

# Case studies and a pitfall for nonlinear variational regularization under conditional stability

Daniel Gerth, Bernd Hofmann, and Christopher Hofmann

Faculty of Mathematics, Chemnitz University of Technology, 09107 Chemnitz,  
Germany

`daniel.gerth/bernd.hofmann/christopher.hofmann@mathematik.tu-chemnitz.de`

**Abstract.** Conditional stability estimates are a popular tool for the regularization of ill-posed problems. A drawback in particular under nonlinear operators is that additional regularization is needed for obtaining stable approximate solutions if the validity area of such estimates is not completely known. In this paper we consider Tikhonov regularization under conditional stability estimates for nonlinear ill-posed operator equations in Hilbert scales. We summarize assertions on convergence and convergence rate in three cases describing the relative smoothness of the penalty in the Tikhonov functional and of the exact solution. For oversmoothing penalties, for which the true solution no longer attains a finite value, we present a result with modified assumptions for a priori choices of the regularization parameter yielding convergence rates of optimal order for noisy data. We strongly highlight the local character of the conditional stability estimate and demonstrate that pitfalls may occur through incorrect stability estimates. Then convergence can completely fail and the stabilizing effect of conditional stability may be lost. Comprehensive numerical case studies for some nonlinear examples illustrate such effects.

**Keywords:** Nonlinear inverse problems, conditional stability, Tikhonov regularization, oversmoothing penalties, Hilbert scales, convergence rates

**MSC 2010:** 47J06, 65J20, 47A52

## 1 Introduction

Regularization theory for nonlinear ill-posed inverse problems is always a challenging endeavor. In contrast to linear inverse problems, where the theory is rather coherent and well-developed (see, for example, the monographs [5, 19]), the nonlinear theory is harder to grasp. Numerous assumptions exist in the literature that restrict the nonlinear behavior of the forward operator in such a way that stable approximate solutions exist which converge to the exact solution in the limit of vanishing data noise. It is important to keep in mind that the nonlinearity conditions only hold locally. A main goal of this paper is to show that

this can be a pitfall, as incorrect localization leads to the loss of the stabilizing property. A second objective of the paper is to verify theoretical convergence results in numerical examples, as well as pointing out some open questions. To this end, we focus here on regularization in Hilbert scales. This technique is a common approach for the regularization of ill-posed problems and we refer, for example, to [20, 25, 26]. Going into detail of our paper, we consider in this paper the stable approximate solution of the nonlinear operator equation

$$F(x) = y \quad (1)$$

by variational (Tikhonov-type) regularization. Equation (1) serves as a model for an inverse problem where the nonlinear forward operator  $F : \mathcal{D}(F) \subseteq X \rightarrow Y$  maps between the infinite dimensional real Hilbert spaces  $X$  and  $Y$  with domain  $\mathcal{D}(F)$ . The symbols  $\|\cdot\|_X$ ,  $\|\cdot\|_Y$  and  $\langle \cdot, \cdot \rangle_X$ ,  $\langle \cdot, \cdot \rangle_Y$  designate the norms and inner products of the spaces  $X$  and  $Y$ , respectively. Instead of the exact right-hand side  $y = F(x^\dagger)$ , with the uniquely determined preimage  $x^\dagger \in \mathcal{D}(F)$ , we assume to know a noisy element  $y^\delta \in Y$  satisfying the noise model

$$\|y - y^\delta\|_Y \leq \delta \quad (2)$$

with some noise level  $\delta > 0$ . Based on this data element  $y^\delta \in Y$  we use as approximations to  $x^\dagger$  global minimizers  $x_\alpha^\delta \in \mathcal{D}(F)$  of the extremal problem

$$T_\alpha^\delta(x) := \|F(x) - y^\delta\|_Y^2 + \alpha \|Bx\|_X^2 \rightarrow \min, \quad \text{subject to } x \in \mathcal{D}(F). \quad (3)$$

Here,  $B : \mathcal{D}(B) \subset X \rightarrow X$  is a densely defined, unbounded, linear, and self-adjoint operator which is strictly positive such that  $\|Bx\|_X \geq c_B \|x\|_X$  holds for all  $x \in \mathcal{D}(B)$ . Such operators  $B$  generate a Hilbert scale  $\{X_\nu\}_{\nu \in \mathbb{R}}$ , where  $X_\nu = \mathcal{D}(B^\nu)$  coincides with the range  $\mathcal{R}(B^{-\nu})$  of the operator  $B^{-\nu}$ . In particular  $X_0 = X$ , and we set  $\|x\|_\nu := \|B^\nu x\|_X$  for the norm of the Hilbert scale element  $x \in X_\nu$ . With this, the specific Tikhonov functional  $T_\alpha^\delta : X \rightarrow [0, \infty]$  in (3) is the weighted sum of the quadratic misfit functional  $\|F(\cdot) - y^\delta\|_Y^2$  and the Hilbert-scale penalty functional  $\|B \cdot\|_X^2 = \|\cdot\|_1^2$ , where the regularization parameter  $\alpha > 0$  acts as weight factor. Note that no generality is lost by considering only the penalty in the 1-norm  $\|\cdot\|_1$ , since one can always rescale the operator  $B$  to obtain  $\|Bx\| = \|(B^{\frac{1}{p}})^p x\| = \|\tilde{B}^p x\|$  for  $p > 0$ , i.e., one obtains a penalty of arbitrary index  $p$  in the Hilbert scale generated by the operator  $\tilde{B} := B^{\frac{1}{p}}$ . Finally, we mention that for  $x \in \mathcal{D}(F)$  we set  $T_\alpha^\delta(x) := +\infty$  if  $x \notin \mathcal{D}(B)$ , and that the Tikhonov functional attains a well-defined value  $0 \leq T_\alpha^\delta(x) < +\infty$  if  $x \in \mathcal{D} := \mathcal{D}(F) \cap \mathcal{D}(B) \neq \emptyset$ .

A typical phenomenon of the nonlinear equation (1) as a model for an inverse problem is *local ill-posedness* at the solution point  $x^\dagger \in \mathcal{D}(F)$  (cf. [16, Def. 2] or [15, Def. 3]), which means that inequalities of the form

$$\|x - x^\dagger\|_X \leq K \varphi(\|F(x) - F(x^\dagger)\|_Y) \quad \text{for all } x \in \mathcal{B}_r^X(x^\dagger) \cap \mathcal{D}(F) \quad (4)$$

cannot hold for any positive constants  $K$ ,  $r$  and any index function  $\varphi$ .<sup>1</sup> However, the inverse problem literature offers numerous examples, where the left-hand term  $\|x - x^\dagger\|_X$  in (4) is replaced with a weaker norm  $\|x - x^\dagger\|_{-a}$  ( $a > 0$ ) and a corresponding conditional stability estimate takes place. In the sequel, we restrict our considerations to the concave index functions  $\varphi(t) = t^\gamma$  of Hölder-type with exponents  $0 < \gamma \leq 1$  and hence to *conditional stability estimates* of the form

$$\|x - x^\dagger\|_{-a} \leq K \|F(x) - F(x^\dagger)\|_Y^\gamma \quad \text{for all } x \in Q \cap \mathcal{D}(F) \quad (5)$$

with some index  $a > 0$ , which can be interpreted as *degree of ill-posedness* of  $F$  at  $x^\dagger$ , a suitable subset  $Q$  in  $X$  which acts as the aforementioned localization of the nonlinearity condition, and a constant  $K > 0$  that may depend on  $Q$ .

Let us consider the situation that  $x^\dagger \in Q$  and  $Q$  is known. Then one may employ a least squares iteration process of minimizing the norm square

$$\|F(x) - y^\delta\|_Y^2 \rightarrow \min, \quad \text{subject to } x \in Q \cap \mathcal{D}(F). \quad (6)$$

The minimizers  $x_{ls}$  of (6) satisfy  $\|F(x_{ls}) - y^\delta\| \leq \delta$  by definition and due to  $x^\dagger \in Q$ . Hence we have convergence  $\|x_{ls}^\delta - x^\dagger\|_{-a} \rightarrow 0$  as  $\delta \rightarrow 0$  of these least squares-type solutions to  $x^\dagger$  in the norm of the space  $X_{-a}$  which is weaker than the one in  $X$ .

To achieve convergence and even convergence rates in the norm of  $X$ , additional smoothness  $x^\dagger \in X_p$  for some  $p > 0$  is needed. If the approximate solutions  $x_{ls}^\delta \in Q \cap \mathcal{D}(F)$  also possess such smoothness with  $\|x_{ls}^\delta\|_p$  uniformly bounded for all  $0 < \delta \leq \bar{\delta}$ , then, with  $-a < t \leq p$  the interpolation inequality in Hilbert scales (see [18]) applies in the form

$$\|x\|_t \leq \|x\|_{\frac{p-t}{p+a}}^{\frac{p-t}{p+a}} \|x\|_p^{\frac{t+a}{p+a}} \quad (7)$$

for all  $x \in X_p$ . Hence we derive from (5) and (7) with  $t = 0$  and by the triangle inequality that

$$\|x_{ls}^\delta - x^\dagger\|_X \leq \bar{K} \delta^{\frac{\gamma p}{p+a}}$$

for sufficiently small  $\delta > 0$  and some constant  $\bar{K}$ . A way to ensure the property that the approximate solutions belong to  $X_p \cap Q \cap \mathcal{D}(F)$  is to use regularized solutions which minimize the Tikhonov functional  $\|F(x) - y^\delta\|_Y^2 + \alpha \|B^s x\|_X^2$ , subject to  $x \in Q \cap \mathcal{D}(F)$ , where  $s \geq p$  is required. Hence, Tikhonov-type regularization is here an auxiliary tool which complements the conditional stability estimate (5) in order to obtain stable approximate solutions measured in the norm of  $X$ .

On the other hand, we have to take into account the frequently occurring situation that the set  $Q$  in (5) is not or not completely known and a minimization process according to (6) is impossible, because of a not completely known

<sup>1</sup> Throughout,  $\mathcal{B}_r^H(\bar{x})$  denotes a closed ball in the Hilbert space  $H$  around  $\bar{x} \in H$  with radius  $r > 0$ . Furthermore, we call a function  $\varphi: [0, \infty) \rightarrow [0, \infty)$  index function if it is continuous, strictly increasing and satisfies the boundary condition  $\varphi(0) = 0$ .

set of constraints for the optimization problem. Nevertheless, a combination of the conditional stability estimate (5) with variational regularization of the form (3) can be successful. For a systematic treatment of convergence results in the context of regularization theory we will distinguish the following cases relating the smoothness of the solution  $x^\dagger$  and of the approximate solutions  $x_\alpha^\delta$  implied by the functional (3):

**Case distinction.**

- (a) *Classical regularization:*  $x^\dagger \in X_p$  for  $p > 1$ , which means that  $\|Bx^\dagger\|_X^2 < +\infty$  and there is some source element  $w \in X_\varepsilon$  ( $\varepsilon > 0$ ) such that  $x^\dagger = B^{-1}w$ ;
- (b) *Matching smoothness:*  $x^\dagger \in X_1$ , i.e.  $\|Bx^\dagger\|_X < \infty$ , but  $x^\dagger \notin X_{1+\varepsilon}$  for all  $\varepsilon > 0$ .
- (c) *Oversmoothing regularization:*  $x^\dagger \in X_p$  for some  $0 < p < 1$ , but  $x^\dagger \notin X_1$ , i.e.  $\|Bx^\dagger\|_X = +\infty$ .

