}\|F(u)-F(u^\dagger)\|_{H^w} &\le \|F(u)-F(u^\dagger)\|^{(a-w)/a}_{\mathbf{L}^2}\|F(u)-F(u^\dagger)\|^{w/a}_{H^a}\\ &\le C_F^{w/a}\|F(u)-F(u^\dagger)\|^{(a-w)/a}_{\mathbf{L}^2}\|u-u^\dagger\|^{w/a}_{\mathbf{L}^2}\end{aligned}$$

for $0<w\le a$ such that Assumption C1 holds true with $\mathcal{Y}_{\mathrm{d}}=H^w(\mathbb{M})$, $\theta=(a-w)/a$, and $r=2$.

We have now verified all assumptions of Theorem 3.1. (Note, in particular, that $\theta\mu<t$ as $\mu=t$ and $w>0$.) Therefore, we obtain the error bound

$$\|\hat{u}_\alpha-u^\dagger\|^2_{\mathbf{L}^2}\le\begin{cases} C\left[\dfrac{\delta^t+h^{wt}}{\alpha}+\left(\dfrac{h^{at}}{\alpha}\right)^{2/(2-t)}\right.\\ \qquad\left.+\rho^{2ta/(ta-(2-t)s)}\alpha^{2s/(ta-(2-t)s)}\right], & \text{if } t<2,\\ C\left[\dfrac{\delta^2+h^{2w}}{\alpha}+\rho^2\alpha^{s/a}\right], & \text{if } t=2 \text{ and } h^{2a}\le c\alpha.\end{cases}$$

For $h=0$, i.e., $F_h=F$, the parameter choice rule

$$\alpha_*\sim\rho^{-2a/(a+s)}\delta^{(ta-(2-t)s)/(a+s)} \tag{4.4}$$

yields the rate

$$\|\hat{u}_{\alpha_*}-u^\dagger\|^2_{L^2}=\mathcal{O}(\rho^{2a/(s+a)}\delta^{2s/(s+a)})\quad\text{as } \delta\searrow 0, \tag{4.5}$$

which is well known to be optimal; see, e.g., [18]. This rate is also achieved if

$$h^w\le C_h\delta. \tag{4.6}$$

In fact, for the first terms in the error bound, (4.6) yields $h^{tw}=\mathcal{O}(\delta^t)$, and for $t=2$ we have $h^{2a}/\alpha_*=\mathcal{O}(\delta^p)$ with $p=2a/w-2a/(a+s)>0$. Similarly, for $t<2$ we find that the term $(h^{at}/\alpha_*)^{2/(2-t)}=\mathcal{O}(\delta^{p'})$ with $p'=(2at/(2-t))(1/w-1/(a+s))+2s/(a+s)$ is of higher order in δ and hence negligible as $w\le a<a+s$.

We summarize our results in the following corollary.

COROLLARY 4.1. *Suppose the operator $F:D(F)\subset\mathbf{L}^2(\Omega)\to\mathbf{L}^2(\mathbb{M})$ is a-times smoothing in the sense that (2.5) and (4.3) hold true, $u^\dagger\in H^s(\Omega)$ with $0<s<a$, and the data $g^{\mathrm{obs}}\in\mathbf{L}^2(\mathbb{M})$ and numerical approximation $F_h:D(F)\to\mathbf{L}^2(\mathbb{M})$ satisfy (4.1) and (4.2) with $0<w\le a$. Then the estimator*

$$\hat{u}_\alpha\in\operatorname*{argmin}_{u\in D(F)}\left[\frac{1}{\alpha t}\|F_h(u)-g^{\mathrm{obs}}\|^t_{\mathbf{L}^2}+\frac{1}{2}\|u\|^2_{\mathbf{L}^2}\right]$$

(if it exists) with $t\le 2$, $\alpha=\alpha_*$ given by (4.4), and h satisfying (4.6) obeys the optimal error bound (4.5).

ETNA
Kent State University and
Johann Radon Institute (RICAM)

REMARK 4.2. Let us compare our assumption (4.2) to the standard assumption

$$\|T - T_h\|_{L(\mathcal{X},\mathcal{Y})} \leq h \tag{4.7}$$

that is often used in Hilbert space analysis; see, e.g., [16, 24, 25]. Note that, for a linear operator F, (4.2) implies $\|F - F_h\|_{L(\mathcal{X},\mathcal{Y})} \leq c_w h^w \|F\|_{L(\mathcal{X},\mathcal{Y}_\mathrm{d})}$, i.e., (4.7) is weaker than our assumption (4.2). On the other hand, as argued above, (4.2) is still easy to verify in many practical examples, and (4.7) is much more difficult to deal with. In particular, one needs to require the function φ in the (spectral) source condition to be operator monotone on some interval in order to obtain an inequality $\|\varphi(A) - \varphi(B)\| \leq C\varphi(\|A - B\|) + C'\|A - B\|$ for any self-adjoint operators A and B; see [24].

4.2. Gaussian white noise. Let $\mathcal{Y}$ be a Hilbert space. We now consider inverse problems with right-hand sides perturbed by Gaussian white noise processes Z, which often appear as limits of discrete noise processes. This is a Hilbert space process Z on $\mathcal{Y}$ satisfying $\mathbb{E}[Z] = 0$ and $\mathbf{Cov}[Z] = \mathrm{id}_\mathcal{Y}$ such that, for each finite number of elements $g_1, \ldots, g_n \in \mathcal{Y}$, the vector

$$(\langle g_1, Z\rangle, \ldots, \langle g_n, Z\rangle) \in \mathbb{R}^n$$

is a multivariate normal random vector. Precisely, the available data g^{obs} is assumed to be of the form

$$g^{\mathrm{obs}} = g^\dagger + \sigma Z. \tag{4.8}$$

If $\dim(\mathcal{Y}) = \infty$, then $Z \notin \mathcal{Y}$ holds with probability 1, and thus the model has (in general) to be understood in a weak sense; see, e.g., [3]. This implies that we have (only) access to dual products of the form

$$\langle g, g^{\mathrm{obs}}\rangle = \langle g, g^\dagger\rangle + \sigma\langle g, Z\rangle$$

for all $g \in \mathcal{Y}$, and thus the choice

$$\mathcal{S}(g; g^{\mathrm{obs}}) := \tfrac{1}{2}\|g\|_\mathcal{Y}^2 - \langle g, g^{\mathrm{obs}}\rangle, \quad \mathcal{T}(g; g^\dagger) := \tfrac{1}{2}\|g - g^\dagger\|_\mathcal{Y}^2 \tag{4.9}$$

ensures well-definedness of the data fidelity term. Note that this choice of $\mathcal{S}$ equals the negative log-likelihood ratio in the Cameron–Martin–Girsanov sense; cf. [42]. In the finite-dimensional case, this equals the classical negative log-likelihood term up to the constant $\|g^{\mathrm{obs}}\|_\mathcal{Y}^2$. For $\mathrm{err}_\mathcal{Y}$ as in (2.7), we compute

$$\begin{aligned}
\mathrm{err}_\mathcal{Y}(g) &= \tfrac{1}{2}\|g - g^\dagger\|_\mathcal{Y}^2 - \tfrac{1}{2}\|g\|_\mathcal{Y}^2 + \langle g^{\mathrm{obs}}, g\rangle + \tfrac{1}{2}\|g^\dagger\|_\mathcal{Y}^2 - \langle g^{\mathrm{obs}}, g^\dagger\rangle \\
&= \|g^\dagger\|_\mathcal{Y}^2 - \langle g, g^\dagger\rangle + \langle g^\dagger + \sigma Z, g - g^\dagger\rangle \\
&= \sigma\langle Z, g - g^\dagger\rangle.
\end{aligned}$$

In the following, let $\mathbb{M}$ denote a smooth bounded domain $\mathbb{M} \subset \mathbb{R}^d$ and $\mathcal{Y} = \mathbf{L}^2(\mathbb{M})$. By Theorem A.1, Gaussian white noise Z on $\mathbf{L}^2(\mathbb{M})$ is a.s. contained in the Besov space $\mathcal{Y}_\mathrm{d}^* = B_{p,\infty}^{-d/2}(\mathbb{M})$ with arbitrary $p \in [1, \infty)$. For the dual spaces

$$\mathcal{Y}_\mathrm{d} = \tilde{B}_{p',1}^{d/2}(\mathbb{M})$$

(note that, due to (A.2), we can also take $B_{p',1}^{d/2}(\mathbb{M})$ if $d/2 - 1/p' \notin \mathbb{Z}$), the following interpolation inequality for any $q \in [1, \infty]$ and $1/p' + 1/p = 1$ holds true by K-interpolation

ETNA
Kent State University and
Johann Radon Institute (RICAM)

theory (see [36, Sections 4.3.2 and 4.11.1])

$$\begin{aligned}\|g\|_{B^{d/2}_{p',1}(\mathbb{M})} &\le C\|g\|^{1-d/(2a)}_{B^0_{p',2}(\mathbb{M})}\|g\|^{d/(2a)}_{B^a_{p',q}(\mathbb{M})}\\ &\le C\|g\|^{1-d/(2a)}_{\mathbf{L}^2(\mathbb{M})}\|g\|^{d/(2a)}_{B^a_{p',q}(\mathbb{M})},\end{aligned}$$

where the second inequality follows by Besov embedding theorems [36, Section 4.6.2] for $p \in [2,\infty)$. This yields the following corollary with error bounds which for an a-smoothing operator are well known to be order-optimal in a minimax sense; see, e.g., [41].

