NONLINEAR TIKHONOV REGULARIZATION IN HILBERT SCALES FOR INVERSE BOUNDARY VALUE PROBLEMS WITH RANDOM NOISE

THORSTEN HOHAGE AND MIHAELA PRICOP

Institut für Numerische und Angewandte Mathematik
Lotzestr. 16-18
D-37083 Göttingen, Germany

ABSTRACT. We consider the inverse problem to identify coefficient functions in boundary value problems from noisy measurements of the solutions. Our estimators are defined as minimizers of a Tikhonov functional, which is the sum of a nonlinear data misfit term and a quadratic penalty term involving a Hilbert scale norm. In this abstract framework we derive estimates of the expected squared error under certain assumptions on the forward operator. These assumptions are shown to be satisfied for two classes of inverse elliptic boundary value problems. The theoretical results are confirmed by Monte Carlo simulations.

1. INTRODUCTION

We consider the problem to estimate a quantity a in a separable Hilbert space $\mathcal{X}$ which is not directly observable. We only have access to a vector u in another Hilbert space $\mathcal{Y}$, and the relation between a and u is described by some nonlinear operator $F : D(F) \subset \mathcal{X} \to \mathcal{Y}$:

$$F(a) = u. \tag{1}$$

F is assumed to be one-to-one, but the inverse of F is not assumed to be continuous, i.e. we study the situation that the nonlinear operator equation (1) is ill-posed. The exact solution will be denoted by $a_\dagger \in D(F)$, and the exact data by $u_\dagger := F(a_\dagger)$.

In this paper we are particularly interested in parameter identification problems where a is an unknown parameter in an elliptic differential equation, and u is the solution to this differential equation subject to some boundary conditions. Here F is the so-called parameter-to-solution operator which is usually nonlinear even if the differential equation is linear.

We assume that $u_\dagger$ is not given exactly, but only a Hilbert-space-valued random variable $\hat{u}$ satisfying

$$\sqrt{\mathbf{E}\|\widehat{u} - u_\dagger\|_{\mathcal{Y}}^2} \leq \delta. \tag{2}$$

1991 *Mathematics Subject Classification.* Primary: 58F15, 58F17; Secondary: 53C35.

Key words and phrases. statistical inverse problem, nonlinear Tikhonov regularization, Hilbert scales, parameter identification problems, inverse boundary value problems.

We would like to thank Axel Munk for many helpful discussions. Financial support of the DFG Graduiertenkolleg 1023 "Identification in mathematical models: synergy of stochastic and numerical methods" is gratefully acknowledged.

$\hat{u} - u_\dagger$ may contain both deterministic errors and random noise, i.e. it is *not* assumed that $\mathbf{E}\widehat{u} = u_\dagger$. We will study the speed of convergence of estimators of $a_\dagger$ constructed from $\widehat{u}$ as $\delta \to 0$.

Usually $\widehat{u}$ is not given immediately in applications, but has to be computed from the data in a first step. If u describes the solution to a differential equation in a domain $\Omega \subset \mathbf{R}^d$, a natural choice for the space $\mathcal{Y}$ is $L^2(\Omega)$. A typical noise model may look as follows: The given data are described by n random variables

$$Y_i = u_\dagger(X_i) + \varepsilon_i, \qquad i = 1, .., n \tag{3}$$

where measurement errors are modeled by independent, identically distributed random variables ε_i satisfying $\mathbf{E}\varepsilon_i = 0$ and $\mathbf{Var}(\varepsilon_i) < \infty$. The measurement points $X_i \in \Omega$ may either be uniformly distributed random variables or deterministic points satisfying some additional conditions. The regression problem to construct an estimator $\widehat{u}$ of $u_\dagger$ from such data and the speed of convergence $\delta \to 0$ as $n \to \infty$ have been studied extensively in the statistical literature (see [30] and references therein). For theoretical purposes it is often convenient to work with a white noise model

$$\mathbf{Y} = u_\dagger + \sigma\xi \tag{4}$$

where ξ is a white noise process on $\mathcal{Y}$ and $\sigma = 1/\sqrt{n}$. For univariate functions it can be shown under mild assumptions that for every nonparametric regression problem there exists a white noise problem which is asymptotically equivalent to it and conversely (see [3]).

The focus of this paper is on the second step to estimate $\widehat{a}$ from $\widehat{u}$. A commonly used procedure known as nonlinear Tikhonov regularization or Method of Regularization consists in adding a stabilizing term weighted by a regularization parameter $\alpha > 0$ in a least squares formulation of the problem:

$$\widehat{a} := \operatorname{argmin}_{a \in D(F)} \left(\|F(a) - \widehat{u}\|_{\mathcal{Y}}^2 + \alpha \|a - a_0\|_{\mathcal{X}}^2 \right) \tag{5}$$

Here $a_0 \in \mathcal{X}$ denotes some initial guess of $a_\dagger$, which offers the possibility to incorporate a-priori information. The estimators (5) arise naturally in a Bayesian description of the problem (see [12]). A global minimum of the functional in (5) exists if F is weakly sequentially closed, but it is not necessarily unique (see [7]). However the method is consistent in the sense that for any choice of minimizing elements $\widehat{a} \in D(F)$ we have $\mathbf{E}\|\widehat{a} - a_\dagger\|_{\mathcal{X}}^2 \to 0$ if $\alpha \to 0$ and $\delta^2/\alpha \to 0$ ([2]). Rates of convergence have been studied for deterministic noise in [8, 19, 26, 28], and for random noise in [2, 14, 22]. In particular, it was shown in [2] that the estimators (5) converge with the rate

$$\sqrt{\mathbf{E}\|\widehat{a} - a_\dagger\|_{\mathcal{X}}^2} = \mathcal{O}\left(\sqrt{\delta}\right) \tag{6}$$

for the choice $\alpha \sim \delta$ if F has a Fréchet derivative F' which is Lipschitz continuous with constant L and if there exists $w \in \mathcal{Y}$ such that

$$a_\dagger - a_0 = F'[a_\dagger]^* w \qquad \text{and} \qquad L\|w\|_{\mathcal{Y}} < 1. \tag{7}$$

In this paper we will improve this result in two respects:

- We consider more general smoothness classes than the range of the adjoint operator $F'(a_\dagger)^*$.
- Since $F'[a_\dagger]^*$ is typically smoothing, the smallness assumption on $\|w\|$ in (7) typically corresponds to a smallness assumption on $a_\dagger - a_0$ in some higher

norm. We will replace this restrictive smallness assumption by a smallness assumption on $\|a_\dagger - a_0\|_{\mathcal{X}}$.

To achieve these goals we will use Hilbert scales as introduced for linear deterministic inverse problems by Natterer [17]. By regularization in Hilbert scales the finite qualification of Tikhonov regularization can be overcome, i.e. the rate of convergence is not bounded by $\|\widehat{a}-a_\dagger\| = O(\delta^{2/3})$, but given sufficient smoothness of the solution, the exponent of δ may be arbitrarily close to 1. Regularization methods for *linear* inverse problems with random noise have been studied in a Hilbert scale setting in [**?**, 10, 15, 16, 21]. Our analysis in section 2, and in particular the second part of the proof of our general convergence theorem is based on the work of Neubauer [19] on nonlinear Tikhonov regularization with deterministic noise. Other regularization methods for nonlinear inverse problems in Hilbert scales with deterministic noise have been studied in [6, 11, 20, 27].

The plan of this paper is as follows: In the following section we introduce Hilbert scales and prove a general result on the rates of convergence of nonlinear Tikhonov regularization in Hilbert scales for random noise. The optimality of the overall two-step method to estimate $a_\dagger$ from data described by (3) or (4) is discussed in section 3. In sections 4 and 5 we study the reconstruction of a reaction coefficient and a diffusion coefficient in elliptic differential equation and show that all assumptions required by our convergence result are satisfied. The paper is completed by some numerical experiments discussed in section 6 which verify and illustrate our theoretical results.

2. General convergence result

First, we briefly recall the definition of Hilbert scales. Let $L : D(L) \to \mathcal{X}$ be an unbounded, self-adjoint, strictly positive operator with a dense domain of definition $D(L) \subset \mathcal{X}$. Then $L^s : D(L^s) \to \mathcal{X}$ is well-defined by spectral theory for $s \in \mathbf{R}$, and the spaces $\mathcal{X}_s := D(L^s)$, $s \geq 0$ equipped with the inner product

$$\langle x, y\rangle_s := \langle L^s x, L^s y\rangle_{\mathcal{X}}, \qquad x, y \in \mathcal{X}_s$$

are Hilbert spaces. For $s < 0$ we define $\mathcal{X}_s$ as completion of $\mathcal{X}$ under the norm $\|x\|_s := \langle x, x\rangle_s^{1/2}$. $(\mathcal{X}_s)_{s\in\mathbf{R}}$ is called the Hilbert scale induced by L. An important tool for the following analysis will be the interpolation inequality

$$\|x\|_r \leq \|x\|_t^{\frac{s-r}{s-t}} \|x\|_s^{\frac{r-t}{s-t}}, \qquad x \in \mathcal{X}_s \tag{8}$$

which holds for any $t < r < s$. For a more detailed introduction to Hilbert scales and their use in regularization theory we refer to the monograph [7].

Following [19] we define the estimator $\widehat{a}$ of $a_\dagger$ by

$$\widehat{a} := \operatorname{argmin}_{a\in D(F)\cap(a_0+\mathcal{X}_s)} \left(\|F(a) - \widehat{u}\|_{\mathcal{Y}}^2 + \alpha\|a - a_0\|_s^2\right) \tag{9}$$

where $s \geq 0$ and $a_0 \in \mathcal{X}$ an initial guess. To bound the error $\mathbf{E}(\|\widehat{a} - a_\dagger\|^2)$ we need the following assumptions:

Assumptions. (1) *(assumptions on F) If $F(a) = F(a_\dagger)$ for some $a \in D(F) \cap (a_0 + \mathcal{X}_s)$, then $a = a_\dagger$. Moreover, $D(F)$ is convex, $F : D(F) \cap (a_0 + \mathcal{X}_s) \to \mathcal{Y}$ is weakly sequentially closed, and $F : D(F) \subset \mathcal{X} \to \mathcal{Y}$ has a Fréchet derivative $F' : D(F) \to L(\mathcal{X}, \mathcal{Y})$.*

(2) *(smoothing properties of F') There exist constants $p \in [0,s]$ and $\lambda, \Lambda > 0$ such that for all $h \in \mathcal{X}$*

$$\lambda \|h\|_{-p} \le \|F'(a_\dagger)h\|_{\mathcal{Y}} \le \Lambda \|h\|_{-p}. \tag{10}$$

(3) *(Lipschitz continuity of F') There exists a constant $C_L > 0$ such that*

$$\|F'(a) - F'(a_\dagger)\|_{\mathcal{Y}\leftarrow\mathcal{X}_{-p}} \le C_L \|a - a_\dagger\|_{\mathcal{X}} \le \frac{\lambda^2}{2\Lambda} \tag{11}$$

for all $a \in D(F) \cap (a_0 + \mathcal{X}_s)$.