The goal of this paper is to discuss the different opportunities and limitations for convergence and rates of regularized solutions  $x_\alpha^\delta$  in the situations (a), (b), and (c), respectively. It is organized as follows: Section 2 recalls assertions on convergence of regularized solutions in cases (a) and (b). Moreover, usual technical assumptions on forward operator, its domain and the exact solution are listed. In Section 3, Hölder rate results under conditional stability estimates are summarized for the cases of classical regularization and matching smoothness. The rate result of Proposition 4 for the oversmoothing case (c) is of specific interest. It requires two-sided inequalities as conditional stability estimates, whereas in cases (a) and (b) only one-sided inequalities are needed. Three inverse model problems of ill-posed nonlinear equations covering all cases (a), (b), and (c) are outlined in Section 4, for which numerical case studies are presented in Section 5. The proof of Proposition 4 is given in the appendix.

## 2 Convergence

In this section we collect properties of the regularized solutions  $x_\alpha^\delta$  obtained as solutions of the optimization problem (3) for the cases (a), (b), and (c) in different ways. Throughout this paper we suppose that the following assumption concerning the nonlinear forward operator  $F$  and the solvability of the operator equation (1) holds true.

**Assumption 1.** *The operator  $F : D(F) \subseteq X \rightarrow Y$  is weak-to-weak sequentially continuous and its domain  $D(F)$  is a convex and closed subset of  $X$ . For the right-hand side  $y = F(x^\dagger) \in Y$  under consideration let  $x^\dagger \in \mathcal{D}(F)$  be the uniquely determined solution to the operator equation (1).*

Under the setting introduced in Section 1, the penalty  $\|Bx\|_X^2$  as part of the Tikhonov functional  $T_\alpha^\delta$  in (3) is a non-negative, convex, and sequentially lower semi-continuous functional. Moreover, this functional is *stabilizing* in the sense that all its sublevel sets are weakly sequentially compact in  $X$ . Taking also into account Assumption 1, the Assumptions 3.11 and 3.22 of [24] are satisfied and

the assertions from [24, Section 4.1.1] apply, which ensure *existence* and *stability* of the regularized solutions  $x_\alpha^\delta$  in our present Hilbert scale setting, consistent for all three cases (a), (b), and (c).

We emphasize at this point that we always have  $x_\alpha^\delta \in X_1$  by definition of the minimizers in (3), but only in the cases (a) and (b) one can take profit of the inequality

$$T_\alpha^\delta(x_\alpha^\delta) \leq T_\alpha^\delta(x^\dagger), \tag{8}$$

which implies for all  $\alpha > 0$  that

$$\|x_\alpha^\delta\|_1 \leq \sqrt{\|x^\dagger\|_1^2 + \frac{\delta^2}{\alpha}}. \tag{9}$$

In the case (c), however, due to  $x^\dagger \notin X_1$  and hence  $\|x^\dagger\|_1 = +\infty$  we have no such uniform bounds of  $\|x_\alpha^\delta\|_1$  from above. On the contrary, in [7] it was shown that  $\|x_\alpha^\delta\|_1 \rightarrow \infty$  as  $\delta \rightarrow 0$  is necessary even for weak convergence of the regularizers  $x_\alpha^\delta$  to  $x^\dagger$ .

In order to obtain *convergence* of the regularized solutions  $x_\alpha^\delta$  to  $x^\dagger$  as  $\delta \rightarrow 0$ , the interplay of the noise level and the choice of the regularization parameter  $\alpha > 0$ , which we choose either a priori  $\alpha = \alpha(\delta)$  or a posteriori  $\alpha = \alpha(\delta, y^\delta)$ , must be appropriate. In the literature, this interplay is typically controlled by the limit conditions

$$\alpha \rightarrow 0 \quad \text{and} \quad \frac{\delta^2}{\alpha} \rightarrow 0 \quad \text{as} \quad \delta \rightarrow 0. \tag{10}$$

In our case (a) this is a sufficient description.

**Proposition 1.** *Let the regularization parameter  $\alpha > 0$  fulfill the conditions (10). Then we have under Assumption 1 and for case (a), i.e. for  $1 < p < \infty$ , by setting  $\alpha_n = \alpha(\delta_n)$  or  $\alpha_n = \alpha(\delta_n, y^{\delta_n})$ ,  $x_n = x_{\alpha_n}^{\delta_n}$ , that for  $\delta_n \rightarrow 0$  as  $n \rightarrow \infty$*

$$\lim_{n \rightarrow \infty} \|x_n\|_1 = \|x^\dagger\|_1,$$

and

$$\lim_{n \rightarrow \infty} \|x_n - x^\dagger\|_\nu = 0 \quad \text{for all} \quad 0 \leq \nu \leq 1.$$

*Proof.* The proof follows along the lines of Theorem 4.3 and Corollary 4.6 from [24].  $\square$

As we will see in Proposition 2 in the next section, the optimal parameter choice fulfills the conditions (10) in case (a). In case (b), where the smoothness of  $x^\dagger$  coincides with the smoothness of the regularization, i.e.,  $p = 1$ , the matter becomes unclear. On one hand, it is easily seen that Proposition 1 holds in the exact same way for case (b), which is a consequence of (9) holding in both cases. Hence, we have the following corollary:

**Corollary 1.** *Under the assumptions of Proposition 1, in particular for the cases (a) and (b) and for a regularization parameter choice satisfying (10), we have that the regularized solutions  $x_\alpha^\delta$  belong to the ball  $\mathcal{B}_r^{X_\nu}(x^\dagger)$  for prescribed values  $r > 0$  and  $0 \leq \nu \leq 1$  whenever  $\delta > 0$  is sufficiently small.*

The surprising difference between the cases (a) and (b) on the other hand, is that the optimal choice of the regularization parameter for (b) (we show in Proposition 3 below that  $\alpha \sim \delta^2$  yields the optimal convergence rate) violates the second condition in (10). Since obviously a convergence rate implies norm convergence, this means that the condition  $\delta^2/\alpha \rightarrow 0$  in (10) is not necessary but sufficient for convergence, at least in case (b).

In case (c) with oversmoothing penalty, the inequality (8) and consequently (9) are missing. Results of Proposition 1 and Corollary 1 in general do not apply in that case. One cannot even show weak convergence  $x_n \rightharpoonup x^\dagger$  in  $X$ , and regularized solutions  $x_\alpha^\delta$  need not belong to a ball  $\mathcal{B}_r^X(x^\dagger)$  with small radius  $r > 0$  if  $\delta > 0$  is sufficiently small. As will be shown in Proposition 4 of Section 3 (see also [13, 14]), convergence rates can be proven under stronger conditions also for (c), where we have some  $0 < p < 1$  such that  $x^\dagger \in X_p$ . The key to these results was the appropriate choice of  $\alpha$  either by an a priori or a posteriori parameter choice. In particular,  $\frac{\delta^2}{\alpha} \rightarrow \infty$  as  $\delta \rightarrow 0$ , which violates (10), is typical there. The interplay of  $\alpha$  and  $\delta$  will be in the focus of our numerical case studies in Section 5 below.

### 3 Convergence rate results

In this section, we are going to discuss convergence rate results for cases (a) and (b) on one hand, but also (c) on the other hand. In addition to Assumption 1 some versions of conditional stability estimates have to be imposed which, in combination with the smoothness assumptions  $x^\dagger \in X_p$ , are essentially hidden forms of source conditions for the solution  $x^\dagger$ .

In Assumption 2 we first consider the situation for the setting  $Q := \mathcal{B}_\rho^{X_1}(0)$ . This model setting was comprehensively discussed and illustrated by examples of associated nonlinear inverse problems in the papers [3, 4, 13, 17]. Here we have evidently  $x^\dagger \in Q$  for the cases (a) and (b) whenever  $\|x^\dagger\|_1 \leq \rho$ .

**Assumption 2.** *Let for fixed  $a > 0$  and  $0 < \gamma \leq 1$  the conditional stability estimates*

$$\|x - x^\dagger\|_{-a} \leq K(\rho) \|F(x) - F(x^\dagger)\|_Y^\gamma \quad \text{for all } x \in \mathcal{B}_\rho^{X_1}(0) \cap \mathcal{D}(F) \quad (11)$$

*hold, where constants  $K(\rho) > 0$  are supposed to exist for all radii  $\rho > 0$ .*

Then the following proposition, which is a direct consequence of [4, Theorem 2.1] when adapting the corresponding proof, yields an order optimal convergence rate in case (a).

**Proposition 2.** *Under Assumptions 1 and 2 and for  $x^\dagger \in X_p$  with  $1 < p \leq a+2$  we have the rate of convergence of regularized solutions  $x_\alpha^\delta \in \mathcal{D}(F) \cap \mathcal{D}(B)$  to the solution  $x^\dagger \in \mathcal{D}(F) \cap \mathcal{D}(B)$  as*

$$\|x_\alpha^\delta - x^\dagger\|_X = \mathcal{O}\left(\delta^{\frac{\gamma p}{p+a}}\right) \quad \text{as } \delta \rightarrow 0, \quad (12)$$

*provided that the regularization parameter  $\alpha = \alpha(\delta)$  is chosen a priori as*

$$\alpha(\delta) \sim \delta^{2-2\gamma\frac{p-1}{p+a}}. \quad (13)$$

We easily see that the convergence results of Proposition 1 apply here for  $p > 1$  and that in particular (13) implies (10). The additional smoothness of  $x^\dagger$ , which is always required to obtain convergence rates in regularization of ill-posed problems appears in Hilbert scales in form  $x^\dagger = B^{-p}v$  with some source element  $v \in X$ .

*Remark 1.* We mention that along the lines of [4, Theorem 2.2] the rate (12) can also be shown under the assumptions of Proposition 2 when the regularization parameter  $\alpha = \alpha(\delta, y^\delta)$  is chosen a posteriori by a sequential discrepancy principle.

The modified version of the rate result for case (b) is as follows:

**Proposition 3.** *Under the Assumptions 1 and 2 and for  $x^\dagger \in X_1$  we have the rate of convergence of regularized solutions  $x_\alpha^\delta \in \mathcal{D}(F) \cap \mathcal{D}(B)$  to the solution  $x^\dagger \in \mathcal{D}(F) \cap \mathcal{D}(B)$  as*

$$\|x_\alpha^\delta - x^\dagger\|_X = \mathcal{O}\left(\delta^{\frac{\gamma}{1+a}}\right) \quad \text{as } \delta \rightarrow 0, \quad (14)$$

*if the regularization parameter  $\alpha = \alpha(\delta)$  is chosen a priori as*

$$\alpha(\delta) \sim \delta^2. \quad (15)$$

*Proof.* By the standard technique of variational regularization under conditional stability estimates (cf. [4, Proof of Theorem 1.1] or [24, Section 4.2.5]) we obtain for the choice (15) of the regularization parameter and by using the conditional stability estimate (11) the inequality

$$\|x_\alpha^\delta - x^\dagger\|_{-a} \leq C\delta^\gamma, \quad (16)$$

where the constant  $C > 0$  via  $\rho$  and  $K(\rho)$  depends on  $\|x^\dagger\|_1$  and on upper and lower bounds of  $\delta^2/\alpha$ . Combining this with the interpolation inequality (7), taking  $t = 0$  and  $s = 1$ , and applying the triangle inequality provides us with the estimate

$$\|x_\alpha^\delta - x^\dagger\|_X \leq C(\|x_\alpha^\delta\|_1 + \|x^\dagger\|_1)^{\frac{a}{1+a}} \delta^{\frac{\gamma}{1+a}}.$$

Due to (9) the norm  $\|x_\alpha^\delta\|_1$  is uniformly bounded by a finite constant for  $\alpha(\delta)$  from (15). This yields the rate (14) and completes the proof. Finally, we should note that the inequality (16) can only be established, because constants  $K(\rho) > 0$  in (11) exist for arbitrarily large  $\rho > 0$ .  $\square$

In the borderline case (b) we have also a borderline a priori choice of the regularization parameter which contradicts the second limit condition in (10) such that the quotient  $\frac{\delta^2}{\alpha}$  is uniformly bounded below by a positive constant and above by a finite constant.