COROLLARY 4.3. *Let $\mathcal{Y} = \mathbf{L}^2(\mathbb{M})$, consider the noise model (4.8) with a Gaussian white noise Z, suppose that $\mathcal{S}$ and $\mathcal{T}$ are given by (4.9), and that $u^\dagger \in D(F)$ satisfies the variational source condition (2.3). Assume furthermore that $F : D(F) \subset \mathcal{X} \to \mathcal{Y}$ satisfies*

$$\|F(u_1) - F(u_2)\|_{B^a_{p',q}(\mathbb{M})} \le C_{\mathrm{L}}\ell(u_1,u_2)^{1/r} \tag{4.10}$$

for some $p \in [2,\infty)$ with $d/2 - 1/p' \notin \mathbb{Z}$, $q \in [1,\infty]$, $r > 2d/(2a+d)$, $C_{\mathrm{L}} > 0$, and all $u_1, u_2 \in D(F)$. If $F_h = F$ on $D(F)$ and $\mathcal{Y}_h = \mathcal{Y}$, then we obtain for every global minimizer $\hat{u}_\alpha$ of the Tikhonov functional (1.1) the a.s. error bound

$$\ell(\hat{u}_\alpha, u^\dagger) \le C\alpha^{-(2a+d)/(2a+d-2d/r)}(\sigma\|Z\|_{B^{-d/2}_{p,\infty}(\mathbb{M})})^{4a/(2a+d(1-2/r))} + 4\varphi_{\mathrm{app}}(2\alpha).$$

In the situation of Example 2.2, i.e., under (2.5) and (4.3), $\mathcal{R}(u) = \frac{1}{2}\|u\|^2_{\mathcal{X}}$, $\ell(u,u^\dagger) = \frac{1}{4}\|u-u^\dagger\|^2_{\mathcal{X}}$, and $\|u\|_{H^s} \le \rho$ with $0 < s < a$, we obtain by $B^a_{2,2}(\mathbb{M}) = H^a(\mathbb{M})$ the upper bound

$$\|\hat{u}_\alpha - u^\dagger\|^2_{\mathcal{X}} \le C[\alpha^{-(2a+d)/(2a)}(\sigma\|Z\|_{B^{-d/2}_{2,\infty}(\mathbb{M})})^2 + \rho^2\alpha^{s/a}] \tag{4.11}$$

a.s. For the choice $\alpha \sim (\sigma/\rho)^{4a/(2s+2a+d)}$, this yields the order-optimal rate of convergence

$$\mathbb{E}[\|\hat{u}_\alpha - u^\dagger\|^2_{\mathcal{X}}]^{1/2} = \mathcal{O}(\rho^{(2a+d)/(s+a+d/2)}\sigma^{s/(s+a+d/2)}) \quad \text{as } \sigma \to 0.$$

Proof. The claimed error bounds follow directly from Theorem 3.1. According to [38, Corollary 3.7], the random variable $\|Z\|_{B^{-d/2}_{2,\infty}(\mathbb{T}^d)}$ has finite second moment on the torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$. By Remark A.2 this also holds true for $\|Z\|_{B^{-d/2}_{2,\infty}(\mathbb{M})}$. As a consequence, we can take the expectation in (4.11) to obtain the final statement on the rate of convergence in expectation. ∎

REMARK 4.4. The above result can be generalized to arbitrary (i.e., non-Gaussian) Hilbert space processes Z on $\mathcal{Y}$ with $\|\mathbf{Cov}[Z]\|_{\mathcal{Y}\to\mathcal{Y}} \le 1$ as long as a space $\mathcal{Y}_{\mathrm{d}}$ is known such that $\|Z\|_{\mathcal{Y}'_{\mathrm{d}}}$ is a.s. finite and has finite moments.

4.3. Operator approximations in statistical inverse problems. In this section, we study statistical inverse problems with noisy operator

$$g^{\mathrm{obs}} = F_h(u^\dagger) + \sigma Z, \tag{4.12}$$

where Z is white noise on $\mathcal{Y} := \mathbf{L}^2(\mathbb{M})$ with some bounded open domain $\mathbb{M} \subset \mathbb{R}^d$, and the approximations of the forward operator are of the form

$$F_h = P_n F_\eta,$$

with a sequence of orthogonal projection operators P_n in $\mathbf{L}^2(\mathbb{M})$ and some approximation $F_\eta \approx F$, e.g., $F_\eta = FQ_n$ with projections operators Q_n in $\mathcal{X}$. Here F_η may also involve other approximations or errors, and it may be nonlinear.

For many useful projection operators P_n (e.g., by finite elements), the range of P_n is not contained in the range of F, and therefore it would be very restrictive to impose the condition (4.10) on F_h instead of on F. However, we will assume that

$$\|F_\eta(u_1) - F_\eta(u_2)\|_{H^a(\mathbb{M})} \leq C_a \|F(u_1) - F(u_2)\|_{H^a(\mathbb{M})}. \tag{4.13}$$

This condition is obviously satisfied if $F_\eta = FQ_n$ with uniformly bounded projection operators Q_n. We further assume the Jackson-type inequality

$$\|(I - P_n)g\|_{\mathcal{Y}} \leq C_P n^{-a} \|g\|_{H^a(\mathbb{M})} \quad \text{for all } g \in H^a(\mathbb{M}). \tag{4.14}$$

Such inequalities, e.g., for trigonometric or spline projections, are standard results from approximation theory; see, e.g., [30, Theorem 2.26 and Corollary 2.47].

We also assume an approximation property with respect to η:

$$\|F_\eta(u) - F(u)\|_{\mathcal{Y}} \leq C_w \eta^w \|u\|_{\mathcal{X}}. \tag{4.15}$$

Note that if $F_\eta = FQ_n$ with orthogonal projections Q_n, then (4.15) follows from Jackson-type inequalities for Q_n applied to the adjoint operators.

Due to the Hilbert space structure and the white noise assumption on Z, it seems appropriate to choose $\mathcal{S}$ and $\mathcal{T}$ as in (4.9), i.e.,

$$\mathcal{S}(g; g^{\mathrm{obs}}) := \tfrac{1}{2}\|g\|_{\mathcal{Y}}^2 - \langle g, g^{\mathrm{obs}}\rangle, \quad \mathcal{T}(g; g^\dagger) := \tfrac{1}{2}\|g - g^\dagger\|_{\mathcal{Y}}^2. \tag{4.16}$$

This yields

$$\begin{aligned}
\mathrm{err}(u) &= \tfrac{1}{2}\|F(u) - F(u^\dagger)\|_{\mathcal{Y}}^2 - \tfrac{1}{2}\|F_h(u)\|_{\mathcal{Y}}^2 + \langle g^{\mathrm{obs}}, F_h(u)\rangle \\
&\quad + \tfrac{1}{2}\|F_h(u^\dagger)\|_{\mathcal{Y}}^2 - \langle g^{\mathrm{obs}}, F_h(u^\dagger)\rangle \\
&= \tfrac{1}{2}\|F(u) - F(u^\dagger)\|_{\mathcal{Y}}^2 - \tfrac{1}{2}\|F_h(u) - F_h(u^\dagger)\|_{\mathcal{Y}}^2 + \sigma\langle Z, F_h(u) - F_h(u^\dagger)\rangle.
\end{aligned} \tag{4.17}$$

To estimate the last term, we move P_n in $F_h = P_n F_\eta$ to the left-hand side. In order to show that $\sup_n \|P_n^* Z\|_{B_{2,\infty}^{-d/2}} < \infty$, we impose the condition

$$C_P := \sup_n \|P_n\|_{B_{p,\infty}^{d/2}(\mathbb{M}) \to B_{p,\infty}^{d/2}(\mathbb{M})} < \infty. \tag{4.18}$$

This condition can, e.g., on the torus $\mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$, be shown by interpolation for orthogonal trigonometric projection operators P_n (as $\sup_n \|P_n\|_{H_p^s(\mathbb{T}^d)} < \infty$ for all $s \in \mathbb{R}$ and $(H_p^s(\mathbb{T}^d), H_p^t(\mathbb{T}^d))_{\theta,\infty} = B_{p,\infty}^{(1-\theta)s+\theta t}(\mathbb{T}^d)$), but also in an analogous manner for hierarchical finite element discretizations (see [28, Theorem 15]), or more directly for wavelet projections truncating all wavelet coefficients above level n using Besov norms in terms of wavelet coefficients; see [12, Section 4.3.1]. Then we obtain

$$\sigma|\langle Z, F_h(u) - F_h(u^\dagger)\rangle| \leq \sigma C_P \|Z\|_{B_{2,\infty}^{-d/2}} \|F_\eta(u) - F_\eta(u^\dagger)\|_{B_{2,1}^{d/2}}.$$