If $\widehat{u}$ is deterministic then the norm $\|\cdot\|_{Y\leftarrow X_{-p}}$ in (11) can be replaced by $\|\cdot\|_{Y\leftarrow X}$. A condition similar to (11) has been used in [20] (in fact, condition (iv) in [20, Assumption 2.1] with $\beta = 1$ and $b = p$ is equivalent to the first inequality in (11)).

We comment on the condition $s \ge p$ in Assumption 2. Due to Lemma 4.3 below and condition (12), it follows that the first equation in the source condition (7) is satisfied. If this source condition is not fulfilled, convergence rate results are only known under much more restrictive assumptions on the operator F (cf. [28]).

Theorem 2.1. *Let $a_\dagger \in D(F)$ be the solution of operator equation $F(a_\dagger) = u_\dagger$, and let $\widehat{u}$ be an estimator of u satisfying (2). If the Assumptions 1–3 hold true and if*

$$a_\dagger - a_0 \in \mathcal{X}_q \qquad \text{for some } q \in [s, 2s+p] \tag{12}$$

then, for the a-priori choice of the regularization parameter $\alpha \sim \delta^{\frac{2(p+s)}{p+q}}$, we obtain

$$\sqrt{\mathbf{E}\|\widehat{a} - a_\dagger\|_{\mathcal{X}}^2} = \mathcal{O}\left(\delta^{\frac{q}{p+q}}\right), \qquad \delta \to 0. \tag{13}$$

Proof. By the definition (9) of $\widehat{a}$ as a solution to the minimization problem we have

$$\|F(\widehat{a}) - \widehat{u}\|_{\mathcal{Y}}^2 + \alpha\|\widehat{a} - a_0\|_s^2 \le \|u_\dagger - \widehat{u}\|_{\mathcal{Y}}^2 + \alpha\|a_\dagger - a_0\|_s^2,$$

and hence

$$\begin{aligned}\|F(\widehat{a}) - \widehat{u}\|_{\mathcal{Y}}^2 + \alpha\|\widehat{a} - a_\dagger\|_s^2 &\le \|u_\dagger - \widehat{u}\|_{\mathcal{Y}}^2 + \alpha\left(\|a_\dagger - a_0\|_s^2 - \|\widehat{a} - a_0\|_s^2 + \|\widehat{a} - a_\dagger\|_s^2\right)\\ &= \|u_\dagger - \widehat{u}\|_{\mathcal{Y}}^2 + 2\alpha\left\langle a_\dagger - a_0, a_\dagger - \widehat{a}\right\rangle_s.\end{aligned}$$

Using the hypothesis $a_\dagger - a_0 \in \mathcal{X}_q$ we obtain

$$\|F(\widehat{a}) - \widehat{u}\|_{\mathcal{Y}}^2 + \alpha\|\widehat{a} - a_\dagger\|_s^2 \le \|u_\dagger - \widehat{u}\|_{\mathcal{Y}}^2 + 2\alpha\|a_\dagger - \widehat{a}\|_{2s-q}\|a_\dagger - a_0\|_q. \tag{14}$$

Introducing the Taylor remainder $e_{\mathrm{tay}} := F(\widehat{a}) - F(a_\dagger) - F'(a_\dagger)(\widehat{a} - a_\dagger)$ we obtain the inequality

$$\begin{aligned}&\|F(\widehat{a}) - \widehat{u}\|_{\mathcal{Y}}^2 = \|u_\dagger - \widehat{u} + e_{\mathrm{tay}} + F'(a_\dagger)(\widehat{a} - a_\dagger)\|_{\mathcal{Y}}^2\\ &\quad= \quad \|u_\dagger - \widehat{u} + e_{\mathrm{tay}}\|_{\mathcal{Y}}^2 + \|F'(a_\dagger)(\widehat{a} - a_\dagger)\|_{\mathcal{Y}}^2 + 2\left\langle F'(a_\dagger)(\widehat{a} - a_\dagger), u_\dagger - \widehat{u} + e_{\mathrm{tay}}\right\rangle\\ &\quad\ge \quad \lambda^2\|\widehat{a} - a_\dagger\|_{-p}^2 - 2\Lambda\|\widehat{a} - a_\dagger\|_{-p}\left(\|u_\dagger - \widehat{u}\|_{\mathcal{Y}} + \|e_{\mathrm{tay}}\|_{\mathcal{Y}}\right).\end{aligned}$$

Plugging this into (14) yields

$$\begin{aligned}\lambda^2\|\widehat{a} - a_\dagger\|_{-p}^2 + \alpha\|\widehat{a} - a_\dagger\|_s^2 \le &\|u_\dagger - \widehat{u}\|_{\mathcal{Y}}^2 + 2\alpha\|a_\dagger - \widehat{a}\|_{2s-q}\|a_\dagger - a_0\|_q\\ &+ 2\Lambda\|\widehat{a} - a_\dagger\|_{-p}\|e_{\mathrm{tay}}\|_{\mathcal{Y}} + 2\Lambda\|\widehat{a} - a_\dagger\|_{-p}\|u_\dagger - \widehat{u}\|_{\mathcal{Y}}.\end{aligned}$$

By the interpolation inequality (8) for $t = -p$, $r = 2s - q$ and $s = s$ we have

$$\|\widehat{a} - a_\dagger\|_{2s-q} \le \|\widehat{a} - a_\dagger\|_{-p}^{\frac{-s+q}{s+p}}\,\|\widehat{a} - a_\dagger\|_s^{\frac{2s-q+p}{s+p}}. \tag{15}$$

Replacing this term in our estimate and taking the mean in the inequality we get

$$\begin{aligned}\lambda^2\mathbf{E}\|\widehat{a}-a_\dagger\|_{-p}^2+\alpha\mathbf{E}\|\widehat{a}-a_\dagger\|_s^2 &\le \mathbf{E}\|u_\dagger-\widehat{u}\|_{\mathcal{Y}}^2\\ &+2\alpha\|a_\dagger-a_0\|_q\mathbf{E}\left(\|\widehat{a}-a_\dagger\|_{-p}^{\frac{-s+q}{s+p}}\|\widehat{a}-a_\dagger\|_s^{\frac{2s-q+p}{s+p}}\right)\\ &+2\Lambda\mathbf{E}\left(\|\widehat{a}-a_\dagger\|_{-p}\|e_{\mathrm{tay}}\|_{\mathcal{Y}}\right)+2\Lambda\mathbf{E}\left(\|\widehat{a}-a_\dagger\|_{-p}\|u_\dagger-\widehat{u}\|_{\mathcal{Y}}\right)\end{aligned} \tag{16}$$

Using the Cauchy-Schwarz inequality and Jensen's inequality we obtain

$$\begin{aligned}\mathbf{E}\left(\|\widehat{a}-a_\dagger\|_{-p}^{\frac{-s+q}{s+p}}\|\widehat{a}-a_\dagger\|_s^{\frac{2s-q+p}{s+p}}\right) &\le \mathbf{E}^{\frac12}\left(\|\widehat{a}-a_\dagger\|_{-p}^{2\frac{-s+q}{s+p}}\right)\mathbf{E}^{\frac12}\left(\|\widehat{a}-a_\dagger\|_s^{2\frac{2s-q+p}{s+p}}\right)\\ &\le \mathbf{E}^{\frac12\frac{-s+q}{s+p}}\left(\|\widehat{a}-a_\dagger\|_{-p}^2\right)\mathbf{E}^{\frac12\frac{2s-q+p}{s+p}}\left(\|\widehat{a}-a_\dagger\|_s^2\right)\end{aligned}$$

since $0\le\frac{-s+q}{s+p}\le 1$ and $0\le\frac{2s-q+p}{s+p}\le 1$ for $s\le q\le 2s+p$.

Next we are going to show that the term involving the Taylor remainder e_{tay} in (16) is dominated by the term on the left hand side for small δ:

$$2\Lambda\mathbf{E}\left(\|\widehat{a}-a_\dagger\|_{-p}\|e_{\mathrm{tay}}\|_{\mathcal{Y}}\right)\le\frac{\lambda^2}{2}\mathbf{E}\|\widehat{a}-a_\dagger\|_{-p}^2 \tag{17}$$