In Assumption 3 we consider alternatively the situation that  $Q := \mathcal{B}_r^X(x^\dagger)$ . This model, which is illustrated by Example 1 in Section 4 below, is typical for conditional stability estimates that arise from nonlinearity conditions imposed on the forward operator  $F$  in a neighbourhood of the solution  $x^\dagger$ . In this context, the radius  $r > 0$  which restricts the validity area of stability estimates can be rather small. In all cases of the Case distinction we have here  $x^\dagger \in Q \cap \mathcal{D}(F)$ , but only for (a) and (b) also  $x^\dagger \in \mathcal{D}(F) \cap \mathcal{D}(B)$ .

**Assumption 3.** *Let for fixed  $a > 0$  and  $0 < \gamma \leq 1$  the conditional stability estimate*

$$\|x - x^\dagger\|_{-a} \leq K(r) \|F(x) - F(x^\dagger)\|_Y^\gamma \quad \text{for all } x \in \mathcal{B}_r^X(x^\dagger) \cap \mathcal{D}(F) \quad (17)$$

hold, where the constant  $K(r) > 0$  depends on the largest admissible radius  $r > 0$ .

**Corollary 2.** *The assertion of Proposition 2 remains true if Assumption 2 is replaced with Assumption 3.*

*Proof.* To see the validity of Proposition 2 under Assumption 3 in case (a) of the Case distinction, where the regularization parameter choice satisfies (10), it is enough to take the assertion of Corollary 1 into account. This assertion implies that for sufficiently small  $\delta > 0$  the regularized solutions  $x_\alpha^\delta$  belong to the ball  $\mathcal{B}_r^X(x^\dagger)$  for prescribed  $r > 0$ . Then the conditional stability estimate (17) applies and yields the convergence rate (12) along the lines of the proof of [4, Theorem 2.1].  $\square$

In case (b), however, for the choice (15) of Proposition 3 the condition (10) fails and even if  $\delta > 0$  is sufficiently small, it cannot be shown that  $x_\alpha^\delta \in \mathcal{B}_r^X(x^\dagger)$  for prescribed  $r > 0$ . Consequently, the conditional stability estimate (17) need not hold for the regularized solutions  $x = x_\alpha^\delta$  and the rate assertion (14) of Proposition 3 is only valid under Assumption 3 if constants  $K(r) > 0$  in (17) exist for arbitrarily large  $r > 0$ . This is, however, the case in the exponential growth model of Example 1 below.

Now we turn to the cases with oversmoothing penalty, where  $x^\dagger \notin X_1$  and restrict ourselves to  $\gamma = 1$  in the conditional stability estimates. As is well-known since the paper by Natterer [21], convergence rates in this case require lower and upper estimates of  $\|F(x) - F(x^\dagger)\|_Y$  by multiples of the term  $\|x - x^\dagger\|_{-a}$ . We start with a corresponding analytical result. The goal of the case studies in Section 5 below is to gain further insight into the behavior of regularized solutions in case (c) for a priori and a posteriori choices of the regularization parameter.

**Assumption 4.** *Let  $a > 0$ . Moreover, let  $x^\dagger$  be an interior point of  $\mathcal{D}(F)$  such that for the radius  $r > 0$  we have  $\mathcal{B}_r^X(x^\dagger) \subset \mathcal{D}(F)$  and the two estimates*

$$\underline{K} \|x - x^\dagger\|_{-a} \leq \|F(x) - F(x^\dagger)\|_Y \quad \text{for all } x \in \mathcal{D}(F) \cap \mathcal{D}(B) = \mathcal{D}(F) \cap X_1 \quad (18)$$



and

$$\|F(x) - F(x^\dagger)\|_Y \leq \overline{K} \|x - x^\dagger\|_{-a} \quad \text{for all } x \in \mathcal{B}_r^X(x^\dagger) \cap X_1 \quad (19)$$

hold true, where  $0 < \underline{K} \leq \overline{K} < \infty$  are constants.

**Proposition 4.** *Let  $x^\dagger \in X_p$  for some  $0 < p < 1$ , but  $x^\dagger \notin X_1$ . Under the Assumptions 1 and 4 we then have the rate of convergence of regularized solutions to the exact solution as*

$$\|x_{\alpha_*}^\delta - x^\dagger\|_X = \mathcal{O}\left(\delta^{\frac{p}{p+a}}\right) \quad \text{as } \delta \rightarrow 0, \quad (20)$$

if the regularization parameter is chosen a priori as

$$\alpha_* = \alpha(\delta) = \delta^{2 - \frac{2(p-1)}{p+a}}. \quad (21)$$

The proof of Proposition 4 is given in the appendix along the lines of [14, Theorem 1], where we set for simplicity  $\bar{x} = 0$ . Note that Theorem 1 in [14] refers to a simplified version of the pair of estimates (18) and (19), which are *ibid* both assumed to hold for all  $x \in \mathcal{D}(F)$ . As the proof in the appendix shows, the upper estimate (19) is only exploited by auxiliary elements  $x_\alpha$ , which belong to  $\mathcal{B}_r^X(x^\dagger) \cap X_1$  for sufficiently small  $\alpha > 0$ . On the other hand, there are no arguments for restricting the noisy regularized solutions  $x_\alpha^\delta$  to small balls. Consequently, the lower estimate (18) needs to hold for all elements in  $\mathcal{D}(F) \cap X_1$ . This is an essential drawback for the application of Proposition 4 to practical problems. An analogue of Proposition 4 for the discrepancy principle as parameter choice rule can be formulated and proven along the lines of the paper [13].

As already mentioned in Section 2, we stress again that, despite the assertion of Proposition 4, norm convergence of regularized solutions cannot be shown in general for case (c), not even weak convergence in  $X$  can be established. Evidently the parameter choice (21) violates (10) since we have

$$\alpha(\delta) \rightarrow 0 \quad \text{and} \quad \frac{\delta^2}{\alpha(\delta)} = \delta^{\frac{2(p-1)}{p+a}} \rightarrow \infty \quad \text{as } \delta \rightarrow 0.$$

It appears that

$$\alpha(\delta) \rightarrow 0 \quad \text{and} \quad \frac{\delta^2}{\alpha(\delta)} \rightarrow \infty \quad \text{as } \delta \rightarrow 0$$

tends to be the typical situation in the oversmoothing case (c), at least for regularization parameters yielding optimal convergence rates. Numerical case studies below support this conjecture. A similar behavior of the regularization parameters was noted for oversmoothing  $\ell^1$ -regularization [7].

To conclude and summarize this section, we stress that in all cases of the Case distinction, we have under the appropriate conditional stability assumption (to show the similarities between the cases, we fix  $\gamma = 1$  for (a), (b), and (c) for the next assertion) and for  $x^\dagger \in X_p$  for some  $p > 0$  the convergence rate

$$\|x_\alpha^\delta - x^\dagger\|_X = \mathcal{O}\left(\delta^{\frac{p}{p+a}}\right) \quad \text{as } \delta \rightarrow 0 \quad (22)$$

under both the discrepancy principle and the a priori parameter choice

$$\alpha(\delta) = \delta^{2-2\frac{p-1}{p+a}} = \delta^{\frac{2(a+1)}{a+p}}. \quad (23)$$

Hence, we obtain the same parameter choice and the same convergence rate as in the case of a linear operator. Namely in [21], (22) and (23) were obtained for a linear operator  $A : X \rightarrow Y$  under a two-sided inequality

$$\underline{K}\|x\|_{-a} \leq \|Ax\|_Y \leq \overline{K}\|x\|_{-a} \quad \text{for all } x \in X,$$

in analogy to the estimates from Assumption 4.

## 4 Examples

In the following, we introduce two nonlinear inverse problems of type (1), for which we will investigate the analytic results from the previous section numerically. Before doing so, we will introduce two similar, but different Hilbert scales used as penalty in the minimization problem (3) and as measure of the solution smoothness. On one hand, we consider the standard Sobolev-scale  $H^p[0, 1]$ . For integer values of  $p \geq 0$ , these function spaces consist of functions whose  $p$ -th derivative is still in  $L^2(0, 1)$ . For real parameters of  $p > 0$ , the spaces can be defined by an interpolation argument [1]. Using Fourier-analysis, one can define a norm in  $H^p[0, 1]$  via

$$\|x\|_{H^p[0,1]}^2 := \int_{\mathbb{R}} (1 + |\xi|^2)^p |\hat{x}(\xi)|^2 d\xi, \quad (24)$$

where  $\hat{x}$  is the Fourier-transform of  $x$ . Then  $x \in H^p[0, 1]$  iff  $\|x\|_{H^p[0,1]} < \infty$ . The Sobolev scale for  $p \geq 0$  does not constitute a Hilbert scale in the strict sense, but for each  $0 < p^* < \infty$  there is an operator  $B : L^2(0, 1) \rightarrow L^2(0, 1)$  such that  $\{X_p\}_{0 \leq p \leq p^*}$  is a Hilbert scale [22]. This is not an issue in numerical experiments. Note that the norm (24) is easy to implement, in particular it allows a precise gauging of the solution smoothness.

The reason why the Sobolev scale does not form a Hilbert scale for arbitrary values of  $p$  lies in the boundary values. In order to generate a full Hilbert scale  $\{X_\tau\}_{\tau \in \mathbb{R}}$ , we exploit the simple integration operator

$$[Jh](t) := \int_0^t h(\tau) d\tau \quad (0 \leq t \leq 1) \quad (25)$$

of Volterra-type mapping in  $X = Y = L^2(0, 1)$  and set

$$B := (J^*J)^{-1/2}. \quad (26)$$

By considering the Riemann-Liouville fractional integral operator  $J^p$  and its adjoint  $(J^*)^p = (J^p)^*$  for  $0 < p \leq 1$  we have that

$$X_p = \mathcal{D}(B^p) = \mathcal{R}((J^*J)^{p/2}) = \mathcal{R}((J^*)^p),$$

cf. [9, 10, 23], and hence by [10, Lemma 8]

$$X_p = \begin{cases} H^p[0, 1] & \text{for } 0 < p < \frac{1}{2} \\ \{x \in H^{\frac{1}{2}}[0, 1] : \int_0^1 \frac{|x(t)|^2}{1-t} dt < \infty\} & \text{for } p = \frac{1}{2} \\ \{x \in H^p[0, 1] : x(1) = 0\} & \text{for } \frac{1}{2} < p \leq 1 \end{cases}, \quad (27)$$

where the fractional Sobolev spaces  $H^p[0, 1]$  occur. One can also show that

$$X_p = \{x \in H^p[0, 1] : x(1) = 0\} \quad \text{for } 1 < p < \frac{3}{2}. \quad (28)$$

On the other hand, it is well-known that  $X_2 \subset X_p \subset X_1$  for  $1 < p < 2$  and that  $X_2$  is characterized by

$$X_2 = \{x \in H^2[0, 1] : x'(0) = 0, x(1) = 0\} \quad (29)$$

in an explicit manner, see for example [19, Beispiel 2.1.5]. We omit discussing higher smoothness spaces since we will not consider those in our examples. In the following we present examples that show and illustrate the occurrence of the cases (a), (b) and (c) of the Case distinction in Section 1. Note that for  $0 < p < \frac{1}{2}$  the Sobolev-scale  $H^p[0, 1]$  and the Hilbert scale  $\{X_p\}_{p>0}$  induced by  $J$  coincide.