Now the first two terms in (4.17) can be expanded as

$$\begin{aligned}
&\tfrac{1}{2}\|F(u) - F(u^\dagger)\|_{\mathcal{Y}}^2 - \tfrac{1}{2}\|F_h(u) - F_h(u^\dagger)\|_{\mathcal{Y}}^2 \\
&\quad = \tfrac{1}{2}\langle F(u) - F(u^\dagger) - F_h(u) + F_h(u^\dagger),\ F(u) - F(u^\dagger) + F_h(u) - F_h(u^\dagger)\rangle \\
&\quad = \tfrac{1}{2}\langle (I - P_n)(F(u) - F(u^\dagger)) + P_n(F(u) - F(u^\dagger) - F_\eta(u) + F_\eta(u^\dagger)), \\
&\qquad\quad 2(F(u) - F(u^\dagger)) + (P_n - I)(F_\eta(u) - F_\eta(u^\dagger)) \\
&\qquad\quad + F_\eta(u) - F_\eta(u^\dagger) - (F(u) - F(u^\dagger))\rangle.
\end{aligned}$$

Using the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}
&\tfrac{1}{2}\|F(u)-F(u^\dagger)\|_{\mathcal{Y}}^2-\tfrac{1}{2}\|F_h(u)-F_h(u^\dagger)\|_{\mathcal{Y}}^2\\
&\quad\le \tfrac{1}{2}\|(I-P_n)(F(u)-F(u^\dagger))+P_n(F(u)-F(u^\dagger)-F_\eta(u)+F_\eta(u^\dagger))\|_{\mathcal{Y}}\\
&\qquad\times\|2(F(u)-F(u^\dagger))+(P_n-I)(F_\eta(u)-F_\eta(u^\dagger))\\
&\qquad\qquad+F_\eta(u)-F_\eta(u^\dagger)-(F(u)-F(u^\dagger))\|_{\mathcal{Y}}.
\end{aligned}$$

Both terms in this product can be estimated by means of (4.14), namely as

$$\begin{aligned}
&\|(I-P_n)(F(u)-F(u^\dagger))+P_n(F(u)-F(u^\dagger)-F_\eta(u)+F_\eta(u^\dagger))\|_{\mathcal{Y}}\\
&\quad\le C_P n^{-a}\|F(u)-F(u^\dagger)\|_{H^a(\mathbb{M})}+\|F(u)-F_\eta(u)\|_{\mathcal{Y}}+\|F(u^\dagger)-F_\eta(u^\dagger)\|_{\mathcal{Y}}\\
&\quad\le C_P n^{-a}\|F(u)-F(u^\dagger)\|_{H^a(\mathbb{M})}+C_\eta\eta^w(\|u\|_{\mathcal{X}}+\|u^\dagger\|_{\mathcal{X}}),
\end{aligned}$$

and by (4.13) furthermore

$$\begin{aligned}
&\|2(F(u)-F(u^\dagger))+(P_n-I)(F_\eta(u)-F_\eta(u^\dagger))\\
&\qquad+F_\eta(u)-F_\eta(u^\dagger)-(F(u)-F(u^\dagger))\|_{\mathcal{Y}}\\
&\quad\le 2\|F(u)-F(u^\dagger)\|_{\mathcal{Y}}+C_P n^{-a}\|F_\eta(u)-F_\eta(u^\dagger)\|_{H^a(\mathbb{M})}\\
&\qquad+\|F(u)-F_\eta(u)\|_{\mathcal{Y}}+\|F(u^\dagger)-F_\eta(u^\dagger)\|_{\mathcal{Y}}\\
&\quad\le 2\|F(u)-F(u^\dagger)\|_{\mathcal{Y}}+C_P C_a n^{-a}\|F(u)-F(u^\dagger)\|_{H^a(\mathbb{M})}\\
&\qquad+C_\eta\eta^w(\|u\|_{\mathcal{X}}+\|u^\dagger\|_{\mathcal{X}}).
\end{aligned}$$

If we now assume furthermore that F is at most a smoothing in the sense that

$$\|u-v\|_{\mathcal{X}}\le C_F\|F(u)-F(v)\|_{H^a(\mathbb{M})}\quad\text{for all } u,v\in D(F) \tag{4.19}$$

(in addition to (2.5)), then we can estimate

$$\|u\|_{\mathcal{X}}\le\|u-u^\dagger\|_{\mathcal{X}}+\|u^\dagger\|_{\mathcal{X}}\le\|u^\dagger\|_{\mathcal{X}}+C_F\|F(u)-F(u^\dagger)\|_{H^a(\mathbb{M})}$$

and obtain by (3.1) with $p=p'=2$ as well as $\left(\sum_{j=1}^n a_j\right)^2\le n\sum_{j=1}^n a_j^2$ for $n=2,3$ the overall bound

$$\begin{aligned}
&\frac{1}{2}\|F(u)-F(u^\dagger)\|_{\mathbf{L}^2(\mathbb{M})}^2-\frac{1}{2}\|F_h(u)-F_h(u^\dagger)\|_{\mathbf{L}^2(\mathbb{M})}^2\\
&\quad\le\frac{C}{\varepsilon}(n^{-2a}+\eta^{2w})\|F(u)-F(u^\dagger)\|_{H^a(\mathbb{M})}^2+\frac{C}{\varepsilon}\eta^{2w}\|u^\dagger\|_{\mathcal{X}}^2+\varepsilon\|F(u)-F(u^\dagger)\|_{\mathbf{L}^2(\mathbb{M})}^2
\end{aligned}$$

for some constant $C>0$ and some $\varepsilon\in(0,1)$ to be determined later.

We summarize the setting of Theorem 3.1 in its generalization of Remark 3.2 in Table 4.1.

Note that for $m=1$ we use Remark 3.3 with $G=F_\eta$. As $t=2$, in order to meet the condition $t(1-\theta_m)\mu_m\ge r(t-\mu_m\theta_m)$ for $m=2$, we must have $r\ge 2$. For simplicity, we only consider the case $r=2$. Then $t(1-\theta_2)\mu_2=r(t-\mu_2\theta_2)$ and $t(1-\theta_3)\mu_3=r(t-\mu_3\theta_3)$, so we have to meet the additional constraints $\eta_2\lesssim\alpha$ and $\eta_3\lesssim 1$ which hold true if $\max\{\eta^{2w}+n^{-2a}\}\alpha^{-1}\lesssim\varepsilon\lesssim 1$. Choosing $\varepsilon=\eta_3>0$ sufficiently small, we obtain the constraint $\alpha\gtrsim\max\{\eta^{2w}+n^{-2a}\}$ on the choice of α.

COROLLARY 4.5. *Let $\mathcal{X}=\mathcal{Y}=\mathbf{L}^2(\mathbb{M})$ and consider the estimator (1.1) with $\mathcal{R}(u)=\frac{1}{2}\|u\|_{\mathbf{L}^2}^2$ and $\mathcal{S}$ given by (4.16) for the noise model (4.12). Suppose that F satisfies (4.10) and (4.19) with $B_{p,q}^a=B_{2,2}^a=H^a$, $\ell(u_1,u_2)=\|u_1-u_2\|_{\mathbf{L}^2}^2$, and $r=2$,* *assume that the approximation $F_h = P_n F_\eta$ to F satisifies* (4.13)*,* (4.14)*,* (4.15)*,* (4.18)*, and that $u^\dagger \in H^s(\Omega)$ with $\|u^\dagger\|_{H^s} \leq \rho$ for some $s \in (0, a)$ and $\rho > 0$. If α is chosen such that $\alpha \gtrsim \max\{\eta^{2w} + n^{-2a}\}$, then we obtain the error bound*

$$\|\hat{u}_\alpha - u^\dagger\|_{\mathbf{L}^2}^2 \leq C\left[\rho^2 \alpha^{s/a} + \alpha^{-2a/(2a+d)} (\sigma \|Z\|_{B^{-d/2}_{2,\infty}})^2 + \frac{\eta^{2w}}{\alpha}\right]. \tag{4.20}$$

In particular, if

$$n^{-2a} \lesssim \alpha \sim \left(\frac{\sigma}{\rho}\right)^{4a/(2a+2s+d)} \quad \textit{and} \quad \eta^{2w} \lesssim \rho^{(d-2a)/(2s+2a+d)} \sigma^{(2s+4a)/(2s+2a+d)},$$

then the second term in (4.20) *dominates the first one (i.e., the errors in the operator are negligible), and the reconstruction error tends to* 0 *with the rate*

$$\mathbb{E}[\|\hat{u}_\alpha - u^\dagger\|_{\mathbf{L}^2}^2]^{1/2} = \mathcal{O}(\rho^{(a+d/2)/(s+a+d/2)} \sigma^{s/(s+a+d/2)}) \quad \textit{as } \sigma \to 0.$$

Note that the latter rate equals the optimal rate of convergence under white noise, as discussed in Corollary 4.3.

TABLE 4.1
Verifications of the assumptions of Theorem 3.1/ Remark 3.2 for Corollary 4.5.