Using Assumption 3, the Taylor remainder e_{tay} can be estimated by

$$\begin{aligned}\|e_{\mathrm{tay}}\|_{\mathcal{Y}} &= \|F(\widehat{a})-F(a_\dagger)-F'(a_\dagger)(\widehat{a}-a_\dagger)\|_{\mathcal{Y}}\\ &= \left\|\int_0^1\left\{F'(a_\dagger+t(\widehat{a}-a_\dagger))-F'(a_\dagger)\right\}(\widehat{a}-a_\dagger)\,dt\right\|_{\mathcal{Y}}\\ &\le \int_0^1\|F'(a_\dagger+t(\widehat{a}-a_\dagger))-F'(a_\dagger)\|_{\mathcal{Y}\leftarrow\mathcal{X}_{-p}}\|\widehat{a}-a_\dagger\|_{-p}\,dt\\ &\le \int_0^1 C_L t\|\widehat{a}-a_\dagger\|_0\|\widehat{a}-a_\dagger\|_{-p}\,dt\le\frac{C_L}{2}\|\widehat{a}-a_\dagger\|_0\|\widehat{a}-a_\dagger\|_{-p},\end{aligned}$$

which implies (17). Altogether, we arrive at the estimate

$$\begin{aligned}\mathbf{E}\|\widehat{a}-a_\dagger\|_{-p}^2+\alpha\mathbf{E}\|\widehat{a}-a_\dagger\|_s^2 =&\mathcal{O}\Big(\delta^2+\alpha\mathbf{E}^{\frac12\frac{-s+q}{s+p}}\left(\|\widehat{a}-a_\dagger\|_{-p}^2\right)\mathbf{E}^{\frac12\frac{2s-q+p}{s+p}}\left(\|\widehat{a}-a_\dagger\|_s^2\right)\\ &+\delta\mathbf{E}^{\frac12}\left(\|\widehat{a}-a_\dagger\|_{-p}^2\right)\Big)\end{aligned}$$

using also assumption (2). In the following, we will make repeated use of the implication

$$c^r\le e+dc^t\Rightarrow c^r=\mathcal{O}\left(e+d^{\frac{r}{r-t}}\right), \tag{18}$$

which holds for $0\le t<r$ and $c,d,e>0$. This inequality is applied to the previous estimate for $c=\mathbf{E}^{\frac12}(\|\widehat{a}-a_\dagger\|_{-p}^2)$ and $r=2$. First we take $t=1$, $d=\delta$ and $e=\delta^2+\alpha\mathbf{E}^{\frac12\frac{-s+q}{s+p}}\left(\|\widehat{a}-a_\dagger\|_{-p}^2\right)\mathbf{E}^{\frac12\frac{2s-q+p}{s+p}}\left(\|\widehat{a}-a_\dagger\|_s^2\right)$ and we obtain

$$\mathbf{E}\|\widehat{a}-a_\dagger\|_{-p}^2=\mathcal{O}\left(\delta^2+\alpha\mathbf{E}^{\frac12\frac{-s+q}{s+p}}\left(\|\widehat{a}-a_\dagger\|_{-p}^2\right)\mathbf{E}^{\frac12\frac{2s-q+p}{s+p}}\left(\|\widehat{a}-a_\dagger\|_s^2\right)\right),$$

then we choose $t=\frac{q-s}{p+s}$, $d=\alpha\mathbf{E}^{\frac12\frac{2s-q+p}{s+p}}\left(\|\widehat{a}-a_\dagger\|_s^2\right)$ and $e=\delta^2$ and we get

$$\mathbf{E}\|\widehat{a}-a_\dagger\|_{-p}^2=\mathcal{O}\left(\delta^2+\alpha^{\frac{2(s+p)}{3s+2p-q}}\mathbf{E}^{\frac{2s-q+p}{3s+2p-q}}\left(\|\widehat{a}-a_\dagger\|_s^2\right)\right).$$

Replacing the term that contains $\mathbf{E}\|\widehat{a}-a_\dagger\|_{-p}^2$ on the right hand side and using the inequality $(x+y)^r \le x^r + y^r$ for $0 \le r \le 1$ we obtain

$$\begin{aligned}\mathbf{E}\|\widehat{a}-a_\dagger\|_s^2 =& \mathcal{O}\Big(\alpha^{-1}\delta^2 + \delta\alpha^{\frac{q-2s-p}{3s+2p-q}}\mathbf{E}^{\frac{p}{2}\frac{2s-q+p}{3s+2p-q}}\left(\|\widehat{a}-a_\dagger\|_s^2\right)\\ &+\delta^{\frac{-s+q}{s+p}}\mathbf{E}^{\frac{1}{2}\frac{2s-q+p}{s+p}}\left(\|\widehat{a}-a_\dagger\|_s^2\right) + \alpha^{\frac{-s+q}{3s+2p-q}}\mathbf{E}^{\frac{2s-q+p}{3s+2p-q}}\left(\|\widehat{a}-a_\dagger\|_s^2\right)\Big).\end{aligned}$$

Applying (18) repeatedly for $c = \mathbf{E}\|\widehat{a}-a_\dagger\|_s^2$, $r=1$ and $t = \frac{2s-q+p}{2(3s+2p-q)}$, $t = \frac{1}{2}\frac{2s-q+p}{s+p}$ and $t = \frac{2s-q+p}{3s+2p-q}$ we obtain

$$\mathbf{E}\|\widehat{a}-a_\dagger\|_s^2 = \mathcal{O}\left(\alpha^{-1}\delta^2 + \delta^{\frac{2(3s+2p-q)}{4s+3p-q}}\alpha^{\frac{2(q-2s-p)}{4s+3p-q}} + \delta^{\frac{2(q-s)}{q+p}}\right) = \mathcal{O}\left(\delta^{2\frac{q-s}{p+q}}\right).$$

As $\alpha \sim \delta^{\frac{2(p+s)}{p+q}}$ it follows that

$$\begin{aligned}\mathbf{E}\|\widehat{a}-a_\dagger\|_{-p}^2 &= \mathcal{O}\left(\delta^2 + \alpha^{\frac{2(s+p)}{3s+2p-q}}\mathbf{E}^{\frac{2s-q+p}{3s+2p-q}}\left(\|\widehat{a}-a_\dagger\|_s^2\right)\right)\\ &= \mathcal{O}\left(\delta^2 + \delta^{4\frac{(p+s)^2}{(p+q)(3s+2p-q)}+2\frac{2s-q+p}{3s+2p-q}\frac{q-s}{p+q}}\right) = \mathcal{O}\left(\delta^2\right).\end{aligned}$$

Taking the mean in the squared interpolation inequality (8) and using Jensen's inequality we get

$$\begin{aligned}\mathbf{E}\left(\|\widehat{a}-a_\dagger\|_0^2\right) &\le \mathbf{E}\left(\|\widehat{a}-a_\dagger\|_{-p}^{\frac{2s}{s+p}}\|\widehat{a}-a_\dagger\|_s^{\frac{2p}{s+p}}\right)\\ &\le \mathbf{E}\left(\|\widehat{a}-a_\dagger\|_{-p}^2\right)^{\frac{s}{s+p}}\mathbf{E}\left(\|\widehat{a}-a_\dagger\|_s^2\right)^{\frac{p}{s+p}}\\ &= \mathcal{O}\left(\delta^{\frac{2s}{p+s}}\delta^{2\frac{q-s}{q+p}\frac{p}{s+p}}\right) = \mathcal{O}\left(\delta^{\frac{2q}{q+p}}\right).\end{aligned}$$

□

Remark 2.2. *In some cases the rates of convergence with respect to other norms in the Hilbert scale are also of interest. Introducing $r \in [-p, s]$ as an additional parameter, a straightforward modification of the very last step of the previous proof yields the estimate*

$$\sqrt{\mathbf{E}\|\widehat{a}-a_\dagger\|_r^2} = \mathcal{O}\left(\delta^{\frac{q-r}{p+q}}\right), \qquad \delta \to 0. \tag{19}$$

The choice of the regularization parameter in Theorem 2.1, $\alpha \sim \delta^{\frac{2(p+s)}{p+q}}$ requires the knowledge of the smoothness index q of the solution, which is typically unknown in practice. Therefore, *a-posteriori* or *adaptive parameter choice rules* have to be used. From the rich literature on this subject we only cite [7, 28, 10]. A Lepskiĭ-type parameter choice rule for the problem at hand has been studied in [24] and will be published elsewhere.

3. Optimality

In the following we show that our two-step method to estimate $a_\dagger$ from noisy data described by (3) and (4) is capable of achieving the optimal rates provided that the direct regression problem in the first step is solved in an optimal way.

For this end we consider a situation where lower bounds on the expected squared error are known. Let $F = T : \mathcal{X} \to \mathcal{Y}$ be a linear, compact injective operator with polynomially decaying singular values. More precisely, we assume that the singular values $\sigma_j(T)$ decay like

$$\sigma_j(T) \sim j^{-p/d}, \qquad j \in \mathbf{N}$$

for some $p, d > 0$, i.e. there exists a constant $c > 0$ such that $j^{-p/d}/c \leq \sigma_j(T) \leq cj^{-p/d}$ for all $j \in \mathbf{N}$. (Later in our applications where T acts in Sobolev scales, d will be the space dimension and p the degree of ill-posedness as in (10).) Consider the Hilbert scale $(\mathcal{X}_s)_{s\in\mathbf{R}}$ generated by the self-adjoint operator $L := (T^*T)^{-1/2p}$ defined on the dense subspace $D(L) := (T^*T)^{1/2p}(\mathcal{X})$. Then the eigenvalues of L behave like

$$\lambda_j(L) \sim j^{1/d}.$$

It is well-known that a lower bound on the expected squared error of an estimator $\widehat{a}$ of $a_\dagger$ for the white noise model (4) with $u_\dagger = Ta_\dagger$ under the smoothness assumption $\|a_\dagger\|_q \leq 1$ is given by

$$\inf_{\widehat{a}} \sup_{\|a_\dagger\|_q \leq 1} \sqrt{\mathbf{E}\|\widehat{a}(\mathbf{Y}) - a_\dagger\|^2_{\mathcal{X}}} \geq c\sigma^{\frac{q}{q+p+d/2}} \tag{20}$$

for some $c > 0$ (see [15, 21]).