**Model problem 1** (Exponential growth model). The following exponential growth model has been previously discussed in the literature, and we refer for more details and properties to [11, Section 3.1] and [12]. To identify the time dependent growth rate  $x(t)$  ( $0 \leq t \leq T$ ) of a population we use observations  $y(t)$  ( $0 \leq t \leq T$ ) of the time-dependent size of the population with initial size  $y(0) = y_0 > 0$ , where the O.D.E. initial value problem

$$y'(t) = x(t)y(t) \quad (0 < t \leq T), \quad y(0) = y_0,$$

is assumed to hold. For simplicity let in the sequel  $T := 1$  and consider the space setting  $X = Y := L^2(0, 1)$ . Then we simply derive the nonlinear forward operator  $F : x \mapsto y$  mapping in the real Hilbert space  $L^2(0, 1)$  as

$$[F(x)](t) = y_0 \exp \left( \int_0^t x(\tau) d\tau \right) \quad (0 \leq t \leq 1), \quad (30)$$

with full domain  $\mathcal{D}(F) = L^2(0, 1)$  and with the Fréchet derivative

$$[F'(x)h](t) = [F(x)](t) \int_0^t h(\tau) d\tau \quad (0 \leq t \leq 1, h \in X).$$

It can be shown that there is some constant  $\hat{K} > 0$  such that for all  $x \in X$  the inequality

$$\|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\|_Y \leq \hat{K} \|F(x) - F(x^\dagger)\|_Y \|x - x^\dagger\|_X \quad (31)$$

is valid. By applying the triangle inequality to (31) we obtain the estimate

$$\|F'(x^\dagger)(x-x^\dagger)\|_Y \leq (\hat{K} \|x-x^\dagger\|_X + 1) \|F(x) - F(x^\dagger)\|_Y \leq \check{K}(r) \|F(x) - F(x^\dagger)\|_Y \quad (32)$$

for all  $x \in \mathcal{B}_r^X(x^\dagger)$ , where the constant  $\check{K}(r) > 0$  attains the form  $\check{K}(r) := r\hat{K} + 1$  for arbitrary  $r > 0$ .

Using the Hilbert scale generated by the operator  $J$  from (25), taking into account that  $\|Jh\|_Y = \|(J^*J)^{1/2}h\|_X = \|B^{-1}h\|_X = \|h\|_{-1}$  for all  $h \in X$ , and that there is some  $0 < \underline{c} < \infty$  such that  $\underline{c} \leq [F(x^\dagger)](t)$  ( $0 \leq t \leq 1$ ) for the multiplier function in  $F'(x^\dagger)$ , there is a constant  $0 < c_0 < \infty$  satisfying

$$c_0 \|x - x^\dagger\|_{-1} = c_0 \|J(x - x^\dagger)\|_Y \leq \|F'(x^\dagger)(x - x^\dagger)\|_Y \quad \text{for all } x \in X.$$

This implies by formula (32) the estimate

$$\|x - x^\dagger\|_{-1} \leq K(r) \|F(x) - F(x^\dagger)\|_Y \quad \text{for all } x \in \mathcal{B}_r^X(x^\dagger). \quad (33)$$

This estimate is of the form (17) with  $a := 1$  and  $K(r) := \frac{r\hat{K}+1}{c_0}$ . But it is specific for this example that there exist constants  $K(r) > 0$  for arbitrarily large radii  $r > 0$  such that (33) is valid.

The case (a) of the Case distinction in Section 1 occurs due to formula (28) if the solution is sufficiently smooth, i.e.  $x^\dagger \in H^p[0, 1]$  for some  $p > 1$  and, for the Hilbert scale induced by  $J$ , it fulfills the necessary boundary conditions. Case (b) will be the subject of Model problem 3 below. The oversmoothing case (c) of the Case distinction occurs either if the solution is insufficiently smooth, i.e.  $x^\dagger \in H^p[0, 1]$  for  $0 < p < 1$ , or in case of the Hilbert scale induced by  $J$ , one might have  $x^\dagger \in H^p[0, 1]$  for  $p \geq 1$  but the boundary condition  $x^\dagger(1) = 0$  fails. Due to formula (27) we then have  $x^\dagger \in X_p$  for all  $p < 1/2$ , but  $x^\dagger \notin X_p$  for all  $p > 1/2$  and consequently also  $x^\dagger \notin X_1$ . This is, for example, the case for the constant function  $x^\dagger(t) = 1$  ( $0 \leq t \leq 1$ ). We complete this example with the remark that due to formula (29) a function  $x^\dagger \in H^2[0, 1]$ , like the function  $x^\dagger(t) = -(t-0.5)^2 + 0.25$  used in the case studies below and satisfying  $x^\dagger(1) = 0$ , does not belong to  $X_2$  whenever its first derivative at  $t = 0$  does not vanish.

**Model problem 2 (Autoconvolution).** As a second problem, we consider under the same space setting  $X = Y := L^2(0, 1)$  the autoconvolution operator on the unit interval defined as

$$[F(x)](s) = \int_0^s x(s-t)x(t)dt \quad (0 \leq s \leq 1), \quad (34)$$

with full domain  $\mathcal{D}(F) = L^2(0, 1)$ . This operator and the associated nonlinear operator equation (1) with applications in statistics and physics have been discussed early in the literature of inverse problems (cf. [8]). Due to extensions in laser optics, the deautoconvolution problem was comprehensively revisited recently (see, e.g., [2] and [6]). Even though  $F$  from (34) is a non-compact operator, we have for all  $x \in X$  a compact Fréchet derivative

$$[F'(x)h](s) = 2 \int_0^s x(s-t)h(t)dt \quad (0 \leq s \leq 1, \quad h \in X).$$

Taking the Hilbert scale  $\{X_\tau\}_{\tau \in \mathbb{R}}$  based on the operator  $B$  from (26) and the integral operator  $J$  from (25), we see for the specific solution

$$x^\dagger(t) = 1 \quad (0 \leq t \leq 1) \quad (35)$$

that

$$\|F'(x^\dagger)h\|_Y = 2\|Jh\|_Y = 2\|B^{-1}h\|_X = \|h\|_{-1} \quad \text{for all } h \in X.$$

Unfortunately no estimate of the form (32) is available, because such estimates with  $F$ -differences on the right-hand side are not known for the autoconvolution operator. However, as a condition characterizing the nonlinearity of  $F$  the inequality

$$\|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\|_Y = \|F(x - x^\dagger)\|_Y \leq \|x - x^\dagger\|_X^2 \quad \text{for all } x \in X$$

is valid. Thus we have for all  $x \in X$  and  $x^\dagger$  from (35), by using the triangle inequality,

$$\begin{aligned} \|x - x^\dagger\|_{-1} &\leq \frac{1}{2}\|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\|_Y + \frac{1}{2}\|F(x) - F(x^\dagger)\|_Y \\ &\leq \frac{1}{2}\|F(x) - F(x^\dagger)\|_Y + \frac{1}{2}\|x - x^\dagger\|_X^2. \end{aligned}$$

Using the interpolation inequality (7) in the form

$$\|h\|_X^2 \leq \|h\|_{-1}\|h\|_1 \quad \text{for all } h \in X_1$$

we derive for  $x - x^\dagger \in X_1$  the inequality

$$\|x - x^\dagger\|_{-1} \leq \frac{1}{2}\|F(x) - F(x^\dagger)\|_Y + \frac{1}{2}\|x - x^\dagger\|_1\|x - x^\dagger\|_{-1}$$

and, if moreover  $\|x - x^\dagger\|_1 \leq \kappa < 2$ , even the conditional stability estimate

$$\|x - x^\dagger\|_{-1} \leq \frac{1}{2 - \kappa}\|F(x) - F(x^\dagger)\|_Y \quad \text{for all } x - x^\dagger \in \mathcal{B}_\kappa^{X_1}(0). \quad (36)$$

The estimate (36) can only unfold a stabilizing effect if approximate solutions  $x$  are such that  $x - x^\dagger \in \mathcal{B}_\kappa^{X_1}(0)$  for some  $\kappa < 2$ . For  $x^\dagger$  from (35) with  $x^\dagger(1) = 1 \neq 0$  we have  $x^\dagger \notin X_1$ , but regularized solutions  $x = x_\alpha^\delta$  solving the extremal problem (3) have by definition the property  $x_\alpha^\delta \in X_1$ , which implies that  $x_\alpha^\delta - x^\dagger \notin X_1$ . This is a pitfall, because convergence assertions for  $x_\alpha^\delta$  as  $\delta \rightarrow 0$  are missing in case (c) and thus the behaviour of  $x_\alpha^\delta$  remains completely unclear.

**Model problem 3** (Situation of  $x^\dagger$  meeting case (b)). It is not straight forward to construct an example for case (b) of Case distinction. We base our construction

on the observation that the series  $\sum_{n=2}^{\infty} \frac{1}{n(\log n)^2}$  is convergent, i.e. it characterizes a finite value, whereas the series  $\sum_{n=2}^{\infty} \frac{n^\varepsilon}{n(\log n)^2}$  is divergent for all  $\varepsilon > 0$ , i.e. we

have  $\sum_{n=2}^{\infty} \frac{n^\varepsilon}{n(\log n)^2} = \infty$ . In order to be able to use the model operators and the Hilbert scale introduced before, we use the following integral formulation.