Term m	Space $\mathcal{Y}_{\mathrm{d},m}$	Power μ_m	Interpol. par. θ_m	Error η_m
0	—	—	—	$(C/\varepsilon)\eta^{2w}\|u^\dagger\|_{\mathcal{X}}^2$
1	$B^{d/2}_{2,1}(\mathbb{M})$	1	$1 - d/(2a)$	$\sigma C_P \|Z\|_{B^{-d/2}_{2,\infty}}$
2	$H^a(\mathbb{M})$	2	0	$(C/\varepsilon)(n^{-2a} + \eta^{2w})$
3	$\mathcal{Y} = \mathbf{L}^2(\mathbb{M})$	2	1	ε

5. Discrete sampled data. Let us continue our discussion of the applicability of Theorems 3.1 and 3.4 with a focus on discrete noise models.

In many applications, the data does actually consist of samples of a function $g^\dagger$. To model this mathematically, suppose that $\mathbb{M} \subset \mathbb{R}^d$ is a bounded Lipschitz domain and suppose that $g^\dagger \in \mathscr{C}(\mathbb{M})$. Let $x_1^{(n)}, \ldots, x_n^{(n)} \in \mathbb{M}$ be points and consider the data model

$$g^{\mathrm{obs}} \in \mathbb{R}^n, \quad g_i^{\mathrm{obs}} = g^\dagger(x_i^{(n)}) + \varepsilon_i, \quad i = 1, \ldots, n \tag{5.1}$$

with scalar noise ε_i. Suppose that

$$Q_n\varphi := \sum_{i=1}^n \alpha_i^{(n)} \varphi(x_i^{(n)}) \approx Q\varphi := \int_{\mathbb{M}} \varphi(x)\,\mathrm{d}x$$

is a quadrature rule and choose

$$\mathcal{S}(g; g^{\mathrm{obs}}) := \frac{1}{2}\|g\|_{\mathbf{L}^2(\mathbb{M})}^2 - \sum_{i=1}^n \alpha_i^{(n)} g_i^{\mathrm{obs}} g(x_i^{(n)}), \tag{5.2a}$$

$$\mathcal{T}(g; g^\dagger) := \frac{1}{2}\|g - g^\dagger\|_{\mathbf{L}^2(\mathbb{M})}^2. \tag{5.2b}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

This choice includes the discretization effects contained in the model (5.1) and, besides that, is similar to (4.9). From that point of view, the data is handled in a least-squares sense, which corresponds to the negative log-likelihood functional if ε_i are independent random normals.

Now we compute that

$$
\begin{aligned}
\mathrm{err}_{\mathcal{Y}}(g) &= \frac{1}{2}\|g^\dagger\|^2 - \frac{1}{2}\|g\|^2 + \frac{1}{2}\|g-g^\dagger\|^2 + \sum_{i=1}^n \alpha_i g_i^{\mathrm{obs}}(g(x_i^{(n)}) - g^\dagger(x_i^{(n)})) \\
&= \frac{1}{2}\|g^\dagger\|^2 - \frac{1}{2}\|g\|^2 + \frac{1}{2}\|g-g^\dagger\|^2 + \langle g^\dagger, g-g^\dagger\rangle \\
&\quad + (Q_n - Q)(g^\dagger(g-g^\dagger)) + \sum_{i=1}^n \alpha_i^{(n)}\varepsilon_i(g(x_i^{(n)}) - g^\dagger(x_i^{(n)})) \\
&= (Q_n - Q)(g^\dagger(g-g^\dagger)) + \sum_{i=1}^n \alpha_i^{(n)}\varepsilon_i(g(x_i^{(n)}) - g^\dagger(x_i^{(n)})). \qquad (5.3)
\end{aligned}
$$

This shows that the effective noise level functional can once again be bounded by a sum of two error terms as in Remark 3.2. The first term on the right-hand side of (5.3) can be controlled depending on the order of the quadrature rule as long as the function $g^\dagger(g-g^\dagger)$ is sufficiently smooth. Therefore, assume that $g^\dagger \in \mathscr{C}^s(\mathbb{M})$ with some $s \in \mathbb{N}$, $s > d/2$, and choose $\mathcal{Y}_{\mathrm{d}}^1 = H^s(\mathbb{M})$. If the quadrature rule furthermore satisfies

$$
|Q_n(g) - Q(g)| \le C_Q n^{-\kappa}\|g\|_{H^s(\mathbb{M})} \quad \text{for all } g \in H^s(\mathbb{M}), \qquad (5.4)
$$

then it follows from the standard estimate

$$
\|g^\dagger(g-g^\dagger)\|_{H^s(\mathbb{M})} \le M\|g^\dagger\|_{\mathscr{C}^s(\mathbb{M})}\|g-g^\dagger\|_{H^s(\mathbb{M})}
$$

that the first term in (5.3) can be bounded by

$$
(Q_n - Q)(g^\dagger(g-g^\dagger)) \le n^{-\kappa}\|g^\dagger(g-g^\dagger)\|_{H^s(\mathbb{M})} \le M\|g^\dagger\|_{\mathscr{C}^s(\mathbb{M})}\|g-g^\dagger\|_{H^s(\mathbb{M})}.
$$

To treat the second term, we recall from the uniform boundedness principle that a necessary condition for convergence, $\lim_{n\to\infty} Q_n\varphi = Q\varphi$ for all $\varphi \in \mathscr{C}(\Omega)$, is that

$$
A := \sup_n \sum_{i=1}^n |\alpha_i^{(n)}| < \infty. \qquad (5.5)
$$

This gives the following lemma.

LEMMA 5.1. *Let $\mathbb{M} \subset \mathbb{R}^d$ be a bounded Lipschitz domain. Suppose that $g^\dagger \in \mathscr{C}^s(\mathbb{M})$ with some $s \in \mathbb{N}$ such that $s > d/2$ and let Q_n be a quadrature rule obeying (5.5) and (5.4). Then the data model (5.1) yields the estimate*

$$
\mathrm{err}_{\mathcal{Y}}(g) \le n^{-\kappa} C_Q M\|g^\dagger\|_{\mathscr{C}^s(\mathbb{M})}\|g^\dagger - g\|_{H^s(\mathbb{M})} + A\delta\|g^\dagger - g\|_{\mathbf{L}^\infty(\mathbb{M})}
$$

with $\delta := \max_{1\le i\le n}|\epsilon_i|$.

As a next step, we need to verify the interpolation inequalities for the spaces $\mathcal{Y}_{\mathrm{d}}^1 = H^s(\mathbb{M})$ and $\mathcal{Y}_{\mathrm{d}}^2 = \mathbf{L}^\infty(\mathbb{M})$ in Remark 3.2. The former is readily obtained by Remark 2.6 with $\theta_1 = 1 - s/a$ as long as $F : D(F) \subset \mathcal{X} \to \mathcal{Y}$ satisfies

$$
\|F(u_1) - F(u_2)\|_{H^a(\mathbb{M})} \le C_{\mathrm{L}}\ell(u_1, u_2)^{1/r} \qquad (5.6)
$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

for some $a > s$, $r \geq 1$, $C_{\mathrm{L}} > 0$, and all $u_1, u_2 \in D(F)$. On $\mathcal{Y}^2_{\mathrm{d}} = \mathbf{L}^\infty(\mathbb{M})$ we exploit Agmon's inequality (see, e.g., [1]), which states that there exists a constant $C > 0$ such that

$$\|g\|_{\mathbf{L}^\infty(\mathbb{M})} \leq C \|g\|^{(2a-d)/(2a)}_{\mathbf{L}^2(\mathbb{M})} \|g\|^{d/(2a)}_{H^a(\mathbb{M})}$$

for all $g \in H^a(\mathbb{M})$. Together with (5.6), this implies the interpolation inequality (2.11) on $\mathcal{Y}^2_{\mathrm{d}}$ with $\theta_2 = (2a-d)/(2a)$. Now we are in position to apply Remark 3.2 to obtain the following corollary.