To show that this lower bound is achieved by our two-step method, we first study the direct regression problem to estimate $u_\dagger$ from $\mathbf{Y}$. Let $\{(v_j, u_j, \sigma_j(T)) : j \in \mathbf{N}\}$ denote a singular system of T, i.e. $Ta = \sum_{j=1}^{\infty} \sigma_j(T) \langle a, v_j\rangle_{\mathcal{X}} u_j$ for $a \in \mathcal{X}$. For theoretical purposes we may assume that $T(\mathcal{X})$ is dense in $\mathcal{Y}$ (otherwise we can use the orthogonal projection onto $\overline{T(\mathcal{X})}$). Then $\{u_j : j \in \mathbf{N}\}$ is a complete orthonormal system in $\mathcal{Y}$. The assumption that $a_\dagger \in \mathcal{X}_q$ and $\|a_\dagger\|_q \leq 1$ is equivalent to the fact that $w := L^q a_\dagger$ is well-defined and $\|w\|_{\mathcal{X}} \leq 1$. The Fourier coefficients of $u_\dagger$ satisfy

$$|\langle u_\dagger, u_j\rangle_{\mathcal{Y}}| = |\langle TL^{-q}w, u_j\rangle_{\mathcal{Y}}| = |\langle w, L^{-q}T^*u_j\rangle_{\mathcal{X}}| \sim j^{-(q+p)/d}|\langle w, v_j\rangle_{\mathcal{X}}|.$$

Therefore, by Pinskers Theorem (see [23, 30]) an optimal estimator $\widehat{u}$ satisfies

$$\sqrt{\mathbf{E}\|\widehat{u} - u_\dagger\|^2_{\mathcal{Y}}} \leq \delta \qquad \text{with} \qquad \delta \sim \sigma^{\frac{p+q}{p+q+d/2}}. \tag{21}$$

In a second step we can compute the estimator $\widehat{a}$ defined in (9) with the natural choice $a_0 = 0$. Since

$$\|Ta\|^2_{\mathcal{Y}} = (a, T^*Ta)_{\mathcal{X}} = \|(T^*T)^{1/2}a\|^2_{\mathcal{X}} = \|L^{-p}a\|^2_{\mathcal{X}} = \|a\|^2_{-p}$$

for all $a \in \mathcal{X}$, Assumption 2 is satisfied. Moreover, Assumptions 1 and 3 are trivially satisfied with $C_L = 0$. Therefore, Theorem 2.1 yields

$$\sqrt{\mathbf{E}\|\widehat{a} - a_\dagger\|^2_{\mathcal{X}}} = \mathcal{O}\left(\delta^{\frac{q}{p+q}}\right) = \mathcal{O}\left(\sigma^{\frac{q}{p+q+d/2}}\right).$$

This coincides with the lower bound (20). Since the two-step procedure described above yields an order-optimal error bound, in particular the error bound for the second step in Theorem 2.1 has to be of optimal order under the given assumptions.

Remark 3.1. *It seems that for nonlinear problems additional assumptions on F are required to guarantee that the direct regression problem can be solved with the accuracy (21). We refrain from discussing this problem in the general functional analytic setting, but show for the problems considered in the following sections that this accuracy can be achieved.*

4. Reconstruction of a reaction coefficient

We consider the problem of identification of the parameter $a \in L^2(\Omega)$ in the boundary value problem

$$\begin{cases} -\Delta u + au = f \text{ in } \Omega \\ u = g \text{ on } \partial\Omega \end{cases} \tag{22}$$

where $\Omega \subseteq \mathbf{R}^d$, $d \in \{1,2,3\}$ is a bounded smooth domain, and $f : \Omega \to \mathbf{R}$ and $g : \partial\Omega \to \mathbf{R}$ are smooth functions. For a given bound $\gamma > 0$ we introduce the set

$$D(F) = \{a \in L^2(\Omega) : 0 \le a \le \gamma\}. \tag{23}$$

It follows from classical results on elliptic partial differential equations (see e.g. [31]) that for $a \in D(F)$ and $g \in H^{\frac{3}{2}}(\partial\Omega)$ there exists a unique solution $u_a \in H^2(\Omega)$ of the direct problem (22). The inverse problem which we consider can be formulated as an operator equation with the parameter-to-solution operator $F : D(F) \to L^2(\Omega)$, $F(a) := u_a$.

In this section the Hilbert scale $(\mathcal{X}_s)_{s\in\mathbf{R}}$ will be generated by the square root $B^{\frac{1}{2}}$ of the positive, self-adjoint operator B defined by

$$B : H_0^1(\Omega) \cap H^2(\Omega) \to L^2(\Omega), \qquad Bv := -\Delta v + v. \tag{24}$$

Remark 4.1. *The first elements of this Hilbert scale with integer index are given by*

$$\mathcal{X}_1 = H_0^1(\Omega), \quad \mathcal{X}_2 = H_2(\Omega) \cap H_0^1(\Omega), \quad \mathcal{X}_3 = \{v \in H_3(\Omega) : v, \Delta v \in H_0^1(\Omega)\}.$$

This sheds some light on role of the initial guess a_0 in Theorem 2.1. For condition (12) to be fulfilled for a large index q, it is not sufficient that $a_\dagger$ and a_0 be smooth (e.g. C^∞), but additionally the boundary values of $a_\dagger$ and its derivatives must be known a-priori. This a-priori knowledge (if available) must be incorporated in the initial guess a_0. We have observed in numerical experiments that such a-priori knowledge of the behavior of $a_\dagger$ at the boundary is also necessary for fast convergence.

In the following analysis we will need operators of the form

$$\begin{aligned} T(a) :& H^2(\Omega) \cap H_0^1(\Omega) \to L^2(\Omega) \\ & T(a)v = -\Delta v + av, \end{aligned}$$

with $a \in D(F)$, which are self-adjoint as a densely defined unbounded operators in $L^2(\Omega)$, will play a dominant role. In the following lemma we are going to establish some properties of these operators, following ideas from [4].

Lemma 4.2. *There exist two strictly positive constants k_1 and k_2 such that*

$$k_1\|v\|_{\mathcal{X}_2} \le \|T(a)v\|_{L^2(\Omega)} \le k_2\|v\|_{\mathcal{X}_2}, \tag{25a}$$

$$k_1\|v\|_{L^2(\Omega)} \le \|T(a)v\|_{\mathcal{X}_{-2}} \le k_2\|v\|_{L^2(\Omega)} \tag{25b}$$

for all $v \in H^2(\Omega) \cap H_0^1(\Omega)$ and all $a \in D(F)$.
Furthermore, there exists a positive constant C such that

$$\|\left(T(a)^{-1} - T(\overline{a})^{-1}\right)\varphi\|_{\mathcal{X}_2} \le C\|a - \overline{a}\|_{L^2(\Omega)}\|\varphi\|_{L^2(\Omega)}. \tag{26}$$

for any $a, \overline{a} \in D(F)$, $\varphi \in L^2(\Omega)$.

Proof. The existence of a constant k_2 independent of a in (25a) is easily seen from the definition of the operator $T(a)$ as

$$\begin{aligned}\|T(a)v\| &\leq \| \triangle v\|_{L^2(\Omega)} + \|av\|_{L^2(\Omega)} \\ &\leq \| \triangle v\|_{L^2(\Omega)} + \gamma\|v\|_{L^2(\Omega)} \\ &\leq \max(1,\gamma)\|v\|_{H^2(\Omega)}.\end{aligned}$$

Now we prove the first inequality in (25a) where the constant k_1 is also independent of $a \in D(F)$. The L^2-norm of $T(a)v$ can be estimated from below by

$$\begin{aligned}\|T(a)v\|^2_{L^2(\Omega)} &= \int_\Omega \left(- \triangle v + av\right)^2 \, dx \\ &= \int_\Omega \left((\triangle v)^2 - 2av \triangle v + a^2v^2\right) \, dx \\ &\geq \| \triangle v\|^2_{L^2(\Omega)} - 2\int_\Omega |av \triangle v| \, dx + \int_\Omega a^2v^2 \, dx. \end{aligned} \tag{27}$$

From the inequality $2|av \triangle v| \leq 4a^2v^2 + \dfrac{(\triangle v)^2}{4}$ we have that

$$\begin{aligned}\|T(a)v\|^2_{L^2(\Omega)} &\geq \| \triangle v\|^2_{L^2(\Omega)} - \int_\Omega \frac{(\triangle v)^2}{4} \, dx - \int_\Omega 4a^2v^2 \, dx + \int_\Omega a^2v^2 \, dx \\ &\geq \frac{3}{4}\| \triangle v\|^2_{L^2(\Omega)} - 3\int_\Omega a^2v^2 \, dx \\ &\geq \frac{3}{4}\| \triangle v\|^2_{L^2(\Omega)} - 3\gamma^2\|v\|^2_{L^2(\Omega)}. \end{aligned} \tag{28}$$

On the other hand

$$\langle T(a)v, v\rangle = \int_\Omega (- \triangle v + av) \, v \, dx = \int_\Omega | \bigtriangledown v|^2 + av^2 \, dx \geq \int_\Omega | \bigtriangledown v|^2 \, dx \geq \frac{1}{k}\|v\|^2_{H^1}.$$

where k is the Poincaré constant and is independent of $v \in H^1_0(\Omega)$. Applying Cauchy-Schwarz inequality we obtain that

$$\frac{1}{k}\|v\|^2_{L^2(\Omega)} \leq \frac{1}{k}\|v\|^2_{H^1} \leq \|T(a)v\|_{L^2(\Omega)}\|v\|_{L^2(\Omega)},$$

and hence

$$\frac{1}{k}\|v\|_{L^2(\Omega)} \leq \|T(a)v\|_{L^2(\Omega)}. \tag{29}$$

Multiplying the squared inequality (29) by $3\gamma^2k^2$ and adding this to (28) yields

$$\left(1 + 3\gamma^2k^2\right) \|T(a)v\|^2_{L^2(\Omega)} \geq \| \triangle v\|^2_{L^2(\Omega)} \geq K^2\|v\|^2_{H^2(\Omega)},$$

using the global regularity result $\| \triangle u\|_{L^2(\Omega)} \geq K\|u\|_{H^2(\Omega)}$ with constant K independent of $u \in H^2(\Omega) \cap H^1_0(\Omega)$. This completes the proof of the first inequality in (25a).
(25b) follows from (25a) and Lemma 4.3 below since $T(a)$ is self- adjoint with respect to the L^2- norm and boundedly invertible from $\mathcal{X}_2$ to $\mathcal{X}_0$.