**Lemma 1.** *The improper integral  $\int_2^\infty \frac{1}{x^\eta \log^2(x)} dx$  converges for  $\eta \geq 1$  and diverges for  $\eta < 1$ .*

*Proof.* It is

$$\int \frac{1}{x^\eta \log^2(x)} dx = (1 - \eta)E((1 - \eta) \log x) - \frac{x^{1-\eta}}{\log x} + C,$$

$C \in \mathbb{R}$ , where  $E(z) := \int_z^\infty \frac{e^{-t}}{t} dt$ . The claim follows since

$$\lim_{x \rightarrow \infty} (1 - \eta)E((1 - \eta) \log x) - \frac{x^{1-\eta}}{\log x} = \begin{cases} 0 & \eta \geq 1 \\ \infty & \eta < 1 \end{cases}.$$

□

Hence, we construct exact solutions  $x^\dagger$  via their Fourier transform

$$\hat{x}^\dagger(\xi) := \begin{cases} 0 & |\xi| < 2 \\ \left( \frac{1}{|\xi|(\log|\xi|)^2(1+|\xi|^2)^p} \right)^{\frac{1}{2}} & |\xi| \geq 2 \end{cases} \quad (37)$$

to obtain, after an inverse Fourier transform, solutions  $x^\dagger \in H^p[0, 1]$ , but for arbitrarily small  $\epsilon > 0$   $x \notin H^{p+\epsilon}[0, 1]$ . Namely, for this  $x^\dagger$ , the  $H^p$ -norm in Fourier-domain (24) reads

$$\|x_s^\dagger\|_{H^p}^2 := \int_{\mathbb{R}} \frac{1}{|\xi|(\log|\xi|)^2} d\xi \leq \infty.$$

Since we have little control over the boundary values through this approach, we will only consider the Hilbert scale induced by (26) for  $0 < p < \frac{1}{2}$ .

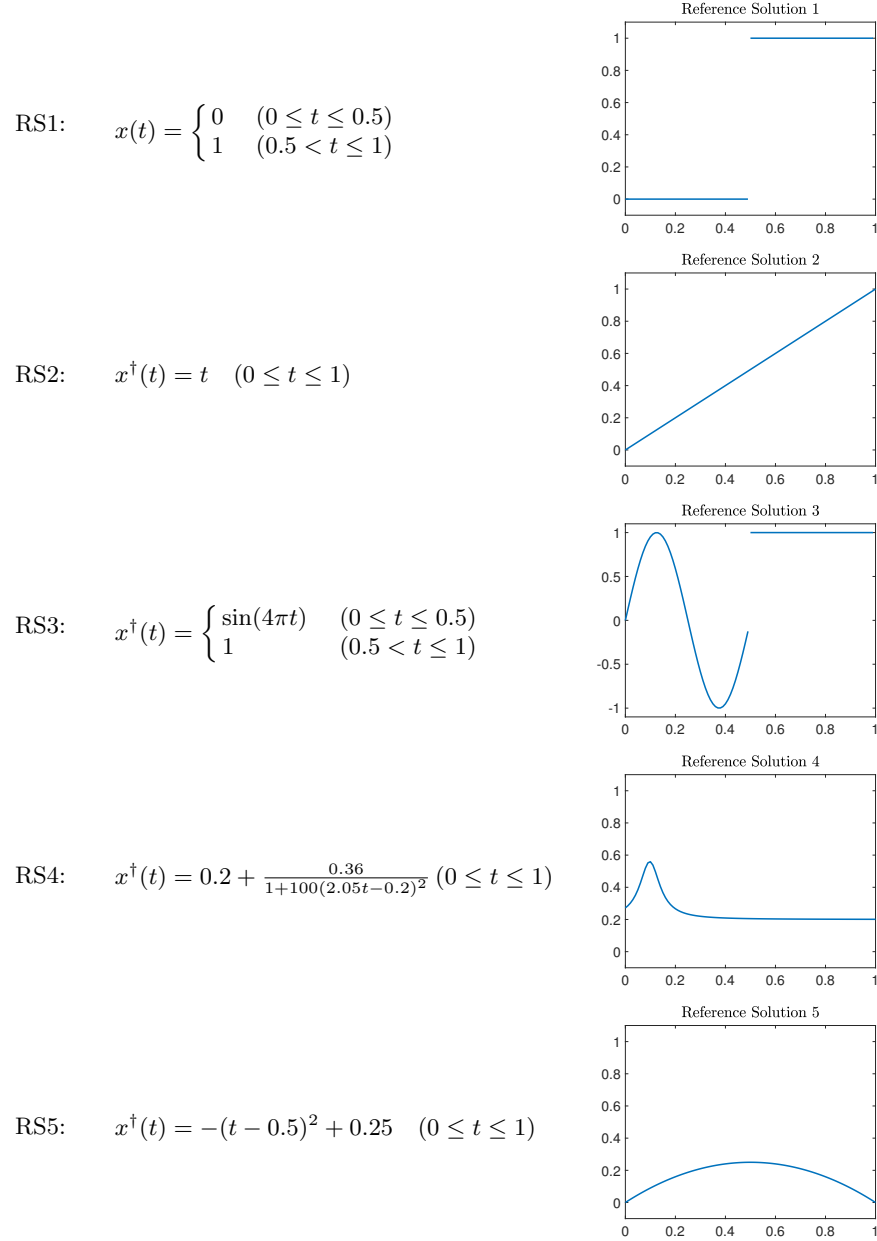
## 5 Case studies

In this section we provide numerical evidence for the behavior of regularized solutions  $x_\alpha^\delta$  with respect to the Case distinction from Section 1 and the Model problems from Section 4.

### 5.1 Numerical studieions for Model Problem 1

We consider the forward operator  $F$  from (30) in the setting  $X = Y = L^2(0, 1)$ ,  $\mathcal{D}(F) = X$ . As was shown, a conditional stability estimate of the form (17) is valid there with  $a = 1$  and  $\gamma = 1$  (cf. formula (33)). It must be emphasized that Assumption 3 applies even in an extended manner, which means that there are finite constants  $K(r) > 0$  for arbitrarily large radii  $r > 0$  such that (33) is valid.

In our first set of experiments, we will investigate the interplay between the value  $p \in (0, 1)$ ,  $\alpha$ -rates of the regularization parameter and error rates of regularized solutions  $x_\alpha^\delta$  using several test cases. To this end, we consider five



**Fig. 1.** Reference solutions used in the first series of experiments. Due to the failure of the respective boundary conditions, we have  $x^\dagger \in X_p$ ,  $0 < p < \frac{1}{2}$  for RS1–RS4, and for RS5 we find  $x^\dagger \in X_p$  with  $p \leq \frac{3}{2}$ .

reference solutions as shown in Figure 1. Of these examples, only RS5 fulfills the boundary condition  $x(1) = 0$ , hence RS1-RS4 can only be an element of  $X_p$  for  $0 < p < \frac{1}{2}$ . Since for RS5  $x'(0) \neq 0$ , we have in this case  $x^\dagger \in H^p[0, 1]$  for  $p \leq \frac{3}{2}$ .

To confirm our theoretical findings of Section 3, we solve (3) using, after discretization via the trapezoidal rule for the integral, the MATLAB®-function `fmincon`. Typically, we use a discretization level  $N = 200$ . To the simulated data  $y = F(x^\dagger)$  we add random noise for which we prescribe the *relative error*  $\bar{\delta}$  such that  $\|y - y^\delta\| = \bar{\delta}\|y\|$ , i.e., we have (2) with  $\delta = \bar{\delta}\|y\|$ . To obtain the  $X_1$  norm in the penalty, we set  $\|\cdot\|_1 = \|\cdot\|_{H^1[0,1]}$  and additionally force the boundary condition  $x(1) = 0$ . The regularization parameter  $\alpha$  is chosen as  $\alpha_{DP} = \alpha(\delta, y^\delta)$  using the discrepancy principle, i.e.,

$$\delta \leq \|F(x_{\alpha_{DP}}^\delta) - y^\delta\|_Y \leq C\delta, \quad (38)$$

with some prescribed multiplier  $C > 1$ . Unless otherwise noted  $C=1.1$  was used. From Section 3, we know that this should yield a  $\alpha$ -rate similar to the a-priori choice

$$\alpha(\delta) \sim \delta^{\frac{4}{p+1}},$$

cf. formula (23), that has already been used by Natterer in [21] for linear problems in the case of oversmoothing penalties. We should also be able to observe the order optimal convergence rate

$$\|x_{\alpha(\delta)}^\delta - x^\dagger\|_X = \mathcal{O}(\delta^{\frac{p}{p+1}}) \quad \text{as} \quad \delta \rightarrow 0.$$

Since  $x^\dagger$  is known, we can compute the regularization errors  $\|x_\alpha^\delta - x^\dagger\|_X$ . We interpret this as a function of  $\delta$  and make a regression for the model function

$$\|x_\alpha^\delta - x^\dagger\|_X \leq c_x \delta^{\kappa_x}, \quad (39)$$

and similarly we estimate the function behind the regularization parameter through the ansatz

$$\alpha \sim c_\alpha \delta^{\kappa_\alpha}. \quad (40)$$

Comparing (39) and the predicted rate (22), we have  $\kappa_x = \frac{p}{a+p}$ , hence we can estimate the smoothness of the solution as  $p = \frac{a\kappa_x}{1-\kappa_x}$ . Recall that  $a = 1$  in this example. Results on regularized solutions  $x_{\alpha_{DP}}^\delta$  with the discrepancy principle for all five reference solutions are summarized in Table 1. From this estimated  $p$ , we can calculate the a-priori parameter choice (23) and compare it to the measured one. Results on regularized solutions  $x_{\alpha_{DP}}^\delta$  with the discrepancy principle for all five reference solutions are summarized in Table 1.

As discussed before, the reference solutions RS1, RS2, RS3 and RS4 all belong to case (c) of the Case distinction whereas RS5 belongs to case (a). Our computed results fit to this narrative. For RS1–RS4 we obtain an estimated  $p < 1$  and  $\kappa_\alpha > 2$ , i.e.,  $\delta^2/\alpha \rightarrow \infty$  as  $\delta \rightarrow 0$ . As expected we have  $0 < \kappa_\alpha < 2$  for RS5 ( $p > 1$ ) together with  $\frac{\delta^2}{\alpha_{DP}} \rightarrow 0$  as  $\delta \rightarrow 0$ . For RS1 we know that  $p$  is bounded above by 0.5 and the estimated value 0.487 fits well. In particular in



RS	$c_x$	$\kappa_x$	$c_\alpha$	$\kappa_\alpha$	est. $p$	$\frac{4}{p+1}$
1	0.9578	0.3276	5.8483	2.7950	0.4871	2.6898
2	0.9017	0.3426	13.5714	2.8609	0.5212	2.6290
3	1.6102	0.4110	0.2782	2.3221	0.6978	2.3560
4	0.2571	0.2582	462.1747	2.7974	0.3481	2.9671
5	0.8868	0.6135	25.8986	1.9546	1.5875	1.5459

**Table 1.** Model Problem 1: Numerically computed convergence rates (39) and  $\alpha$ -rates (40) for the five test cases with estimated values  $p$  from the index  $\kappa_x$ , characterizing approximately the smoothness of the exact solution in these test cases.

the oversmoothing cases, we have an excellent fit between the  $\alpha$ -rates from the discrepancy principle and the a-priori choice based on our estimate of  $p$ .