COROLLARY 5.2. *Consider the data model (5.1) on a bounded Lipschitz domain $\mathbb{M} \subset \mathbb{R}^d$, and suppose that $u^\dagger$ satisfies the variational source condition (2.3) and is such that $g^\dagger = F(u^\dagger) \in \mathscr{C}^s(\mathbb{M})$ with some $s \in \mathbb{N}$, $s > d/2$. Furthermore let Q_n be a quadrature rule satisfying (5.4) and (5.5), and choose $\mathcal{S}$ and $\mathcal{T}$ as in (5.2). If F satisfies (5.6) and if $2s/a < r(1 + s/a)$ and $d/a < r(1 + d/(2a))$, then there exists a constant $C > 0$ such that*

$$\begin{aligned}\ell(\hat{u}_\alpha, u^\dagger) \leq C\big(&n^{-2a\kappa/(a+s(1-2/r))}\alpha^{-(a+s)/(a+s(1-2/r))} \\ &+ \alpha^{-(2a+d)/(2a+d(1-2/r))}\delta^{4a/(2a+d(1-2/r))} + \varphi_{\mathrm{app}}(2\alpha)\big).\end{aligned}$$

In the situation of Example 2.2, i.e., under (2.5) and (4.3), $\mathcal{R}(u) = \frac{1}{2}\|u\|^2_{\mathcal{X}}$, $\ell(u, u^\dagger) = \frac{1}{4}\|u - u^\dagger\|^2_{\mathcal{X}}$, and $\|u\|_{H^t} \leq \rho$ with $\max\{0, s - a + d/2\} \leq t < a$, we obtain the upper bound

$$\|\hat{u}_\alpha - u^\dagger\|^2_{\mathcal{X}} \leq C(n^{-2\kappa}\alpha^{-(a+s)/a} + \alpha^{-(2a+d)/(2a)}\delta^2 + \rho^2\alpha^{t/a}).$$

If we choose $\alpha \sim \max\{(\delta/\rho)^{4a/(2a+2t+d)}, (1/(n^\kappa\rho))^{2a/(t+a+s)}\}$, we obtain the convergence rate

$$\begin{aligned}&\|\hat{u}_\alpha - u^\dagger\|^2_{\mathcal{X}} \\ &\quad = \mathcal{O}(\max\{\rho^{(2a+d)/(t+a+d/2)}\delta^{s/(s+a+d/2)},\ \rho^{(2a+2s+t)/(a+s+t)}n^{-2t\kappa/(a+s+t)}\}) \\ &\qquad\qquad \text{as } \delta \to 0,\ n \to \infty,\end{aligned}$$

where $\delta := \max_{1\leq i\leq n}|\epsilon_i|$. Thus, as long as $n^{-\kappa/(a+s+t)} \lesssim \delta^{2/(2t+2a+d)}$, the discretization error is negligible.

REMARK 5.3. With our techniques, it is also possible to allow for random noise ϵ_i in (5.1). However, in this case one should not estimate the right-hand side of (5.3) by δ as above, but again by a smoother (discrete) norm of $g - g^\dagger$ and a weaker (discrete) norm of the noise contributions ϵ_i. This then requires one to control the supremum of a corresponding empirical process, which is possible by standard techniques such as chaining (see, e.g., [37, Corollary 2.2.5]), but beyond the scope of this paper.

6. Impulsive noise. In [19] and [22] we have shown that the techniques discussed in this paper can also be used to derive rates of convergence for inverse problems with impulsive noise, i.e., in situations where the noise can be understood as a function $\xi : \mathbb{M} \to \mathbb{R}$ which is large on a small part of $\mathbb{M}$, and small or even zero elsewhere. Here and in what follows, $\mathbb{M}$ is a manifold with finite Lebesgue measure $|\mathbb{M}|$ in $\mathbb{R}^d$. In this section, we extend the previously mentioned results by additionally allowing for a perturbed operator $F_h : D(F) \subset \mathcal{X} \to \mathcal{Y}_h := \mathbf{L}^1(\mathbb{M}_h)$ and consider with $\mathcal{Y} = \mathbf{L}^1(\mathbb{M})$ the noise model

$$g^{\mathrm{obs}} = F_h(u^\dagger) + \xi, \tag{6.1}$$

where ξ is a function on a (possibly discrete) finite measure space $(\mathbb{M}_h, \Sigma, \nu)$. If F_h not only is a projection of F, i.e., $F_h = P_h F$ for some operator $P_h : L^1(\mathbb{M}) \to L^1(\mathbb{M}_h)$, but also contains modeling or numerical approximation errors, then ξ contains not only measurement but also modeling errors.

ASSUMPTION IN. *Suppose that the noise function* $\xi \in \mathbf{L}^1(\mathbb{M}_h)$ *and there exists a measurable set* $\mathbb{M}_c \subset \mathbb{M}_h$ *such that*

$$\|\xi\|_{\mathbf{L}^1(\mathbb{M}_h \setminus \mathbb{M}_c)} \le \varepsilon, \quad \nu(\mathbb{M}_c) \le \eta$$

for two noise parameters $\varepsilon, \eta \ge 0$.

We refer to [19] for a detailed motivation of this assumption. It exactly describes the setting that ξ is small on a large part $\mathbb{M}_h \setminus \mathbb{M}_c$ of the domain $\mathbb{M}$, but might be arbitrarily large on the corrupted part $\mathbb{M}_c$. Note that we do not impose a bound on $|\xi|$ on $\mathbb{M}_c$, but only assume that $\xi \in \mathbf{L}^1(\mathbb{M}_h)$ in total.

Moreover, we assume that the perturbed operator F_h satisfies the error bounds

$$\begin{aligned} \|F(u) - F(u^\dagger)\|_{\mathbf{L}^1(\mathbb{M})} - \|F_h(u) - F_h(u^\dagger)\|_{\mathbf{L}^1(\mathbb{M}_h)} &\le C_w h^w \|F(u) - F(u^\dagger)\|_{\mathcal{Y}_d}, \\ \|F_h(u) - F_h(u^\dagger)\|_{\mathbf{L}^\infty(\mathbb{M}_h)} &\le C_w \|F(u) - F(u^\dagger)\|_{\mathbf{L}^\infty(\mathbb{M})} \end{aligned} \tag{6.2}$$

for all $u \in D(F)$ with some constants $C_w, w \ge 0$ and a norm $\|\cdot\|_{\mathcal{Y}_d}$.

As argued in [8, 19, 22], the natural choice of $\mathcal{S}$ would be

$$\mathcal{S}(g; g^{\mathrm{obs}}) = \|g - g^{\mathrm{obs}}\|_{\mathbf{L}^1(\mathbb{M}_h)},$$

as the $\mathbf{L}^1$-norm is more robust to the outliers in $\mathbb{M}_c$ compared to a standard least-squares approach involving an $\mathbf{L}^2$-norm. Unfortunately, the corresponding optimization problem (1.1) is non-smooth due to the $\mathbf{L}^1$-norm. Following [8] one can also consider the Huber-type smoothing

$$\mathcal{H}_\beta(g) := \int\limits_{\mathbb{M}_h} h_\beta(g(x))\,\mathrm{d}\nu(x), \quad h_\beta(v) := \begin{cases} v - \beta/2 & \text{if } v > \beta, \\ -v - \beta/2 & \text{if } v < -\beta, \\ (1/(2\beta))v^2 & \text{if } |v| \le \beta, \end{cases}$$

with some parameter $\beta \ge 0$. As $\nu(\mathbb{M}_h) < \infty$, one clearly has

$$\|g\|_{\mathbf{L}^1(\mathbb{M}_h)} - \frac{\beta}{2}\nu(\mathbb{M}_h) \le \mathcal{H}_\beta(g) \le \|g\|_{\mathbf{L}^1(\mathbb{M}_h)}, \quad g \in \mathbf{L}^1(\mathbb{M}_h),\ \beta \ge 0, \tag{6.3}$$

and hence in particular $\mathcal{H}_0(g) = \|g\|_{\mathbf{L}^1(\mathbb{M}_h)}$ for all $g \in L^1(\mathbb{M}_h)$. If we choose

$$\mathcal{S}(g; g^{\mathrm{obs}}) = \mathcal{H}_\beta(g - g^{\mathrm{obs}}), \quad \mathcal{T}(g; g^\dagger) = \|g - g^\dagger\|_{\mathbf{L}^1(\mathbb{M})}, \tag{6.4}$$

the corresponding problem (1.1) might be smooth depending on $\mathcal{R}$ and hence the non-differentiability due to the $\mathbf{L}^1$-norm can be overcome. Note that neither $\mathbb{M}_h \setminus \mathbb{M}_c$ nor $\mathbb{M}_c$ need to be known for computing the Tikhonov regularizer $\hat{u}_\alpha$ in (1.1) for these choices, but they are rather of theoretical interest. Again, the error introduced by the operator approximation F_h is reflected in the choice of $\mathcal{S}$ and $\mathcal{T}$, namely by the $\mathbf{L}^1(\mathbb{M})$-norm in $\mathcal{T}$ compared to the $\mathbb{M}_h$ integrals in $\mathcal{S}$.