In order to prove (26) we consider the difference

$$v = \left(T(a)^{-1} - T(\overline{a})^{-1}\right) \varphi = v_a - v_{\overline{a}},$$

where $v_a = T(a)^{-1}\varphi$ and $v_{\overline{a}} = T(\overline{a})^{-1}\varphi$. As

$$T(a)v = \varphi - T(a)v_{\overline{a}} = (T(\overline{a}) - T(a))v_{\overline{a}} = (\overline{a} - a) \, v_{\overline{a}},$$

it follows that

$$\begin{aligned}\|v\|_{\mathcal{X}_2} &= \|T(a)^{-1}\,(\overline{a}-a)\,v_{\overline{a}}\|_{\mathcal{X}_2} \le k_1^{-1}\|\,(\overline{a}-a)\,v_{\overline{a}}\|_{L^2(\Omega)} \\ &\le k_1^{-1}\|a-\overline{a}\|_{L^2(\Omega)}\|v_{\overline{a}}\|_\infty \le ck_1^{-1}\|a-\overline{a}\|_{L^2(\Omega)}\|v_{\overline{a}}\|_{\mathcal{X}_2} \\ &\le ck_1^{-2}\|a-\overline{a}\|_{L^2(\Omega)}\|\varphi\|_{L^2(\Omega)}.\end{aligned}$$

□

In this section we repeatedly use the following elementary result from functional analysis:

Lemma 4.3. *Let $V_j \subset H_j \subset V_j'$ ($j=1,2$) be Gelfand triples and $A \in L\,(V_1, V_2)$ an operator with adjoint $A^* \in L\,(V_2', V_1')$ i.e. $\langle A\varphi_1, \varphi_2\rangle_{H_2} = \langle \varphi_1, A^*\varphi_2\rangle_{H_1}$ for all $\varphi_1 \in V_1$ and $\varphi_2 \in V_2'$. Then*

$$\|A\|_{V_2\leftarrow V_1} = \|A^*\|_{V_1'\leftarrow V_2'}. \tag{30}$$

Moreover, if A has a bounded inverse $A^{-1} : V_2 \to V_1$ then $\left(A^{-1}\right)^* = \left(A^*\right)^{-1}$ and

$$\|A^{-1}\|_{V_1\leftarrow V_2} = \|\left(A^{-1}\right)^*\|_{V_2'\leftarrow V_1'}. \tag{31}$$

We will assume that the exact solution $u_\dagger = u_{a_\dagger}$ of the equation (22) fulfills the condition

$$c_u = \inf_{x\in\Omega} u_\dagger\,(x) > 0 \tag{32}$$

which is satisfied if $g > 0$ and $f \ge 0$ by the maximum principle.

Theorem 4.4. *If (32) holds true, then F satisfies Assumption 1 for $s \ge 2$.*

Proof. The Fréchet differentiability of F has been established in [4, 5]. Injectivity follows easily from (32) as the relation $u_a = u_\dagger$ implies $(a - a_\dagger)u_\dagger = 0$ for any $a \in D(F)$. Since $D(F) \cap (a_0 + \mathcal{X}_s)$ is closed and convex for $s \ge 2$, it is weakly sequentially closed. Hence, for a weakly convergent sequence $(a_n)_{n\in\mathbf{N}}$ in $D(F) \cap (a_0 + \mathcal{X}_s)$ the limit point a belongs to $D(F) \cap (a_0 + \mathcal{X}_s)$. Since $d \le 3$, it follows from the compactness of the embedding $\mathcal{X}^s \hookrightarrow \mathcal{X}^{7/4}$ and the continuity of the embedding $\mathcal{X}^{7/4} \hookrightarrow L^\infty(\Omega)$ that there exists a weakly convergent subsequence (a_{n_k}) in $L^\infty(\Omega)$. Since F is continuous (even Fréchet differentiable) on $L^\infty(\Omega)$, it follows that $\|F(a_{n_k}) - F(a)\|_{\mathcal{Y}} \to 0$ as $k \to \infty$. This shows that F is weakly sequentially closed. □

Differentiating (22) with respect to a shows that the Fréchet derivative $F'(a)h = v_h$ is the unique solution v_h of the equation

$$T(a)v_h = -hu_a, \tag{33}$$

where u_a is the solution of (22). Hence, the adjoint of the Fréchet derivative is given by

$$F'(a)^*\varphi = -u_aT(a)^{-1}\varphi, \tag{34}$$

where $\varphi \in L^2(\Omega)$.

Theorem 4.5. *If assumption (32) holds true and $(\mathcal{X}_s)_{s\in\mathbf{R}}$ is the Hilbert scale generated by $\sqrt{B}$, then F' satisfies Assumption 2 with $p = 2$. Moreover, Assumption 3 is satisfied if γ in (23) is sufficiently small.*

Proof. To prove Assumption 2 we use the chain of inequalities (25b) and the relation (33). It follows that

$$\|F'(a)h\|_{L^2(\Omega)} = \|T(a)^{-1}(-hu_a)\|_{L^2(\Omega)} \sim \|hu_a\|_{\mathcal{X}_{-2}}, \tag{35}$$

with constants independent of a. From the Banach algebra property for the Sobolev space $H^m(\Omega)$(see [1] for details) it holds that

$$\|vw\|_{H^m(\Omega)} \leq K\|v\|_{H^m(\Omega)}\|w\|_{H^m(\Omega)}, \tag{36}$$

where $u, v \in H^m(\Omega)$, Ω is a domain in $\mathbf{R}^d$ satisfying the cone condition, $2m > d$ and the constant K depends on m, d and the cone determining the cone condition for Ω. Choosing $m = 2$ and $d = 1, 2, 3$ in (36) we have

$$\|vw\|_{H^2(\Omega)} \leq K\|v\|_{H^2(\Omega)}\|w\|_{H^2(\Omega)}, \tag{37}$$

and from relation (30) the multiplication operator $M_{u_a} : h \to hu_a$ has bounded norm

$$\|M_{u_a}\|_{\mathcal{X}_{-2}\leftarrow\mathcal{X}_{-2}} = \|M_{u_a}\|_{\mathcal{X}_2\leftarrow\mathcal{X}_2} \leq K\|u_a\|_{H^2(\Omega)}. \tag{38}$$

Due to (32), $\left\|\frac{1}{u_a}\right\|_{H^2(\Omega)}$ is finite, and hence

$$\|M_{u_a}^{-1}\|_{\mathcal{X}_{-2}\leftarrow\mathcal{X}_{-2}} = \|M_{u_a}^{-1}\|_{\mathcal{X}_2\leftarrow\mathcal{X}_2} \leq K\left\|\frac{1}{u_a}\right\|_{H^2(\Omega)} < \infty.$$

This implies $\|hu_a\|_{\mathcal{X}_{-2}} \sim \|h\|_{\mathcal{X}_{-2}}$. Together with (35) this finishes the proof of Assumption 2.

Due to (30), it suffices to show Lipschitz continuity of the mapping $a \to F'(a)^*$ from $\mathcal{X}_0$ to $L\left(L^2(\Omega), \mathcal{X}_2\right)$ to prove Assumption 3. From eq. (34) we can write

$$[F'(a)]^* = M_{u_a} \circ T(a)^{-1}.$$

It is going to be proved now that

$$\|(M_{u_a} \circ T(a)^{-1} - M_{u_{\overline{a}}} \circ T(\overline{a})^{-1})\varphi\|_{\mathcal{X}_2} \leq c\|\varphi\|_{L^2(\Omega)}\|a - \overline{a}\|_{L^2(\Omega)},$$

where c is a constant which can vary from now on. We split the difference in two terms which can be estimated

$$\begin{aligned}\|(M_{u_a} \circ T(a)^{-1} - M_{u_{\overline{a}}} \circ T(\overline{a})^{-1})\varphi\|_{\mathcal{X}_2} &\leq \|M_{u_a}[T(a)^{-1} - T(\overline{a})^{-1}]\varphi\|_{\mathcal{X}_2} \\ &\quad + \|(M_{u_a} - M_{u_{\overline{a}}})T(\overline{a})^{-1}\varphi\|_{\mathcal{X}_2}.\end{aligned} \tag{39}$$

The first term on the right hand side of the inequality (39) can be bounded using (26) and (38). If $Z : H^{\frac{3}{2}}(\partial\Omega) \to H^2(\Omega)$ is a right inverse of the trace operator, we have $F(a) = u_a = \tilde{u}_a + Zg$ with $\tilde{u}_a = T(a)^{-1}(f - Zg)$ and $F(\overline{a}) = u_{\overline{a}} = \tilde{u}_{\overline{a}} + Zg$ and $\tilde{u}_{\overline{a}} = T(\overline{a})^{-1}(f - Zg)$. It follows from Lemma 4.2 that

$$\begin{aligned}\|u_a - u_{\overline{a}}\|_{\mathcal{X}_2} &= \|\left(T(a)^{-1} - T(\overline{a})^{-1}\right)(f - Zg)\|_{\mathcal{X}_2} \\ &\leq C\|a - \overline{a}\|_{L^2(\Omega)}\|f - Zg\|_{L^2(\Omega)}.\end{aligned}$$

Together with the Banach algebra property (37) and (25a) we obtain

$$\begin{aligned}\|(M_{u_a} - M_{u_{\overline{a}}})T(\overline{a})^{-1}\varphi\|_{\mathcal{X}_2} &= \|(u_a - u_{\overline{a}})T(\overline{a})^{-1}\varphi\|_{\mathcal{X}_2} \leq \frac{K}{k_1}\|u_a - u_{\overline{a}}\|_{\mathcal{X}_2}\|\varphi\|_{L^2(\Omega)} \\ &\leq \left(\frac{CK}{k_1}\|f - Zg\|_{L^2(\Omega)}\right)\|a - \overline{a}\|_{L^2(\Omega)}\|\varphi\|_{L^2(\Omega)}.\end{aligned}$$

This proves that our operator satisfies Assumption 3. ∎

For the direct regression problem in the first step we have to determine the precise regularity of $u_\dagger$:

Proposition 4.6. *If $a_\dagger \in H^q(\Omega)$ with $q > d/2$, then $u_\dagger \in H^{q+2}(\Omega)$.*

Proof. The regularity result can be derived as follows from standard elliptic regularity estimates where smoothness of $a_\dagger$ is measured in terms of L^∞ based Sobolev spaces of integer order: By Sobolev's embedding theorem we have $a_\dagger \in W^{q-d/2-\epsilon,\infty}(\Omega)$ for any $\epsilon > 0$, and in particular $a_\dagger \in W^{k,\infty}(\Omega)$ with $k := \sup\{\tilde{k} \in \mathbf{N}_0 : \tilde{k} < q - d/2\}$ (see [25]). By standard elliptic regularity estimates we get $u_\dagger \in H^{k+2}(\Omega)$. To improve this regularity, we rewrite the differential equation as

$$-\Delta u = \tilde{f} \qquad \text{with} \qquad \tilde{f} := f - au.$$

Since f is assumed to be smooth and $q, k+2 > d/2$, we have $\tilde{f} \in H^{\min(k+2,q)}(\Omega)$ due to (36). First assume that $k + 2 \geq q$. Then $u_\dagger \in H^{q+2}(\Omega)$ by regularity results for Poisson's equation. If $k + 2 < q$, we have to repeat the last argument to further improve the regularity of $u_\dagger$. □