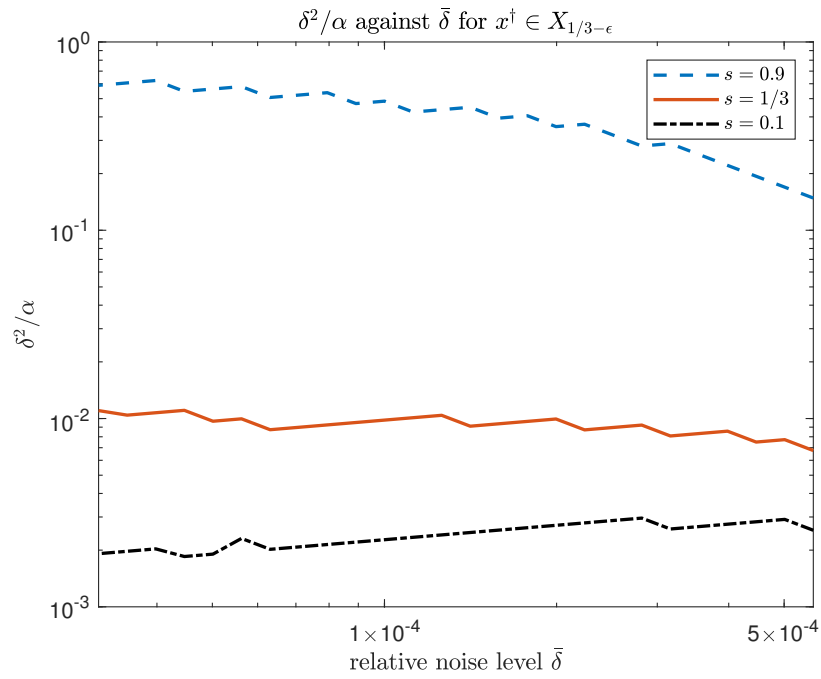
As a second scenario for this model problem, we use the Sobolev scale  $H^p[0, 1]$ ,  $0 < p < \frac{1}{2}$  to investigate a particular case of (c) in our Case distinction. Using the Fourier transform, we construct our solutions  $x^\dagger$  such that  $\hat{x}^\dagger(\xi) := (1 + |\xi|^2)^{-\frac{s}{2} - \frac{1}{4}}$  which yields solutions  $x^\dagger \in H^{p-\epsilon}[0, 1]$  for all  $\epsilon > 0$ , but  $x^\dagger \notin H^p[0, 1]$ . This follows, because  $\int_{\mathbb{R}} (1 + |\xi|^2)^\nu d\xi$  converges for  $\nu < -1/2$  (but diverges for  $\nu \geq -1/2$ ) and the definition of the  $H^p$ -norm (24) in Fourier domain. We take  $p = \frac{1}{3}$  and in principal repeat the previous experiments with the new solutions. The main difference is that we now minimize a Tikhonov functional with variable penalty smoothness,

$$T_{\alpha,s}^\delta(x) := \|F(x) - y^\delta\|_Y^2 + \alpha \|x\|_{H^s[0,1]}^2 \rightarrow \min, \text{ subject to } x \in \mathcal{D}(F). \quad (41)$$

From Section 2 we would expect  $\delta^2/\alpha \rightarrow 0$  for  $s < p = \frac{1}{3}$ ,  $\delta^2/\alpha \approx \text{const}$  for  $s = p = \frac{1}{3}$ , and  $\delta^2/\alpha \rightarrow \infty$  for  $s > p = \frac{1}{3}$ . The numerical results confirm this behavior, see Figure 2 for a plot and Table 2 for the regression results along (39) and (40). Note that in particular the exponent in the convergence rate  $\kappa_x$  remains approximately constant as predicted by the theory.

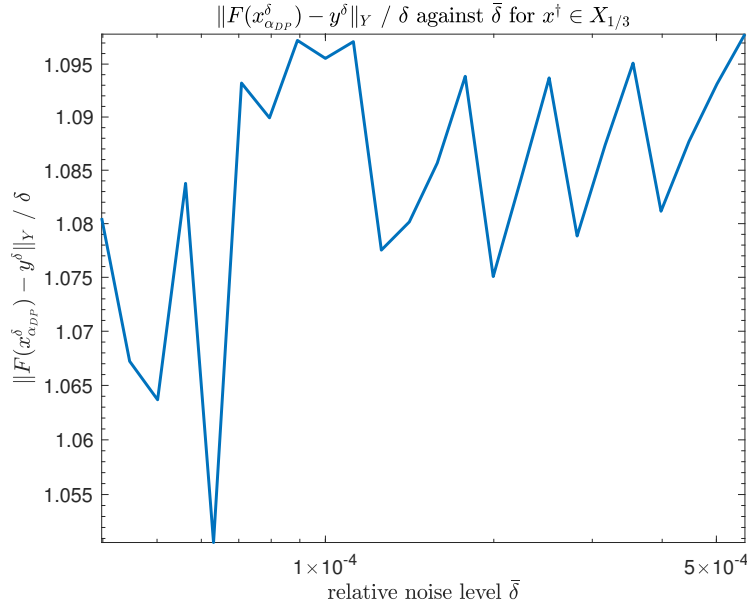
s	$c_x$	$\kappa_x$	$c_\alpha$	$\kappa_\alpha$
0.1	0.9460	0.2647	86.90	1.8168
0.33	1.1492	0.2828	337.42	2.1324
0.9	1.2633	0.2919	250.08	2.5319

**Table 2.** Model Problem 1: numerically computed convergence rates (39) and  $\alpha$ -rates (40) for various  $s$  in (41) for given  $x^\dagger \in X_{1/3-\epsilon}$ .



**Fig. 2.** Model Problem 1:  $\frac{\delta^2}{\alpha}$  for decreasing noise level and various  $s$  in (41);  $x^\dagger \in X_{1/3-\epsilon}$ . In subcase (a) (black, dash-dotted) we have  $\frac{\delta^2}{\alpha} \rightarrow 0$  as  $\delta \rightarrow 0$ , in subcase (b) (red, solid) the quotient stays constant, and in subcase (c) (blue, dashed) we observe the predicted blow up  $\frac{\delta^2}{\alpha} \rightarrow \infty$  as  $\delta \rightarrow 0$ .

Note that the “bumpy” structure in Figure 2 and related plots below are due to the discrepancy principle as exemplified in Figure 3 for  $s = 0.9$ .



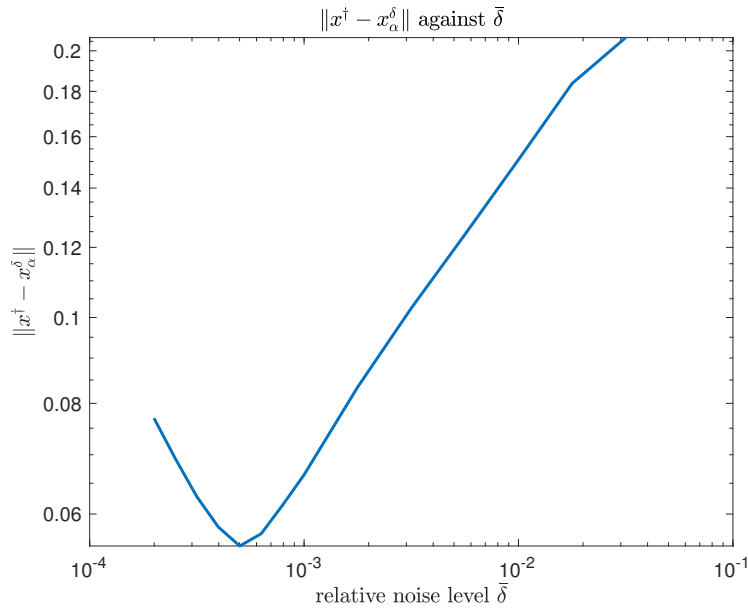
**Fig. 3.** Model Problem 1: Discrepancy constant  $C = \|F(x_{\alpha_{DP}}^\delta) - y^\delta\|_Y / \delta$  for decreasing noise level and  $s=0.9$  in (41);  $x^\dagger \in X_{1/3-\epsilon}$ .

### 5.2 Numerical studies for Model Problem 2

We now turn to the autoconvolution operator  $F$  from (34). In this context, we consider only the specific solution  $x^\dagger(t) = 1$  ( $0 \leq t \leq 1$ ), where  $x^\dagger \notin X_1$ , and the minimization problem (3). It must be emphasized that here, in contrast to model problem 1, Assumption 3 does not apply for any radius  $r > 0$ . It is therefore completely unclear which behavior the regularized solutions  $x_\alpha^\delta$  show when the noise level  $\delta$  tends to zero. It is a pitfall for exploiting Tikhonov regularization to get stable approximations for  $x^\dagger$  when the regularized solutions  $x_\alpha^\delta$  from case (c), i.e.,  $0 < p < 1$ , do not meet the validity area of the conditional stability estimate even if  $\delta > 0$  is sufficiently small. Then due to  $x_\alpha^\delta \notin Q$  estimates of type (5) are useless, convergence  $x_\alpha^\delta \rightarrow x^\dagger$  as  $\delta \rightarrow 0$  cannot be ensured, and the behaviour of the regularized solutions remains unclear. Such a situation occurs, as shown before, in Model problem 2 with  $Q = \mathcal{B}_\kappa^{X_1}(0)$  and  $x^\dagger \notin Q$  for  $x^\dagger \equiv 1$ . The following numerical case studies for that situation enlighten

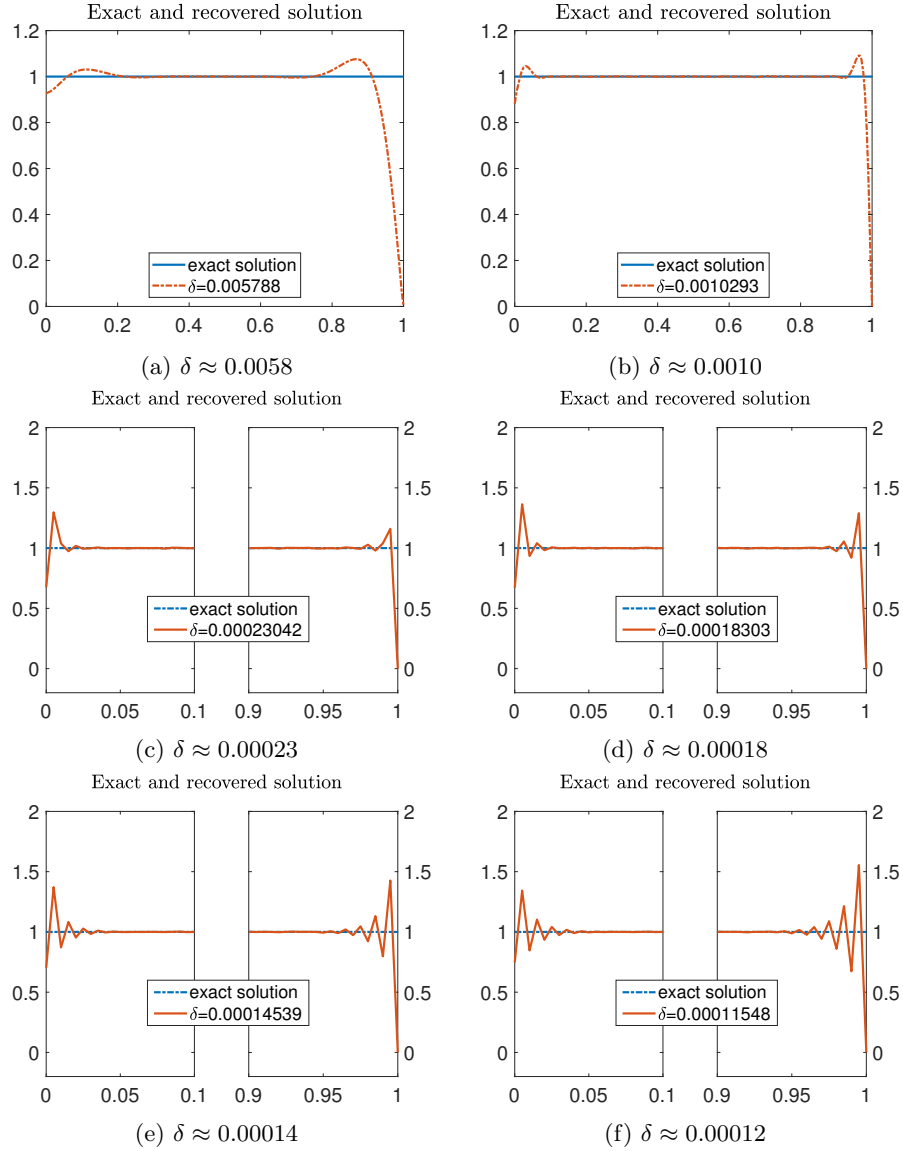
the properties of  $x_\alpha^\delta$ . For the test computations, again the discrepancy principle  $\alpha_{DP} = \alpha(\delta, y^\delta)$  according to (38) has been used with  $C = 1.3$  and a discretization of  $N = 200$  grid points over the interval  $[0, 1]$ . In particular, we demonstrate that  $\|x_{\alpha_{DP}}^\delta - x^\dagger\|_X$  does not tend to zero for  $\delta \rightarrow 0$ .