We now argue that the additional error due to the smoothing parameter $\beta \ge 0$ can be included in the effective noise level. For $\beta = 0$, we have shown in [19, equation (3.1)] that Assumption IN implies that the noise model (6.1) is $\mathbf{L}^\infty(\mathbb{M}_h)$-admissible with $\eta_0 = 2\varepsilon$, $\eta_1 = 2\eta$, and $\mu = 1$, i.e., we have the estimate

$$\|g - F_h(u^\dagger)\|_{\mathbf{L}^1(\mathbb{M}_h)} - (\|g - g^{\mathrm{obs}}\|_{\mathbf{L}^1(\mathbb{M}_h)} - \|\xi\|_{\mathbf{L}^1(\mathbb{M}_h)}) \le 2\varepsilon + 2\eta\|g - F_h(u^\dagger)\|_{\mathbf{L}^\infty(\mathbb{M}_h)}$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

for all $g \in \mathbf{L}^\infty(\mathbb{M}_h)$. For general $\beta \geq 0$, this together with the estimate (6.3) implies

$$\begin{aligned}
\operatorname{err}(u) &= \|F(u) - F(u^\dagger)\|_{\mathbf{L}^1(\mathbb{M})} - \mathcal{H}_\beta(F_h(u) - g^{\mathrm{obs}}) + \mathcal{H}_\beta(\xi) \\
&\leq \|F(u) - F(u^\dagger)\|_{\mathbf{L}^1(\mathbb{M})} - \|F_h(u) - g^{\mathrm{obs}}\|_{\mathbf{L}^1(\mathbb{M}_h)} + \|\xi\|_{\mathbf{L}^1(\mathbb{M}_h)} + \frac{\beta}{2}\nu(\mathbb{M}_h) \\
&\leq 2\varepsilon + \frac{\beta}{2}\nu(\mathbb{M}_h) + 2\eta\|F_h(u) - F_h(u^\dagger)\|_{\mathbf{L}^\infty(\mathbb{M}_h)} + C_w h^w \|F(u) - F(u^\dagger)\|_{\mathcal{Y}_d} \\
&\leq 2\varepsilon + \frac{\beta}{2}\nu(\mathbb{M}_h) + 2C_w\eta\|F(u) - F(u^\dagger)\|_{\mathbf{L}^\infty(\mathbb{M})} + C_w h^w \|F(u) - F(u^\dagger)\|_{\mathcal{Y}_d}
\end{aligned}$$

for all $u \in D(F)$ after adding and subtracting $\|F_h(u) - F_h(u^\dagger)\|_{\mathbf{L}^1(\mathbb{M}_h)}$ and using (6.2).

This shows that, even with Huber-type smoothing, the noise model (6.1) is $(\mathbf{L}^\infty(\mathbb{M}), \mathcal{Y}_d)$-admissible (in the sense of Remark 2.5) with parameters $\eta_0 = 2\varepsilon + (\beta/2)\nu(\mathbb{M}_h)$, $\eta_1 = 2C_w\eta$, $\eta_2 = C_w h^w$, and $\mu_1 = \mu_2 = 1$.

In [19, Lemma 3.2] we have furthermore shown that Ehrling's lemma implies for bounded Lipschitz domains $\mathbb{M}$ an additive interpolation inequality of the form (2.15) with $\mathcal{Y}_{\mathrm{d}} = \mathbf{L}^\infty(\mathbb{M})$, $\mathcal{Y}_{F,\ell} = W^{k,p}(\mathbb{M})$, and $\gamma_1(\delta) = \delta^{k/d-1/p}$ whenever $k > d/p$. Consequently, it follows from Remark 2.7 that Lipschitz continuity

$$\|F(u) - F(u^\dagger)\|_{W^{k,p}(\mathbb{M})} \leq L\ell(u, u^\dagger)^{1/r} \quad \text{for all } u \in \mathcal{X} \tag{6.5}$$

of F with respect to ℓ implies that the operator F is additively $(\mathbf{L}^\infty(\mathbb{M}), \gamma_1, \ell)$-smoothing. As a corollary of our main result in Theorem 3.4, we obtain the following.

COROLLARY 6.1. *Suppose that Assumptions IN and A are satisfied and that the operator F maps $\mathcal{X}$ Lipschitz continuously into $W^{k,p}(\mathbb{M})$ in the sense of (6.5) with $k > d/p$. Furthermore let the approximation F_h satisfy (6.2) and assume that F is additively $(\mathcal{Y}_{\mathrm{d}}, \gamma_2, \ell)$-smoothing. If now $\hat{u}_\alpha$ is a minimizer of the Tikhonov functional (1.1) with $\mathcal{S}$ as in (6.4) with some $\beta \geq 0$, then there exists a constant $C > 0$ such that the error bound*

$$\ell(\hat{u}_\alpha, u^\dagger) \leq \frac{16\varepsilon + \beta\nu(\mathbb{M}_h)}{2\alpha} + C\frac{\eta^{r'\vartheta}}{\alpha^{r'}} + C\left(\frac{h^w}{\alpha}\right)^{r'} \gamma_2(4h^w)^{r'} + 4\varphi_{\mathrm{app}}(2\alpha) \tag{6.6}$$

holds true with $\vartheta = k/d - 1/p + 1$ and the conjugate exponent $r' = r/(r-1)$ of r.

In [22] we have furthermore shown that this result can be generalized to exponentially smoothing operators, i.e., situations in which the range of F consists of analytic functions. In this case, the function γ can be described as the Fenchel conjugate of a function measuring the growth of the holomorphic extension along the imaginary axis; cf. [22, equation (3.3)]. There we have furthermore discussed the examples of the backwards heat equation (where we obtained $\gamma(\delta) \sim \exp(-1/\delta^2)$) and three-dimensional satellite gradiometry (where we obtained $\gamma(\delta) \sim \delta^{-5/2} R^{-\sqrt{4\pi/\delta - 4}}$ with the radius $R > 1$ of the measurement shell).

To specify the result from Corollary 6.1 further, let us consider the following natural noise model.

EXAMPLE 6.2. We consider the following random-valued impulsive noise (RVIN) model; see, e.g., [8]. Suppose that $\mathbb{M}_h = \{x_{1,h}, \ldots, x_{n,h}\}$ is discrete with $\nu(\{x_i\}) = |\mathbb{M}|/n$ for all i and

$$\xi(x_{i,h}) = (1 - B_i)\xi_{i,h} + B_i\zeta_{i,h}, \quad i = 1, \ldots, n, \tag{6.7}$$

where B_i are independent Bernoulli-distributed random variables with parameter p_n (i.e., $P[B_i = 0] = 1 - p_n$ and $P[B_i = 1] = p_n$), and $\xi_h = (\xi_{i,h})$ and $\zeta_h = (\zeta_{i,h})$ are arbitrary deterministic or random vectors with $\|\xi\|_{L^1(\mathbb{M}_h)} \leq \varepsilon$.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

In the setting of Example 6.2, it is natural to consider $\mathbb{M}_c = \{x_{i,h} : B_i = 1\}$ such that $\eta \sim n^{-1}|\mathbb{M}| \operatorname{Bin}(n, p_n)$. This yields the following bound.

COROLLARY 6.3. *Suppose that the observations are discrete as described in Example 6.2 and let Assumption A be satisfied. Assume furthermore that F maps $\mathcal{X}$ Lipschitz continuously into $W^{k,p}(\mathbb{M})$ in the sense of (6.5) with $k > d/p$. Furthermore let the approximation F_h satisfy (6.2) and assume that F is additively $(\mathcal{Y}_{\mathrm{d}}, \gamma_2, \ell)$-smoothing. If now $\hat{u}_\alpha$ is a minimizer of the Tikhonov functional (1.1) with $\mathcal{S}$ as in (6.4) with some $\beta \geq 0$, then there exists a constant $C > 0$ such that the error bound*

$$\ell(\hat{u}_\alpha, u^\dagger) \leq \frac{16\varepsilon + \beta|\mathbb{M}|}{2\alpha} + C\left(\frac{h^w}{\alpha}\right)^{r'} \gamma_2(4h^w)^{r'} + C\frac{\eta^{r'\vartheta}}{\alpha^{r'}} + 4\varphi_{\mathrm{app}}(2\alpha) \tag{6.8}$$

holds true a.s. with the random variable $\eta \sim n^{-1}|\mathbb{M}| \operatorname{Bin}(n, p_n)$ and the exponents $\vartheta = k/d - 1/p + 1$ and $r' = r/(r-1)$.

Due to the explicitly known distribution of η, this allows us to derive convergence rates in expectation for $\ell(\hat{u}_\alpha, u^\dagger)$. As the moments of a binomially distributed random variable $X \sim \operatorname{Bin}(n, p)$ satisfy

$$\mathbb{E}[X^c] \leq (np)^c \exp\left(\frac{c^2}{2np}\right)$$

(see [2]), we obtain

$$\begin{aligned}\mathbb{E}[\ell(\hat{u}_\alpha, u^\dagger)] \leq{}& \frac{16\varepsilon + \beta|\mathbb{M}|}{2\alpha} + C\frac{(np_n)^{r'\vartheta} \exp[(r'\vartheta)^2/(2np_n)]}{\alpha^{r'}} \\ &+ C\left(\frac{h^w}{\alpha}\right)^{r'} \gamma_2(4h^w)^{r'} + 4\varphi_{\mathrm{app}}(2\alpha).\end{aligned}$$

Furthermore, as a binomially distributed random variable $X \sim \operatorname{Bin}(n, p)$ is sub-Gaussian with parameter $\sigma = n/2$, it satisfies the deviation inequality $\mathbb{P}[X \geq \delta] \leq \exp[-2(\delta - np)^2/n^2]$ for all $\delta \geq np$. As a consequence, also deviation results for $\ell(\hat{u}_\alpha, u^\dagger)$ can be derived from (6.8).

REMARK 6.4. In view of (6.6) and (6.8), it seems on first glance that the smoothing parameter $\beta \geq 0$ has only negative effects on the derived error bounds. However, if the noise functions ξ or ξ_h on $\mathbb{M}_c$ are huge, it might numerically be helpful to choose $\beta > 0$ not too small. The error bounds (6.6) and (6.8) describe a worst-case effect of this approximation.