Corollary 4.7. *Assume the noise model (4), let the assumptions of Theorem 4.5 hold true, let $s \geq 2$, $q \in [s, 2s+2]$, and assume that $a_\dagger - a_0 \in \mathcal{X}_q$ and $a_\dagger \in H^q(\Omega)$. Then $u_\dagger$ can be estimated such that (2) holds true with $\delta \sim \sigma^{\frac{q}{q+2+d/2}}$. Moreover, for the choice $\alpha \sim \sigma^{\frac{2q(2+s)}{(q+2+d/2)(2+q)}}$ we get*

$$\sqrt{\mathbf{E}\|\widehat{a} - a_\dagger\|^2_{L^2(\Omega)}} = \mathcal{O}\left(\sigma^{\frac{q}{q+2+d/2}}\right), \qquad \sigma \to 0.$$

Proof. The first statement follows from Proposition 4.6 and general results from nonparametric regression (see [30]). The second statement is a consequence of Theorems 2.1, 4.4 and 4.5. □

5. Reconstruction of a diffusion coefficient

Let us consider now the problem of identifying the parameter a in the boundary value problem

$$\left\{ \begin{array}{l} -(au')' = f \text{ on } (0,1), \\ u(0) = g_0, \quad u(1) = g_1. \end{array} \right. \tag{40}$$

Here the parameter a belongs to the set of admissible functions $D(F) = \{q \in H^2((0,1)) : \underline{\gamma} \leq q, \|q\|_{H^2} \leq \gamma\}$ with strictly positive constants $\underline{\gamma}$ and γ and f is a smooth function. Under these assumptions there exists a unique weak solution u_a to (40) for a given parameter a, and u_a belongs to the Sobolev space of functions $H^2 = H^2(0,1)$ (see [31]). In the following we will always omit the domain $(0,1)$ in function spaces. Our inverse problem is formulated as an operator equation (1) with the parameter-to-solution operator $F : D(F) \to L^2$, $F(a) = u_a$.

It turns out that Assumption 2 is neither satisfied for the Hilbert scale generated by the Dirichlet Laplacian nor the Hilbert scale generated by the shifted Neumann Laplacian. Therefore, we introduce a different Hilbert scale starting from the space

$$L^2_\diamond := \left\{ v \in L^2 : \int_0^1 v \, dx = 0 \right\}$$

and the operator

$$B : D(B) = \left\{ u \in H^2 : u'(0) = u'(1) = 0, \int_0^1 u\, dx = 0 \right\} \to L^2_\diamond, \tag{41}$$
$$B\phi := -\phi'' + \phi.$$

B is well defined as

$$\int_0^1 B\phi\, dx = \int_0^1 (-\phi'' + \phi)\, dx = -\phi'(1) + \phi'(0) + \int_0^1 \phi\, dx = 0.$$

B is also densely defined, as H^2 is dense in L^2. It is also easy to see that B is symmetric, as for any $\varphi, \phi \in D(B)$ we have

$$\begin{aligned}\langle B\phi, \varphi \rangle &= \int_0^1 (-\phi'' + \phi)\varphi\, dx = (-\phi'\varphi)\mid_0^1 + \int_0^1 \phi'\varphi'\, dx + \int_0^1 \phi\varphi\, dx \\ &= (\phi\varphi')\mid_0^1 - \int_0^1 \phi\varphi''\, dx + \int_0^1 \phi\varphi\, dx = \langle \phi, B\varphi \rangle.\end{aligned}$$

It follows that the operator $(D(B^*), B^*)$ is an extension of $(D(B), B)$, so $D(B) \subseteq D(B^*)$. B is self-adjoint if $D(B) = D(B^*)$. As the operator $(D(B), B)$ is bijective (see [29, Prop. 5.7.5]), it follows from [29, Prop. A 8.3] that B is also self-adjoint. Moreover, B is strictly positive as

$$\langle B\phi, \phi \rangle = \int_0^1 (\phi'^2 + \phi^2)\, dx \geq \|\phi\|_{L^2}.$$

First we consider the Hilbert scale $\left(\widehat{X}_s\right)_{s\in\mathbf{R}}$ generated by the operator $B^{\frac{1}{2}}$. However, to prove Assumption 3 we need $\mathcal{X}$ to be a subspace of H^1, so we choose the shifted scale

$$\mathcal{X}_s = \widehat{\mathcal{X}}_{s+1}, \tag{42}$$

which is also a Hilbert scale (see e.g. [7, Prop. 8.19]). The first integer elements of this scale are given by

$$\begin{aligned}&\mathcal{X}_{-1} = L^2_\diamond, \qquad \mathcal{X}_0 = H^1_\diamond := H^1 \cap L^2_\diamond, \qquad \mathcal{X}_1 = D(B), \\ &\mathcal{X}_2 = H^3 \cap D(B), \qquad \mathcal{X}_3 = \{\phi \in H^4 \cap D(B) : \phi'''(0) = \phi'''(1) = 0\}.\end{aligned} \tag{43}$$

We introduce the orthogonal projection $P_\diamond : L^2 \to L^2_\diamond$, $v \mapsto v - \int_0^1 v\, dx$ and the embedding operator $J_\diamond : H^1_\diamond \hookrightarrow L^2_\diamond$. $J_\diamond$ is injective and, due to the density of H^1 in L^2, has also dense range. This also implies that its adjoint $J_\diamond^* : L^2_\diamond \to H^1_\diamond$ is injective and has dense range. By [18, Prop. 2.1] it follows that

$$J_\diamond^* f = B^{-1} f \tag{44}$$

for $f \in L^2_\diamond$.

We assume that the exact solution $u_\dagger$ of the direct problem (40) fulfills the condition

$$\inf_{x\in(0,1)} |u_\dagger'(x)| > 0. \tag{45}$$

Theorem 5.1. *If (45) holds true, then F satisfies Assumption 1 for $s \geq 2$.*

Proof. Under assumption (45) the injectivity of the operator F follows immediately from [13, Theorem 2.1]. Due to similar arguments as in section 4 the operator F is weakly sequentially closed between $D(F) \cap (a_0 + \mathcal{X}_s)$, $s \geq 2$ and L^2 for $s \geq 2$ (see [2]). The Fréchet differentiability of F has been established in [5]. $\square$

In order to prove the Assumptions 2 and 3 we introduce the operator

$$T(a) : H^2 \cap H_0^1 \to L^2, \qquad T(a)v = -\left(av'\right)'$$

which is self-adjoint as a densely defined operator on L^2. The Fréchet derivative $F'(a)h = \eta$ for a given perturbation $h \in H^1$ of a is the weak solution of the equation

$$\begin{cases} -\left(a\eta'\right)' = \left(hu_a'\right)', \\ \eta\left(0\right) = \eta\left(1\right) = 0 \end{cases} \tag{46}$$

(see [5]). Hence, F' can be written in operator form as

$$F'(a)h = S(a)D_x M_{u_a'} J_\diamond h, \tag{47}$$

where $M_{u_a'} : L^2_\diamond \to L^2$, $h \mapsto u_a' h$ is a multiplication operator, $D_x : L^2 \to H^{-1}$ is the differentiation operator, and $S(a) : H^{-1} \to L^2$ maps $f \in H^{-1}$ to the solution of the boundary value problem (46) with $\left(hu_a'\right)'$ replaced by f. (Note that actually the range of $S(a)$ is contained in H_0^1!) To compute $F'(a)^* : L^2 \to H^1_\diamond$ we have to find the adjoint of each factor in (47). Since $T(a)$ is self-adjoint in L^2, the adjoint of $S(a)$ with respect to the Gelfand triple $H^{-1} \subset L^2 \subset H_0^1$ is given by $S(a)^* : L^2 \to H_0^1$, $S(a)^*\varphi = T(a)^{-1}\varphi$. Moreover, $M^*_{u_a'} v = P_\diamond(u_a' \cdot v)$. From eq. (44) we get

$$F'(a)^*\varphi = -B^{-1}P_\diamond\left\{u_a'[T(a)^{-1}\varphi]'\right\}, \qquad \varphi \in L^2. \tag{48}$$

In order to prove the Assumptions 2 and 3 we need the following properties of $T(a)$.

Lemma 5.2. *For all $a \in H^2 \cap D(F)$ there exists two positive constants k_1 and k_2 such that*

$$k_1\|v\|_{H^2} \leq \|T(a)v\|_{L^2} \leq k_2\|v\|_{H^2}. \tag{49}$$

for $v \in H^2$, where k_1 and k_2 can be chosen independent of a. Furthermore, there exists a positive constant C such that

$$\|\left(T(a)^{-1} - T(\overline{a})^{-1}\right)\varphi\|_{H^2} \leq C\|a - \overline{a}\|_{H^1}\|\varphi\|_{L^2}. \tag{50}$$

for any $a, \overline{a} \in D(F)$, $\varphi \in L^2\left(\Omega\right)$.

Proof. The second inequality is easily seen from

$$\| -\left(av'\right)'\|_{L^2} \leq \|av'\|_{H^1} \leq K\|a\|_{H^1}\|v'\|_{H^1} \leq K\gamma\|v\|_{H^2},$$

where we used that

$$\|uv\|_{H^1} \leq K\|u\|_{H^1}\|v\|_{H^1} \tag{51}$$

for $u, v \in H^1$ and $K > 0$ independent of u and v (see (36)). The constant $k_2 = K\gamma$ is independent of a.