Figure 4 shows the regularization error  $\|x_{\alpha_{DP}}^\delta - x^\dagger\|_X$  depending on the relative noise level  $\bar{\delta}$ . It can be seen, that  $\|x_{\alpha_{DP}}^\delta - x^\dagger\|_X$  decreases for decreasing noise levels  $\bar{\delta}$  whenever  $\bar{\delta} \geq 2.9 \cdot 10^{-4}$ . If  $\bar{\delta}$  falls below the value  $2.9 \cdot 10^{-4}$ , then the monotonicity turns around and  $\|x^\dagger - x_\alpha^\delta\|_X$  begins to grow. As illustrated in the overview of Figure 5, the regularized solutions tend to oscillate for small  $\delta > 0$ , especially near the left and right boundaries of the interval  $[0, 1]$  in the sense of the Gibbs phenomenon. The Gibbs phenomenon at the right boundary  $t = 1$  accompanies the required jump from one to zero between  $x^\dagger \notin X_1$  and  $x_\alpha^\delta \in X_1$ . The oscillations blow up for small values of  $\delta$  (Figure 5 (c)–(f)) and indicate non-convergence of  $x_\alpha^\delta$  for  $\delta \rightarrow 0$ . Note that the Gibbs phenomenon starts to appear around the minimum of  $\delta^2/\alpha$ , compare to Figure 4.



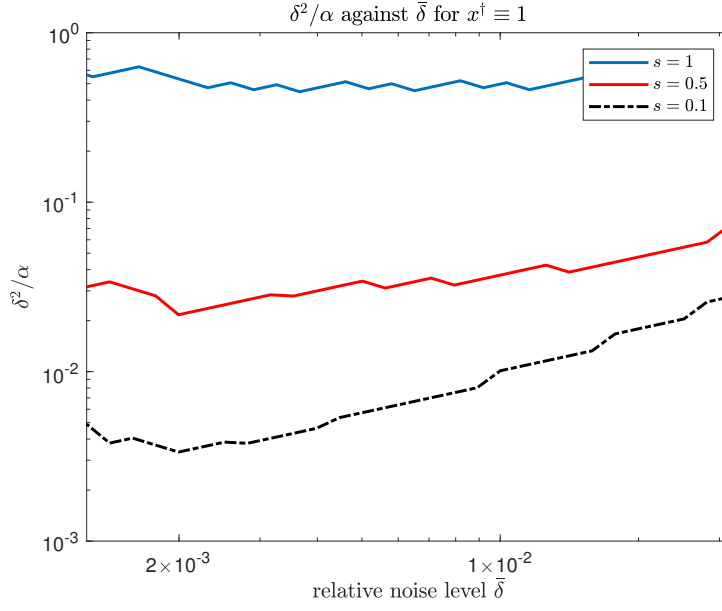
**Fig. 4.** Model Problem 2: Plot of  $\|x_{\alpha_{DP}}^\delta - x^\dagger\|_X$  against  $\delta$  on a logarithmic scale for  $x^\dagger \equiv 1$ . For small values of  $\bar{\delta}$  start to diverge, since  $x^\dagger$  does not belong to the stability set  $Q = \mathcal{B}_\kappa^{X_1}$ .

To confirm that this phenomenon is inherent to the oversmoothing situation, we consider again the Tikhonov functional (41) with  $H^s$ -penalty for  $x^\dagger \equiv 1$ ,  $s = 0.1$  and  $s = 0.5$  respectively. As  $x^\dagger \equiv 1 \in X_p$  for  $0 < p < 1/2$  we expect



**Fig. 5.** Model Problem 2: Regularized and exact solutions with various noise levels,  $x^\dagger \equiv 1$ . To improve the visibility of the blow-up at the boundaries, we omitted the middle part of the functions in the cases (c)-(f).

similar asymptotic behavior of  $\frac{\delta^2}{\alpha}$  for  $\delta \rightarrow 0$  as at the end of Section 5.1. Figure 6 shows the result.

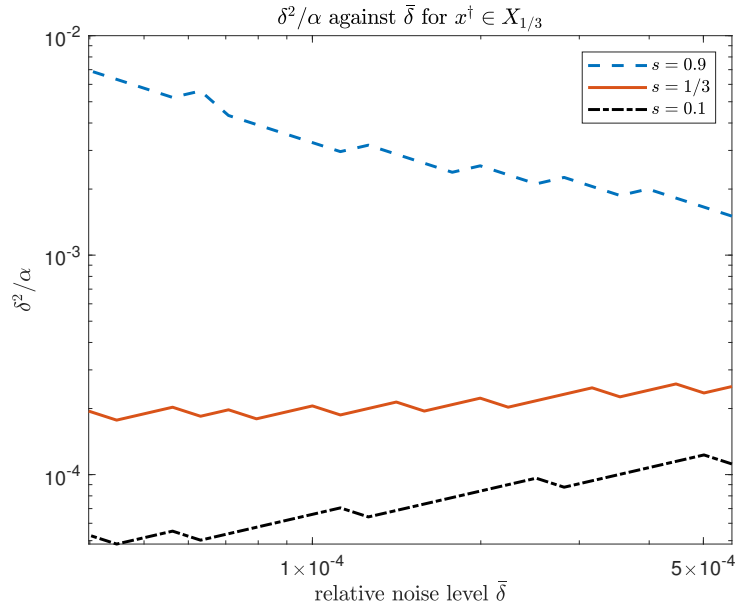


**Fig. 6.** Model Problem 2: Realized values of  $\frac{\delta^2}{\alpha}$  for decreasing noise level and various  $s$  with  $x^\dagger \equiv 1$ .

### 5.3 Numerical studies for Model Problem 3

Based on the case distinction in Section 1 we now study the convergence rates and properties of  $\frac{\delta^2}{\alpha}$  as  $\delta$  decays to zero for the Model Problem 3 in case (b) of Case distinction. Using the Sobolev-scale with norm (24) we define  $x^\dagger \in X_p$ , but  $x^\dagger \notin X_{p+\epsilon}$  via (37). For given  $x^\dagger \in X_p$  in the above sense, we then turn to the Tikhonov functional (41) with penalty in  $H^s[0, 1]$ . Again we choose  $p = \frac{1}{3}$  which means  $x^\dagger \in X_{1/3}$  such that we can employ the theory from Model Problem 1. For  $s > p$  we are in the classical setting and therefore expect  $\frac{\delta^2}{\alpha} \rightarrow 0$  as  $\delta \rightarrow 0$ , for  $s < p$  we are in a oversmoothing situation and expect that  $\frac{\delta^2}{\alpha} \rightarrow \infty$ . Letting  $s = p$  yields precisely case (b) of the Case distinction, and  $\frac{\delta^2}{\alpha}$  should remain approximately constant. The numerical results, see Figure 7 and Table 3, confirm this. Note that, since  $a = 1$  and  $p = \frac{1}{3}$ , we expect and obtain  $\kappa_x = 0.25$ . We also see that the  $\kappa_\alpha < 2$  for  $s=0.1$ ,  $\kappa > 2$  for  $s = 0.9$ , i.e. in the oversmoothing

situation, and  $k \approx 2$  and therefore  $\frac{\delta^2}{\alpha}$  approximately constant for the situation where  $x^\dagger$  and penalty term are of the same smoothness.



**Fig. 7.** Model Problem 3: Realized values of  $\frac{\delta^2}{\alpha}$  for decreasing noise level and various  $s$  in (41);  $x^\dagger \in X_{1/3}$ . In subcase (a) of the Case distinction (black, dash-dotted) we have  $\frac{\delta^2}{\alpha} \rightarrow 0$  as  $\delta \rightarrow 0$ , in subcase (b) (red, solid) the quotient stays constant, and in subcase (c) (blue, dashed) we observe the predicted blow up  $\frac{\delta^2}{\alpha} \rightarrow \infty$  as  $\delta \rightarrow 0$ .

$s$	$c_x$	$\kappa_x$	$c_\alpha$	$\kappa_\alpha$
0.1	0.1275	0.2531	6.04e+02	1.6479
0.33	0.1228	0.2461	1.66e+03	1.8805
0.9	0.1229	0.2427	3.84e+04	2.5385

**Table 3.** Model Problem 3: Numerically computed convergence rates (39) and  $\alpha$ -rates (40) in case (b) for various  $s$  in (41) and  $x^\dagger \in X_{1/3}$ .

## Appendix: Proof of Proposition 4

In this proof we set  $E := \|x^\dagger\|_p$ . To prove the convergence rate result (20) under the a priori parameter choice (21) it is sufficient to show that for sufficiently small  $\delta > 0$  there are two constants  $K > 0$  and  $\tilde{E} > 0$  such that the inequalities

$$\|x_{\alpha_*}^\delta - x^\dagger\|_{-a} \leq K\delta \quad (42)$$

and

$$\|x_{\alpha_*}^\delta - x^\dagger\|_p \leq \tilde{E} \quad (43)$$

hold. Namely, the convergence rate (20) follows directly from inequality chain

$$\|x_{\alpha_*}^\delta - x^\dagger\|_X \leq \|x_{\alpha_*}^\delta - x^\dagger\|_{-a}^{\frac{p}{a+p}} \|x_{\alpha_*}^\delta - x^\dagger\|_p^{\frac{a}{a+p}} \leq K^{\frac{p}{a+p}} \tilde{E}^{\frac{a}{a+p}} \delta^{\frac{p}{a+p}},$$

which is valid for sufficiently small  $\delta > 0$  as a consequence of (42), (43) and of the interpolation inequality for the Hilbert scale  $\{X_\tau\}_{\tau \in \mathbb{R}}$ .

As an essential tool for the proof we use auxiliary elements  $x_\alpha$ , which are for all  $\alpha > 0$  the uniquely determined minimizers over all  $x \in X$  of the artificial Tikhonov functional

$$T_{-a,\alpha}(x) := \|x - x^\dagger\|_{-a}^2 + \alpha \|Bx\|_X^2. \quad (44)$$

Note that the elements  $x_\alpha$  are independent of the noise level  $\delta > 0$  and belong by definition to  $X_1$ , which is in strong contrast to  $x^\dagger \notin X_1$ .