In the model described in Example 6.2, we do not impose any assumptions on the distributions of ζ_h and ξ_h besides the requirement $\|\xi\|_{L^1(\mathbb{M}_h)} \leq \varepsilon$ for the function ξ as in (6.7). The reason is that we cannot make use of their distributions, in particular potential independence of the error components in our current analysis. For such situations, a positive β might also improve error bounds.

7. Conclusion and outlook. We have proposed a flexible unified framework which allows one to bound the effects of noise both in the data and in the forward operator in variational regularization of inverse problems.

We have shown this approach at work for several examples with deterministic and stochastic data, in both continuous and discrete settings. Our analysis provided optimal rates of convergence in the data noise level and conditions on additional discretization and other error parameters which guarantee that these additional sources of error do not become dominant. Whereas errors in the measured data are usually difficult or impossible to improve, discretization parameters are at our disposal, and our analysis provides guidelines on how they should be chosen asymptotically.

ETNA
Kent State University and
Johann Radon Institute (RICAM)

Further investigations involving the theory of stochastic processes are required to analyze discrete sampled stochastic data. A further interesting topic of future research are a-posteriori stopping rules for general noise models.

Appendix A. Besov spaces and Gaussian white noise. In [38] it has been shown that white noise on the torus $\mathbb{T}^d$ almost surely belongs to the Besov spaces $B^{-d/2}_{p,\infty}(\mathbb{T}^d)$ for $p \in [1,\infty)$, but not to any smaller Besov spaces. In this appendix we will use this result to derive Besov regularity of Gaussian white noise on smooth, bounded domains $\mathbb{M} \subset \mathbb{R}^d$.

Recall that there are different possibilities to define Besov spaces on domains, in particular

$$B^s_{p,q}(\mathbb{M}) := \{f|_{\mathbb{M}} : f \in B^s_{p,q}(\mathbb{R}^d)\}, \qquad \|g\|_{B^s_{p,q}(\mathbb{M})} := \inf_{f|\Omega = g} \|f\|_{B^s_{p,q}(\mathbb{R}^d)},$$
$$\tilde{B}^s_{p,q}(\mathbb{M}) := \{f \in B^s_{p,q}(\mathbb{R}^d) : \operatorname{supp} f \subset \overline{\mathbb{M}}\}, \qquad \|f\|_{\tilde{B}^s_{p,q}(\mathbb{M})} = \|f\|_{B^s_{p,q}(\mathbb{M})}$$

(see [36, Sections 4.1 and 4.3]). For negative s, the spaces $\tilde{B}^s_{p,q}(\mathbb{M})$ may contain distributions f with $\operatorname{supp} f \subset \partial\mathbb{M}$, and such elements f are not contained in $B^s_{p,q}(\mathbb{M})$. On the other hand, for positive s, the space $\tilde{B}^s_{p,q}(\mathbb{M})$ is a subspace of $B^s_{p,q}(\mathbb{M})$. More precisely, if ν is the outer unit normal vector on $\partial\mathbb{M}$, $s > 1/p - 1$, $s - 1/p \notin \mathbb{Z}$, $p \in (1,\infty)$, and $q \in [1,\infty]$, then there exists a bounded linear trace operator

$$R : B^s_{p,q}(\mathbb{M}) \to \prod_{k=0}^{K} B^{s-1/p-k}_{p,q}(\partial\mathbb{M}), \quad f \mapsto \left(\left.\frac{\partial^k f}{\partial \nu^k}\right|_{\partial\mathbb{M}} \right)_{k=0,\dots,K}, \tag{A.1}$$

with $K := \max\{k \in \mathbb{Z} : s - 1/p - k > 0\}$, and

$$\tilde{B}^s_{p,q}(\mathbb{M}) = \{f \in B^s_{p,q}(\mathbb{M}) : Rf = 0\} \tag{A.2}$$

(see [36, Sections 4.3 and 4.7]). In particular, $\tilde{B}^s_{p,q}(\mathbb{M}) = B^s_{p,q}(\mathbb{M})$ for $s \in (1/p - 1, 1/p)$ as $K < 0$ and hence $\operatorname{range}(R) = \{0\}$. We also have

$$\tilde{B}^s_{p,q}(\mathbb{M})' = B^{-s}_{p',q'}(\mathbb{M}) \tag{A.3}$$

for $s \in \mathbb{R}$, $p \in (1,\infty)$, and $q \in [1,\infty)$ with $1/p + 1/p' = 1/q + 1/q' = 1$; see [36, Section 4.8.1].

THEOREM A.1. *Let $\mathbb{M} \subset \mathbb{R}^d$ be a smooth bounded domain and let Z be Gaussian white noise on $L^2(\mathbb{M})$. Then $Z \in B^{-d/2}_{p,\infty}(\mathbb{M})$ for all $p \in (1,\infty)$ with $d/2 + 1/p \notin \mathbb{N}$ almost surely.*

Proof. Choose R sufficiently large such that $\overline{\mathbb{M}} \subset (-R/2, R/2)^d$ and consider $\mathbb{M}$ as a subset of the torus $\mathbb{T}^d_R := (\mathbb{R} \setminus (R\mathbb{Z}))^d$. Let $p' \in (1,\infty)$ be such that $1/p + 1/p' = 1$. As $d/2 - 1/p' \notin \mathbb{Z}$, it follows from [36, Sections 2.9.3 and 4.7] that the trace operator R in (A.1) is a retraction from $B^{d/2}_{p',1}(\mathbb{T}^d_R)$ to $Z := \prod_{k=0}^{K} B^{d/2-1/p'-k}_{p',1}(\partial\mathbb{M})$; in particular, it has a bounded right inverse $R^\dagger : Z \to B^{d/2}_{p',1}(\mathbb{T}^d_R)$ and $RR^\dagger = I$. Using (A.2) this yields a topological decomposition

$$B^{d/2}_{p',1}(\mathbb{T}^d_R) = \tilde{B}^{d/2}_{p',1}(\mathbb{M}) \oplus \operatorname{range}(R^\dagger) \oplus \tilde{B}^{d/2}_{p',1}(\mathbb{T}^d_R \setminus \overline{\mathbb{M}})$$

with bounded projections $P_{\mathbb{M}} f := 1_{\mathbb{M}} \cdot (f - R^\dagger R f)$, $P_{\partial\mathbb{M}} f := R^\dagger R f$, and $P_{\mathbb{M}^c} := 1_{\mathbb{T}^d_R \setminus \overline{\mathbb{M}}} \cdot (f - R^\dagger R f)$; in fact, $P_{\mathbb{M}} + P_{\partial\mathbb{M}} + P_{\mathbb{M}^c} = I$, idempotency follows from $RR^\dagger = I$, and boundedness from (A.2). We also obtain a corresponding topological decomposition of the dual space (see (A.3)):

$$B^{-d/2}_{p,\infty}(\mathbb{T}^d_R) = B^{-d/2}_{p,\infty}(\mathbb{M}) \oplus \operatorname{range}(R^\dagger)' \oplus B^{-d/2}_{p,\infty}(\mathbb{T}^d_R \setminus \overline{\mathbb{M}})$$

ETNA
Kent State University and
Johann Radon Institute (RICAM)

with bounded projections $P'_{\mathbb{M}}$, $P'_{\partial\mathbb{M}}$, and $P'_{\mathbb{M}^c}$. To see this, note that, for $F \in B^{-d/2}_{p,\infty}(\mathbb{T}^d_R)$, we have $\langle P'_{\mathbb{M}}F, f\rangle = \langle F, P_{\mathbb{M}}f\rangle = 0$ for $f \in \tilde{B}^{d/2}_{p',1}(\mathbb{T}^d_R \setminus \overline{\mathbb{M}})$ and $f \in \operatorname{range}(R^\dagger)$, and $\langle P'_{\mathbb{M}}F, f\rangle = \langle F, f\rangle$ for $f \in \tilde{B}^{d/2}_{p',1}(\mathbb{M})$, so $P'_{\mathbb{M}}F = F|_{\mathbb{M}} \in B^{-d/2}_{p,\infty}(\mathbb{M})$. Similarly, $P'_{\mathbb{M}^c}F \in B^{-d/2}_{p,\infty}(\mathbb{T}^d_R \setminus \overline{\mathbb{M}})$. Moreover, $\operatorname{supp} P'_{\partial\mathbb{M}}F \subset \partial\mathbb{M}$ as $Rf = 0$ for all $f \in C^\infty_0(\mathbb{T}^d_R \setminus \partial\mathbb{M})$. (Note, however, that $P'_{\partial\mathbb{M}}F \neq 0$ in general even for $F \in C^\infty(\mathbb{T}^d_R)$.)