For the first inequality we use the elliptic a-priori estimate for the solution of equation $T(a)v = f$

$$\|v\|_{H^2} \leq C\left(\|v\|_{H^1} + \|f\|_{L^2}\right)a_{n_k},$$

with C depending on $\underline{\gamma}$ and $\|a\|_{W^{1,\infty}}$ (see e.g. [9, Theorem 8.13]). Since H^2 is continuously embedded in $W^{1,\infty}$, C depends only on $\underline{\gamma}$ and γ, but not on a. Multiplying in (40) by a test function $w \in H_0^1$ and integrating by parts, yields

$$l(v, w; a) = \int_0^1 f w \, dx, \tag{52}$$

with the bilinear form $l(.,.;a) : H_0^1 \times H_0^1 \to \mathbf{R}$ given by

$$l(v, w; a) = \int_0^1 a\,(x)\, v' \cdot w' \, dx.$$

It follows from Poincaré's inequality that l satisfies the inequality

$$k\|v\|_{H^1}^2 \le l(v, v; a) \tag{53}$$

for all $a \in D(F)$, $v \in H_0^1$, where k is a constant that depends on $\underline{\gamma}$ and the Poincaré constant. Moreover, $l(v, v; a) = \int_0^1 f v \, dx \le \|f\|_{L^2}\|v\|_{L^2} \le \|f\|_{L^2}\|v\|_{H^1}$. Hence, the solution v to (40) fulfills

$$\|v\|_{H^1} \le k^{-1}\|f\|_{L^2},$$

and the first statement is proved. (50) can be proved considering the difference

$$v = \left(T(a)^{-1} - T(\overline{a})^{-1}\right)\varphi = v_a - v_{\overline{a}},$$

where $v_a = T(a)^{-1}\varphi$ and $v_{\overline{a}} = T(\overline{a})^{-1}\varphi$. As

$$T(a)v = \varphi - T(a)v_{\overline{a}} = (T(\overline{a}) - T(a))v_{\overline{a}} = \left((\overline{a} - a)\, v_{\overline{a}}'\right)',$$

it follows that

$$\begin{aligned}\|v\|_{H^2} &= \|T(a)^{-1}\left((\overline{a} - a)\, v_{\overline{a}}'\right)'\|_{H^2} \le k_1^{-1}\|\left((\overline{a} - a)\, v_{\overline{a}}'\right)'\|_{L^2} \le Kk_1^{-1}\|a - \overline{a}\|_{H^1}\|v_{\overline{a}}'\|_{H^1} \\ &\le Kk_1^{-1}\|a - \overline{a}\|_{H^1}\|v_{\overline{a}}\|_{H^2} \le Kk_1^{-2}\|a - \overline{a}\|_{H^1}\|\varphi\|_{L^2}.\end{aligned}$$

□

As a preparation for the next theorem we rewrite (48) as

$$F'(a)^*\varphi = -B^{-1}AT(a)^{-1}\varphi, \qquad \varphi \in L^2, \tag{54}$$

with the operator

$$A : H^2 \cap H_0^1 \to H_\diamond^1, \qquad v \mapsto P_\diamond(u_\dagger' \cdot v').$$

Lemma 5.3. *The operator A is an isomorphism between $H^2 \cap H_0^1$ and $H_\diamond^1$.*

Proof. From the definition of A as the composition of three linear and bounded operators, we have that A is also linear and bounded. We prove now that A is injective. Let $v \in H^2 \cap H_0^1$ be such that $Av = 0$. It follows that $P_\diamond(u_\dagger' v') = 0$, which means $u_\dagger' v' = c$ with $c := \int_0^1 u_\dagger'(x)v'(x)\,dx$. Then $v' = \frac{c}{u_\dagger'}$. But from the Mean Value Theorem, as $v(0) = v(1) = 0$ there exists $\xi \in (0, 1)$ such that $v'(\xi) = 0$. Hence, $c = 0$, $v = 0$ and A is injective.

Next we prove that A is a Fredholm operator of index zero. The projection $P_\diamond : H^1 \to H_\diamond^1$ is a Fredholm operator of index 1 since it is surjective and its kernel consists only of the constant functions. As the operator $M_{u_\dagger'}$ is bijective due to the assumption (45), it is a Fredholm operator, and its index is zero. To prove that $D_x : H^2 \cap H_0^1 \to H^1$ is a Fredholm operator we consider the bounded linear operator

$$E : H^1 \to H_0^1 \cap H^2, \qquad (Ev)(x) := \int_0^x (P_\diamond v)(t)\, dt.$$

Since $P_\diamond D_x u = u' - \int_0^1 u'\, dx = u' - u(1) + u(0) = u'$ for $u \in H^2 \cap H_0^1$, we have

$$ED_x u = u, \qquad \text{and} \qquad D_x E v = P_\diamond v$$

for $v \in H^1$. Since E is surjective and its kernel consists of the constant functions, E is Fredholm with $\operatorname{ind}(E) = 1$. Moreover, E is a Fredholm inverse of D_x, and hence D_x is Fredholm with $\operatorname{ind}(D_x) = -\operatorname{ind}(E_x) = -1$.
As a composition of Fredholm operators A is also a Fredholm operator, and its index is

$$\operatorname{ind}(A) = \operatorname{ind}(P_\diamond) + \operatorname{ind}(M_{u'_\dagger}) + \operatorname{ind}(D_x) = 1 + 0 - 1 = 0.$$

Together with the injectivity of A shown in first part of the proof this yields the assertion. □

Theorem 5.4. *If (45) holds true and the Hilbert scale is defined as in (42) then Assumptions 2 holds true for $p = 2$. Moreover, Assumption 3 is satisfied if the diameter of $D(F) \cap (a_0 + \mathcal{X}_s)$ with respect to the $\mathcal{X}$-norm is sufficiently small.*

Proof. To prove Assumption 2 we actually show that $\|F'(a_\dagger)^*\varphi\|_{\mathcal{X}_2} \sim \|\varphi\|_{L^2}$ using Lemma 5.3 and (49). As $F'(a_\dagger)^*$ is a composition of three isomorphisms, the norm-equivalence follows immediately. The proof of Assumption 3 follows the steps of the proof of Theorem 4.5. Using (48) we get

$$\begin{aligned} \|F'(a)^*\varphi - F'(\overline{a}^*)\varphi\|_{\mathcal{X}_2} &= \left\| P_\diamond u'_a \left(T(a)^{-1}\varphi\right)' - P_\diamond u'_{\overline{a}} \left(T(\overline{a})^{-1}\varphi\right)' \right\|_{\mathcal{X}_0} \\ &\le \|P_\diamond\|_{\mathcal{X}_0 \leftarrow H^1} \left\{ \left\| u'_a \left(T(a)^{-1}\varphi - T(\overline{a})^{-1}\varphi\right)' \right\|_{\mathcal{X}_0} \right. \\ &\quad \left. + \left\| (u'_a - u'_{\overline{a}}) \left(T(a)^{-1}\varphi\right)' \right\|_{\mathcal{X}_0} \right\}. \end{aligned} \tag{55}$$

Since $u_a = \tilde{u}_a + g$ and $u_{\overline{a}} = \tilde{u}_{\overline{a}} + g$ with $g(x) := g_0 + x(g_1 - g_0)$, $\tilde{u}_a := T(a)^{-1}(f - g)$, and $\tilde{u}_{\overline{a}} := T(\overline{a})^{-1}(f - g)$, it follows from Lemma 5.2 that

$$\begin{aligned} \|u'_a - u'_{\overline{a}}\|_{H^1} \le \|u_a - u_{\overline{a}}\|_{H^2} &= \| \left(T(a)^{-1} - T(\overline{a})^{-1}\right)(f - g) \|_{H^2} \\ &\le C\|a - \overline{a}\|_{H^1}\|f - g\|_{L^2}. \end{aligned}$$

Therefore, the second term on the right hand side of (55) is bounded by a multiple of $\|a - \overline{a}\|_{H^1}\|\varphi\|_{L^2}$. Using the Banach algebra property (51) and (50) the first term on the right hand side of (55) can be bounded by $c\|a - \overline{a}\|_{H^1}\|\varphi\|_{L^2}$ as well. □

Proposition 5.5. *If $a_\dagger \in H^{q+1}$ with $q > \frac{1}{2}$, then $u_\dagger \in H^{q+2}$.*

Proof. It follows from the differential equation (40) that

$$u'_\dagger = \frac{\overline{f}}{a_\dagger},$$

where $\overline{f}$ is a primitive of f. As $a_\dagger \in D(F)$ is bounded from below, it follows that $1/a_\dagger \in H^{q+1}$. Hence, $u'_\dagger = \overline{f}/a_\dagger \in H^{q+1}$ since f is assumed to be smooth (see (36)). It follows that $u_\dagger \in H^{q+2}$. □

Let us summarize our results:

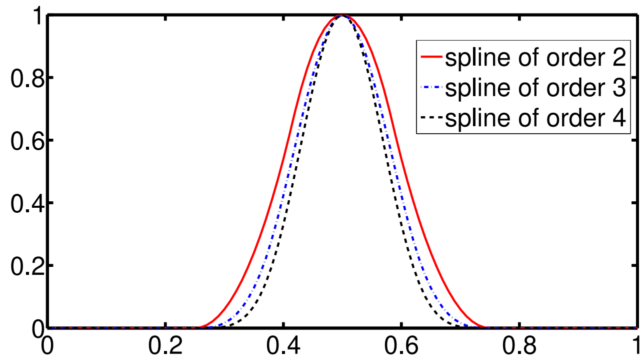

FIGURE 1. Parameter functions $a_\dagger$ with different degrees of smoothness used in the tests

Corollary 5.6. *In the case of the noise model (4), if the assumptions of Theorem (5.4) hold true and if $s \geq 2$, $q \in [s, 2s+2]$, $a_\dagger - a_0 \in \mathcal{X}_q$, and $a_\dagger \in H^{q+1}$, then $u_\dagger$ can be estimated such that (2) holds true with $\delta \sim \sigma^{\frac{q}{q+2+1/2}}$. Moreover, for the choice $\alpha \sim \sigma^{\frac{2q(2+s)}{(q+2+1/2)(2+q)}}$ we get*

$$\begin{aligned}\sqrt{\mathbf{E}\|\widehat{a} - a_\dagger\|^2_{H^1(\Omega)}} &= \mathcal{O}\left(\sigma^{\frac{q}{q+2+1/2}}\right),\\ \sqrt{\mathbf{E}\|\widehat{a} - a_\dagger\|^2_{L^2(\Omega)}} &= \mathcal{O}\left(\sigma^{\frac{q+1}{q+2+1/2}}\right).\end{aligned}$$

Proof. The first statement follows from Proposition 5.5 and well-known results from nonparametric regression (see [30]). The second statement is a consequence of Remark 2.2 with $r = 0$ and $r = -1$ and Theorems 5.1 and 5.4. □

From the characterization of the elements of the Hilbert scale $(\mathcal{X}_s)_{s\in\mathbf{R}}$ (see 43), it follows that for fast rates of convergence $a_\dagger$ must be smooth, and additionally the mean value of $a_\dagger$ and its odd derivatives at the boundary must be known and incorporated in the initial guess a_0.