The following lemma is an immediate consequence of [13, Prop. 2], see also [14, Prop. 3].

**Lemma 2.** *Let  $\|x^\dagger\|_p = E$  and  $x_\alpha$  be the minimizer of the functional  $T_{-a,\alpha}$  from (44) over all  $x \in X$ . Given  $\alpha_* = \alpha_*(\delta) > 0$  as defined by formula (21) the resulting element  $x_{\alpha_*}$  obeys the bounds*

$$\|x_{\alpha_*} - x^\dagger\|_X \leq E\delta^{p/(a+p)}, \quad (45)$$

$$\|B^{-a}(x_{\alpha_*} - x^\dagger)\|_X \leq E\delta, \quad (46)$$

$$\|Bx_{\alpha_*}\|_X \leq E\delta^{(p-1)/(a+p)} = E \frac{\delta}{\sqrt{\alpha_*}} \quad (47)$$



and

$$\|x_{\alpha_*} - x^\dagger\|_p \leq E.$$

Due to (45) we have  $\|x_{\alpha_*} - x^\dagger\|_X \rightarrow 0$  as  $\delta \rightarrow 0$ . Hence by Assumption 4, in particular because  $x^\dagger$  is an interior point of  $\mathcal{D}(F)$ , for sufficiently small  $\delta > 0$  the element  $x_{\alpha_*}$  belongs to  $\mathcal{B}_r^X(x^\dagger) \subset \mathcal{D}(F)$  and moreover with  $x_{\alpha_*} \in X_1$  the inequality (19) applies for  $x = x_{\alpha_*}$  and such small  $\delta$ .

Instead of the inequality (8), which is missing in case of oversmoothing penalties, we can use here the inequality

$$T_{\alpha_*}^\delta(x_{\alpha_*}^\delta) \leq T_{\alpha_*}^\delta(x_{\alpha_*}). \quad (48)$$

as minimizing property for the Tikhonov functional. Using (48) it is enough to bound  $T_{\alpha_*}^\delta(x_{\alpha_*})$  by  $\bar{C}^2 \delta^2$  with

$$\bar{C} := ((\bar{K}E + 1)^2 + E^2)^{1/2} \quad (49)$$

in order to obtain the estimates

$$\|F(x_{\alpha_*}^\delta) - y^\delta\|_Y \leq \bar{C} \delta \quad (50)$$

and

$$\|Bx_{\alpha_*}^\delta\|_X \leq \bar{C} \frac{\delta}{\sqrt{\alpha_*}}. \quad (51)$$

Since the inequality (19) applies for  $x = x_{\alpha_*}$  and sufficiently small  $\delta > 0$ , we can estimate for such  $\delta$  as follows:

$$\begin{aligned} T_{\alpha_*}^\delta(x_{\alpha_*}) &\leq (\|F(x_{\alpha_*}) - F(x^\dagger)\|_Y + \|F(x^\dagger) - y^\delta\|_Y)^2 + \alpha_* \|Bx_{\alpha_*}\|_X^2 \\ &\leq (\bar{K}\|x_{\alpha_*} - x^\dagger\|_{-a} + \delta)^2 + E^2 \alpha_* \delta^{2(p-1)/(a+p)} \\ &\leq (\bar{K}E\delta + \delta)^2 + E^2 \delta^2 \\ &= ((\bar{K}E + 1)^2 + E^2) \delta^2. \end{aligned}$$

This ensures the estimates (50) and (51). Based on this we are going now to show that an inequality (42) is valid for some  $K > 0$ . Here, we use the inequality (18) of Assumption 4, which applies for  $x = x_{\alpha_*}^\delta$ , and we find

$$\begin{aligned} \|x_{\alpha_*}^\delta - x^\dagger\|_{-a} &\leq \frac{1}{\underline{K}} \|F(x_{\alpha_*}^\delta) - F(x^\dagger)\|_Y \\ &\leq \frac{1}{\underline{K}} (\|F(x_{\alpha_*}^\delta) - y^\delta\|_Y + \|F(x^\dagger) - y^\delta\|_Y) \\ &\leq \frac{1}{\underline{K}} (\bar{C}\delta + \delta) = \frac{1}{\underline{K}} (\bar{C} + 1) \delta = K\delta, \end{aligned}$$

where  $\bar{C}$  is the constant from (49) and we derive  $K := \frac{1}{\underline{K}} (\bar{C} + 1)$ .

Secondly, we still have to show the existence of a constant  $\tilde{E} > 0$  such that the inequality (43) holds. By using the triangle inequality in combination with (51) and (47) we find that

$$\|B(x_{\alpha_*}^\delta - x_{\alpha_*})\|_X \leq \|Bx_{\alpha_*}^\delta\|_X + \|Bx_{\alpha_*}\|_X \leq (\bar{C} + E) \frac{\delta}{\sqrt{\alpha_*}}.$$

Again, we use the interpolation inequality and can estimate further as

$$\begin{aligned} \|x_{\alpha_*}^\delta - x_{\alpha_*}\|_p &\leq \|x_{\alpha_*}^\delta - x_{\alpha_*}\|_1^{\frac{a+p}{a+1}} \|x_{\alpha_*}^\delta - x_{\alpha_*}\|_{-a}^{\frac{1-p}{a+1}} \\ &\leq \left( (\bar{C} + E) \frac{\delta}{\sqrt{\alpha_*}} \right)^{\frac{a+p}{a+1}} (\|x_{\alpha_*}^\delta - x^\dagger\|_{-a} + \|x^\dagger - x_{\alpha_*}\|_{-a})^{\frac{1-p}{a+1}} \\ &\leq \left( (\bar{C} + E) \frac{\delta}{\sqrt{\alpha_*}} \right)^{\frac{a+p}{a+1}} ((K + E)\delta)^{\frac{1-p}{a+1}} \\ &= \left( (\bar{C} + E)\delta^{(p-1)/(a+p)} \right)^{\frac{a+p}{a+1}} ((K + E)\delta)^{\frac{1-p}{a+1}} =: \bar{E}. \end{aligned}$$

Finally, we have now

$$\|x_{\alpha_*}^\delta - x^\dagger\|_p \leq \|x_{\alpha_*}^\delta - x_{\alpha_*}\|_p + \|x_{\alpha_*} - x^\dagger\|_p \leq \bar{E} + E =: \tilde{E}.$$

This shows (43) and thus completes the proof of Proposition 4.  $\square$

## Acknowledgment

We thank the colleagues Volker Michel and Robert Plato from the University of Siegen for a hint to the series that allowed us to formulate Model problem 3. The research was financially supported by Deutsche Forschungsgemeinschaft (DFG-grant HO 1454/12-1).

## References

1. R. A. Adams and J. F. J. Fournier. *Sobolev spaces*. Elsevier/Academic Press, Amsterdam, 2003.
2. S. Bürger and B. Hofmann. About a deficit in low order convergence rates on the example of autoconvolution. *Applicable Analysis*, 94:477–493, 2015.
3. J. Cheng and M. Yamamoto. On new strategy for *a priori* choice of regularizing parameters in Tikhonov’s regularization. *Inverse Problems*, 16:L31–L38, 2000.
4. H. Egger and B. Hofmann. Tikhonov regularization in Hilbert scales under conditional stability assumptions. *Inverse Problems*, 34:115015 (17pp), 2018.
5. H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

6. J. Flemming. *Variational Source Conditions, Quadratic Inverse Problems, Sparsity Promoting Regularization – New Results in Modern Theory of Inverse Problems and an Application in Laser Optics*. Birkhäuser, Basel, 2018.
7. D. Gerth and B. Hofmann. Oversmoothing regularization with  $\ell^1$ -penalty term *AIMS Mathematics*, accepted for publication.
8. R. Gorenflo and B. Hofmann. On autoconvolution and regularization. *Inverse Problems*, 10:353–373, 1994.
9. R. Gorenflo, Y. Luchko, and M. Yamamoto. Time-fractional diffusion equation in the fractional Sobolev spaces. *Fract. Calc. Appl. Anal.*, 18(3):799–820, 2015.
10. R. Gorenflo and M. Yamamoto. Operator-theoretic treatment of linear Abel integral equations of first kind. *Japan J. Indust. Appl. Math.*, 16(1):137–161, 1999.
11. C. W. Groetsch. *Inverse Problems in the Mathematical Sciences*. Vieweg Mathematics for Scientists and Engineers. Vieweg, Braunschweig, 1993.
12. B. Hofmann. A local stability analysis of nonlinear inverse problems. In: *Inverse Problems in Engineering - Theory and Practice* (Eds.: D. Delaunay et al.). The American Society of Mechanical Engineers, New York 1998, pp. 313–320.
13. B. Hofmann and P. Mathé. Tikhonov regularization with oversmoothing penalty for non-linear ill-posed problems in Hilbert scales. *Inverse Problems*, 34:015007 (14pp), 2018.
14. B. Hofmann and P. Mathé. A priori parameter choice in Tikhonov regularization with oversmoothing penalty for non-linear ill-posed problems. To appear in the same volume. Preprint available under arXiv:1904.02014 [math.NA].
15. B. Hofmann and R. Plato. On ill-posedness concepts, stable solvability and saturation. *J. Inverse Ill-Posed Probl.*, 26:287–297, 2018.
16. B. Hofmann and O. Scherzer. Factors influencing the ill-posedness of nonlinear problems. *Inverse Problems*, 10:1277–1297, 1994.
17. B. Hofmann and M. Yamamoto. On the interplay of source conditions and variational inequalities for nonlinear ill-posed problems. *Appl. Anal.*, 89:1705–1727, 2010.
18. S. Krein and Y. Petunin. Scales of Banach spaces. *Russian Mathematical Surveys*, 21:85–159, 1966.
19. A. K. Louis *Inverse und schlecht gestellte Probleme*. Teubner, Stuttgart, 1989.
20. M. T. Nair, S. Pereverzev and U. Tautenhahn. Regularization in Hilbert scales under general smoothing conditions. *Inverse Problems*, 21(6):1851–1869, 2005.
21. F. Natterer. Error bounds for Tikhonov regularization in Hilbert scales. *Applicable Anal.*, 18(1-2):29–37, 1984.
22. A. Neubauer. When do Sobolev spaces form a Hilbert scale? *Proc. Amer. Math. Soc.*, 103:557–562, 1988.
23. R. Plato, B. Hofmann and P. Mathé. Optimal rates for Lavrentiev regularization with adjoint source conditions. *Math. Comp.*, 87(310):785–801, 2018.
24. T. Schuster, B. Kaltenbacher, B. Hofmann, and K. S. Kazimierski. *Regularization methods in Banach spaces*, vol. 10 of Radon Series on Computational and Applied Mathematics. Walter de Gruyter, Berlin/Boston, 2012.
25. U. Tautenhahn. Error estimates for regularized solutions of nonlinear ill-posed problems. *Inverse Problems*, 10:485–500, 1994.
26. U. Tautenhahn. On a general regularization scheme for nonlinear ill-posed problems II: Regularization in Hilbert scales. *Inverse Problems*, 14:1607–1616, 1998.