Let Z be Gaussian white noise on $L^2(\mathbb{T}^d_R)$. As mentioned above, $Z \in B^{-d/2}_{p,\infty}(\mathbb{T}^d_R)$ a.s.; see [38]. Then the restriction $Z_{\mathbb{M}}$ of Z to the closed subspace $L^2(\mathbb{M}) \subset L^2(\mathbb{T}^d_R)$ (where functions in $L^2(\mathbb{M})$ are extended by 0 to $\mathbb{T}^d_R \setminus \mathbb{M}$) is Gaussian white noise on $L^2(\mathbb{M})$ since the covariance operator of $Z_{\mathbb{M}}$ is the identity. As $\langle Z_{\mathbb{M}}, f\rangle = \langle P'_{\mathbb{M}}Z, f\rangle$ for all f in the dense subset $\tilde{B}^{d/2}_{p',1}(\mathbb{M}) \subset L^2(\mathbb{M})$, we have $Z_{\mathbb{M}} = P'_{\mathbb{M}}Z$, and hence $Z_{\mathbb{M}} \in B^{-d/2}_{p,\infty}(\mathbb{M})$ a.s. $\square$

REMARK A.2. Note that $\|Z_{\mathbb{M}}\|_{B^{-d/2}_{p,\infty}(\mathbb{M})} \leq C\|Z\|_{B^{-d/2}_{p,\infty}(\mathbb{T}^d_R)}$ with $C := \|P'_{\mathbb{M}}\|$.

REFERENCES

[1] S. AGMON, *Lectures on Elliptic Boundary Value Problems*, AMS Chelsea, Providence, 2010.

[2] T. D. AHLE, *Sharp and simple bounds for the raw moments of the binomial and Poisson distributions*, Statist. Probab. Lett., 182 (2022), Art. 109306, 5 pages.

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

[4] T. BONESKY, K. S. KAZIMIERSKI, P. MAASS, F. SCHÖPFER, AND T. SCHUSTER, *Minimization of Tikhonov functionals in Banach spaces*, Abstr. Appl. Anal., (2008), Art. 192679, 19 pages.

[5] S. C. BRENNER AND L. R. SCOTT, *The Mathematical Theory of Finite Element Methods*, Springer, New York, 1994.

[6] L. BUNGERT, M. BURGER, Y. KOROLEV, AND C.-B. SCHÖNLIEB, *Variational regularisation for inverse problems with imperfect forward operators and general noise models*, Inverse Problems, 36 (2020), Art. 125014, 32 pages.

[7] M. BURGER, Y. KOROLEV, AND J. RASCH, *Convergence rates and structure of solutions of inverse problems with imperfect forward models*, Inverse Problems, 35 (2019), Art. 024006, 33 pages.

[8] C. CLASON AND B. JIN, *A semismooth Newton method for nonlinear parameter identification problems with impulsive noise*, SIAM J. Imaging Sci., 5 (2012), pp. 505–538.

[9] R. A. DEVORE AND V. A. POPOV, *Interpolation of Besov spaces*, Trans. Amer. Math. Soc., 305 (1988), pp. 397–414.

[10] J. FLEMMING, *Generalized Tikhonov Regularization – Basic Theory and Comprehensive Results on Convergence Rates*, PhD. Thesis, Faculty of Math., Chemnitz University of Technology, Chemnitz, 2011.

[11] ———, *Generalized Tikhonov Regularization and Modern Convergence Rate Theory in Banach Spaces*, Shaker, Aachen, 2012.

[12] E. GINÉ AND R. NICKL, *Mathematical Foundations of Infinite-Dimensional Statistical Models*, Cambridge University Press, New York, 2016.

[13] M. GRASMAIR, *Generalized Bregman distances and convergence rates for non-convex regularization methods*, Inverse Problems, 26 (2010), Art. 115014, 16 pages.

[14] M. HOFFMANN AND M. REISS, *Nonlinear estimation for linear inverse problems with error in the operator*, Ann. Statist., 36 (2008), pp. 310–336.

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

[16] T. HOHAGE, *Regularization of exponentially ill-posed problems*, Numer. Funct. Anal. Optim., 21 (2000), pp. 439–464.

[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] T. HOHAGE AND F. WEIDLING, *Characterizations of variational source conditions, converse results, and maxisets of spectral regularization methods*, SIAM J. Numer. Anal., 55 (2017), pp. 598–620.

[19] T. HOHAGE AND F. WERNER, *Convergence rates for inverse problems with impulsive noise*, SIAM J. Numer. Anal., 52 (2014), pp. 1203–1221.

[20] B. KALTENBACHER, A. NEUBAUER, AND O. SCHERZER, *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*, de Gruyter, Berlin, 2008.

[21] B. KALTENBACHER, F. SCHÖPFER, AND T. SCHUSTER, *Iterative methods for nonlinear ill-posed problems in Banach spaces: convergence and applications to parameter identification problems*, Inverse Problems, 25 (2009), Art. 065003, 19 pages.
[22] C. KÖNIG, F. WERNER, AND T. HOHAGE, *Convergence rates for exponentially ill-posed inverse problems with impulsive noise*, SIAM J. Numer. Anal., 54 (2016), pp. 341–360.
[23] S. LU AND J. FLEMMING, *Convergence rate analysis of Tikhonov regularization for nonlinear ill-posed problems with noisy operators*, Inverse Problems, 28 (2012), Art. 104003, 19 pages.
[24] P. MATHÉ AND S. V. PEREVERZEV, *Moduli of continuity for operator valued functions*, Numer. Funct. Anal. Optim., 23 (2002), pp. 623–631.
[25] ———, *Discretization strategy for linear ill-posed problems in variable Hilbert scales*, Inverse Problems, 19 (2003), pp. 1263–1277.
[26] A. NEUBAUER, *Tikhonov regularization for nonlinear ill-posed problems: optimal convergence and finite-dimensional approximation*, Inverse Problems, 5 (1989), pp. 541–557.
[27] A. NEUBAUER AND O. SCHERZER, *Finite-dimensional approximation of Tikhonov regularized solutions of nonlinear ill-posed problems*, Numer. Funct. Anal. Optim., 11 (1990), pp. 85–99.
[28] P. OSWALD, *Multilevel Finite Element Approximation*, Teubner, Stuttgart, 1994.
[29] C. PÖSCHL, E. RESMERITA, AND O. SCHERZER, *Discretization of variational regularization in Banach spaces*, Inverse Problems, 26 (2010), Art. 105017, 18 pages.
[30] S. PRÖSSDORF AND B. SILBERMANN, *Numerical Analysis for Integral and Related Operator Equations*, Birkhäuser, Basel, 1991.
[31] O. SCHERZER, M. GRASMAIR, H. GROSSAUER, M. HALTMEIER, AND F. LENZEN, *Variational Methods in Imaging*, Springer, New York, 2009.
[32] T. SCHUSTER, B. KALTENBACHER, B. HOFMANN, AND K. S. KAZIMIERSKI, *Regularization Methods in Banach Spaces*, de Gruyter, Berlin, 2012.
[33] B. SIXOU AND C. MORY, *Kullback–Leibler residual and regularization for inverse problems with noisy data and noisy operator*, Inverse Probl. Imaging, 13 (2019), pp. 1113–1137.
[34] B. SPRUNG, *Upper and lower bounds for the Bregman divergence*, J. Inequal. Appl., (2019), Art. 4, 12 pages.
[35] M. E. TAYLOR, *Partial Differential Equations II. Qualitative Studies of Linear Equations*, 2nd ed., Springer, New York, 2011.
[36] H. TRIEBEL, *Interpolation Theory, Function Spaces, Differential Operators*, North-Holland, Amsterdam, 1978.
[37] A. W. VAN DER VAART AND J. A. WELLNER, *Weak Convergence and Empirical Processes*, Springer, New York, 1996.
[38] 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.
[39] F. WEIDLING, *Variational Source Conditions and Conditional Stability Estimates for Inverse Problems in PDEs*, PhD. Thesis, University of Göttingen, Göttingen, 2019.
[40] F. WEIDLING AND T. HOHAGE, *Variational source conditions and stability estimates for inverse electromagnetic medium scattering problems*, Inverse Probl. Imaging, 11 (2017), pp. 203–220.
[41] 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.
[42] J. A. WELLNER, *Gaussian white noise models: some results for monotone functions*, in Crossing Boundaries: Statistical Essays in Honor of Jack Hall, J. E. Kolassa and D. Oakes, eds., vol. 43 of IMS Lecture Notes Monogr. Ser., Inst. Math. Statist., Beachwood, 2003, pp. 87–104.
[43] F. WERNER, *Inverse Problems with Poisson Data: Tikhonov-Type Regularization and Iteratively Regularized Newton Methods*, PhD. Thesis, University of Göttingen, Göttingen, 2012.
[44] F. WERNER AND B. HOFMANN, *Convergence analysis of (statistical) inverse problems under conditional stability estimates*, Inverse Problems, 36 (2020), Art. 015004, 23 pages.
[45] F. WERNER AND T. HOHAGE, *Convergence rates in expectation for Tikhonov-type regularization of inverse problems with Poisson data*, Inverse Problems, 28 (2012), Art. 104004, 15 pages.
[46] Z. B. XU AND G. F. ROACH, *Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces*, J. Math. Anal. Appl., 157 (1991), pp. 189–210.