6. Numerical experiments

In this final section we illustrate the influence of the smoothness of the parameter $a_\dagger$ on the rate of convergence of $\widehat{u}$ as shown in Theorem 2.1 by numerical simulation using the problem studied in section 4. We chose Ω to be the interval $[0, 1]$, and the unknown parameter $a_\dagger$ to be a B-spline of order 2, 3 or 4 which corresponds to smoothness up to $H^{2.5}$, $H^{3.5}$ and $H^{4.5}$ respectively (as the splines of order q belong to the intersection of H^r spaces with $0 < r < q$). Figure 1 presents the three different choices of the true parameter $a_\dagger$.

The direct problem was solved by finite differences. We used the noise model (3) with equidistant points $X_1, \ldots X_n$ and Gaussian errors ϵ_i. The direct regression problem was solved by a local polynomial estimator with Gaussian kernel and an optimally a-priori chosen bandwidth (see [30]). Fig. 2 shows the exact data $u_\dagger$ if $a_\dagger$ is a spline of order two and a simulated data sample of size $n = 398$ obeying the noise model (3) with $\epsilon_i \sim N(0, 0.01^2)$.

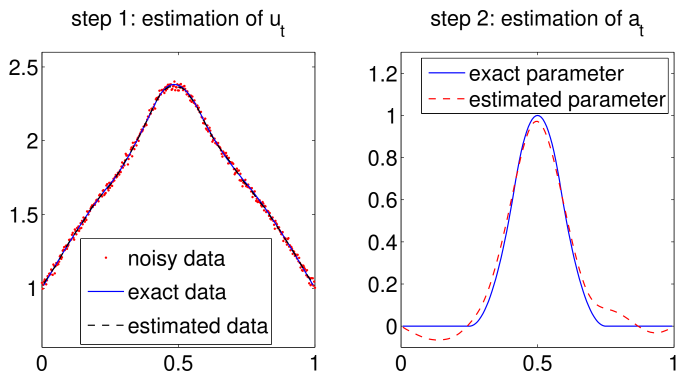

FIGURE 2. Illustration of the method for the spline of order 2 and the noise model (3) with $\epsilon_i \sim N(0, 0.01^2)$. Left: step 1, estimation of the solution $u_\dagger$ to the differential equation, Right: step 2, estimation of the unknown parameter $a_\dagger$.

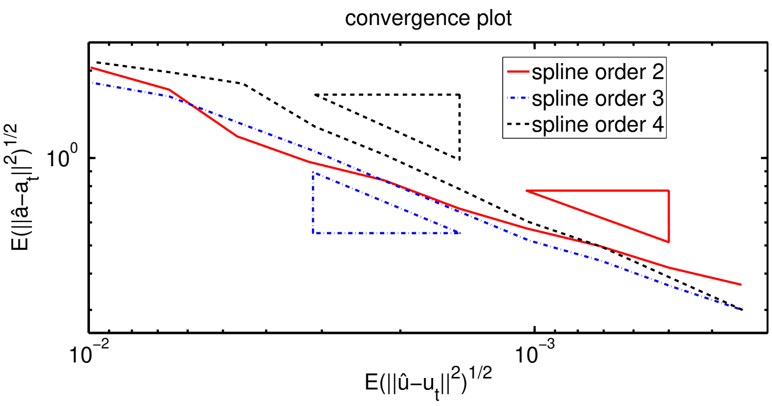

FIGURE 3. Convergence plots for the parameter functions shown in Fig. 1. The triangles indicate the slopes predicted by Theorem 2.1.

The estimators $\widehat{a}$ of the parameter $a_\dagger$ were computed as solutions to the minimizations problem (9) for the initial guess $a_0 = 0$ with the help of the Levenberg-Marquardt algorithm. The linearized least squares minimization problems were solved using the conjugate gradient method. We chose the H^2-norm for the regularization term in the Tikhonov functional so $s = 2$, $p = 2$ and $q \approx 2.5, 3.5$, and 4.5, respectively. This norm was calculated using a finite difference approximation of the Laplacian. We estimated the expected square errors $\sqrt{\mathbf{E}\|\widehat{a} - a_\dagger\|^2}$ and $\sqrt{\mathbf{E}\|\widehat{u} - u_\dagger\|^2}$ for the noise model (3) for different sample size $n \in \{63, 158, 398, 1000, 2512, 6310, 15849, 39811, 100000, 251189\}$ roughly equidistant on a logarithmic scale and $\epsilon_i \sim N(0, 0.01^2)$ by Monte-Carlo simulations running 200 simulations for each sample size n. In Fig. 3 we plot the empirical estimate of $\sqrt{\mathbf{E}\|\widehat{a} - a_\dagger\|^2}$ over the empirical estimate of $\sqrt{\mathbf{E}\|\widehat{u} - u_\dagger\|^2}$ for the three choices of $a_\dagger$ shown in Fig. 1. On both axes a logarithmic scale is used such that according to Theorem 2.1 the plots are expected to be straight lines. The slopes of these straight lines predicted by Theorem 2.1 are indicated by triangles. The empirical slopes of the convergence plots are indeed close to the predicted slopes except for very small sample size. For the example considered here, the asymptotic regime covered by Theorem 2.1 seems to start roughly at $n = 398$.

References

[1] R. A. Adams and J. J. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics*. Elsevier Science, Oxford, 2nd edition, 2003.

[2] N. Bissantz, T. Hohage, and A. Munk. Consistency and rates of convergence of nonlinear Tikhonov regularization with random noise. *Inverse Problems*, 20:1773–1789, 2004.

[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.*, at press.

[4] D. L. Brown and M. G. Low. Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, 24:2384–2398, 1996.

[5] F. Colonius and K. Kunisch. Stability for parameter estimation in two-point boundary value problems. *J. Reine Angew. Math.*, 370:1–29, 1986.

[6] F. Colonius and K. Kunisch. Output least squares stability in elliptic systems. *Appl. Math. Optim.*, 19:33–63, 1989.

[7] H. Egger and A. Neubauer. Preconditioning Landweber iteration in Hilbert scales. *Numer. Math.*, 101:643–662, 2005.

[8] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems.* Kluwer Academic Publishers Group, Dordrecht, 1996.

[9] H. W. Engl, K. Kunisch, and A. Neubauer. Convergence rates for Tikhonov regularisation of nonlinear ill-posed problems. *Inverse Problems*, 5:523–540, 1989.

[10] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order.* Springer, Berlin, 2nd edition, 1983.

[11] A. Goldenshluger and S. V. Pereverzev. On adaptive inverse estimation of linear functionals in Hilbert scales. *Bernoulli*, 9:783–807, 2003.

[12] Q.-n. Jin. Error estimates of some Newton-type methods for solving nonlinear inverse problems in Hilbert scales. *Inverse Problems*, 16(1):187–197, 2000.

[13] J. P. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems.* Springer, New York, 2004.

[14] K. Kunisch. Inherent identifiability of parameters in elliptic differential equations. *J. Math. Anal. Appl.*, 132:453–472, 1988.

[15] J.-M. Loubes and C. Ludena. Penalized estimators for nonlinear inverse problems. *arXiv.org:math*, 1, 2005.

[16] B. A. Mair and F. H. Ruymgaart. Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, 56:1424–1444, 1996.

[17] P. Mathé and S. Pereverzev. Optimal discretization of inverse problems in Hilbert scales. regularization and self-regularization of projection methods. *SIAM J. Numer. Anal.*, 38:1999–2021, 2001.

[18] F. Natterer. Error bounds for Tikhonov regularization in Hilbert scales. *Applicable Anal.*, 18:29–37, 1984.

[19] A. Neubauer. When do Sobolev spaces form a Hilbert scale? *Proc. Amer. Math. Soc.*, 103:557–562, 1988.

[20] A. Neubauer. Tikhonov regularization of nonlinear ill-posed problems in Hilbert scales. *Appl. Anal.*, 46:59–72, 1992.

[21] A. Neubauer. On Landweber iteration for nonlinear ill-posed problems in Hilbert scales. *Numer. Math.*, 85:309–328, 2000.

[22] M. Nussbaum and S. Pereverzev. The degree of ill-posedness in stochastic and deterministic noise models. Technical report, WIAS, Berlin, 1999.

[23] F. O'Sullivan. Convergence characteristics of methods of regularization estimators for nonlinear operator equations. *SIAM J. Numer. Anal.*, 27:1635–1649, 1990.

[24] M. S. Pinsker. Optimal filtering of square-integrable signals in Gaussian white noise. *Probl. Inf. Transm.*, 16:52–68, 1980.
[25] M. Pricop. *Tikhonov regularization in Hilbert scales for nonlinear statistical inverse problems.* PhD thesis, University of Göttingen, 2007.
[26] M. Renardy and R. C. Rogers. *An introduction to partial differential equations.* Springer, New York, 2nd edition, 2004.
[27] O. Scherzer, H. W. Engl, and K. Kunisch. Optimal a-posteriori parameter choice for Tikhonov regularization for solving nonlinear ill-posed problems. *SIAM J. Numer. Anal.*, 30:1796–1838, 1993.
[28] U. Tautenhahn. On a general regularization scheme for nonlinear ill-posed problems. II. Regularization in Hilbert scales. *Inverse Problems*, 14(6):1607–1616, 1998.
[29] U. Tautenhahn and Q. nian Jin. Tikhonov regularization and a posteriori rules for solving nonlinear ill posed problems. *Inverse Problems*, 19:1–21, 2003.
[30] M. Taylor. *Partial Differential Equations: Basic Theory*, volume 1. Springer, New York, 1996.
[31] A. B. Tsybakov. *Introduction à l'estimation non-paramétrique.* Springer, Berlin, 2004.
[32] J. Wloka. *Partial differential equations.* Cambridge University Press, Cambridge, 1987.

E-mail address: hohage@math.uni-goettingen.de
E-mail address: pricop@math.uni-goettingen.de