The Impact of the Discrepancy Principle on the Tikhonov-Regularized Solutions with Oversmoothing Penalties

Bernd Hofmann * and **Christopher Hofmann**

Faculty of Mathematics, Chemnitz University of Technology, 09107 Chemnitz, Germany; christopher.hofmann@mathematik.tu-chemnitz.de

* Correspondence: bernd.hofmann@mathematik.tu-chemnitz.de; Tel.: +49-371-531-34125

check for updates

Abstract: This paper deals with the Tikhonov regularization for nonlinear ill-posed operator equations in Hilbert scales with oversmoothing penalties. One focus is on the application of the discrepancy principle for choosing the regularization parameter and its consequences. Numerical case studies are performed in order to complement analytical results concerning the oversmoothing situation. For example, case studies are presented for exact solutions of Hölder type smoothness with a low Hölder exponent. Moreover, the regularization parameter choice using the discrepancy principle, for which rate results are proven in the oversmoothing case in in reference (Hofmann, B.; Mathé, P. Inverse Probl. 2018, 34, 015007) is compared to Hölder type a priori choices. On the other hand, well-known analytical results on the existence and convergence of regularized solutions are summarized and partially augmented. In particular, a sketch for a novel proof to derive Hölder convergence rates in the case of oversmoothing penalties is given, extending ideas from in reference (Hofmann, B.; Plato, R. ETNA. 2020, 93).

1. Introduction

This paper tries to complement the theory and practice of Tikhonov regularization with oversmoothing penalties for the stable approximate solution of nonlinear ill-posed problems in a Hilbert scale setting. Thus, we consider the operator equation

$$F(x) = y \tag{1}$$

with a nonlinear forward operator $F : \mathcal{D}(F) \subseteq X \to Y$, possessing the domain $\mathcal{D}(F)$ and mapping between the infinite dimensional real Hilbert spaces X and Y. In this context, $\| \cdot \|_X$ and $\| \cdot \|_Y$ denote the norms in X and Y, respectively. Throughout the paper, let $x^\dagger \in \mathcal{D}(F)$ be a solution to Equation (1) for a given right-hand side y. We restrict our considerations to problems that are locally ill-posed at $x^\dagger$. This means that the replacement of the exact right-hand side y by noisy data $y^\delta \in Y$, obeying the deterministic noise model

$$\|y - y^\delta\|_Y \leq \delta \tag{2}$$

with noise level $\delta > 0$, may lead to significant errors in the solution of Equation (1) measured by the X-norm, even if δ tends to zero (cf. ([1], Def. 2) for details).

For finding approximate solutions, we apply a Hilbert scale setting, where the densely defined, unbounded and self-adjoint linear operator $B \colon \mathcal{D}(B) \subset X \to X$ with domain $\mathcal{D}(B)$ generates the Hilbert scale. This operator is assumed to be strictly positive such that we have for some $\underline{m} > 0$

$$\|Bx\|_X \geq \underline{m}\|x\|_X, \quad \text{for all} \quad x \in \mathcal{D}(B). \tag{3}$$

In this sense, we exploit the Hilbert scale $\{X_\tau\}_{\tau \in \mathbb{R}}$ generated by B with $X_0 = X$, $X_\tau = \mathcal{D}(B^\tau)$, and with corresponding norms $\|x\|_\tau := \|B^\tau x\|_X$.

As approximate solutions to $x^\dagger$, we use Tikhonov-regularized solutions $x_\alpha^\delta \in \mathcal{D}(F)$ that are minimizers of the extremal problem

$$T_\alpha^\delta(x) := \|F(x) - y^\delta\|_Y^2 + \alpha\|Bx\|_X^2 \to \min, \quad \text{subject to} \quad x \in \mathcal{D} := \mathcal{D}(F) \cap \mathcal{D}(B), \tag{4}$$

where $\alpha > 0$ is the regularization parameter and $\|F(x) - y^\delta\|_Y^2$ characterizes the misfit or fidelity term. The penalty functional $\|Bx\|_X^2 = \|x\|_1^2$ in the Tikhonov functional T_α^δ is adjusted to the level one of Hilbert scale such that all regularized solutions have the property $x_\alpha^\delta \in \mathcal{D}(B)$. A more general form of the penalty functional in the Tikhonov functional would be $\|B(x - \bar{x})\|_X^2$, where $\bar{x} \in \mathcal{D}$ denotes a given smooth reference element. $\bar{x}$ then plays the role of the origin (the point of central interest), which can be very different for nonlinear problems. Without a loss of generality, we set in the sequel $\bar{x} := 0$, which makes the formulas simpler.

In our study, the discrepancy principle named after Morozov (cf. [2]) as the most prominent a posteriori choice of the regularization parameter $\alpha > 0$ plays a substantial role. On the one hand, the simplified version

$$\alpha_{discr} := \alpha(\delta, y^\delta) \colon \quad \|F(x_{\alpha_{discr}}^\delta) - y^\delta\|_Y = C\,\delta \tag{5}$$

of the discrepancy principle in equation form with a prescribed constant $C > 1$ is important for theory (cf. [3]). However, it is well known that there are nonlinear problems, where that version is problematic due to duality gaps that prevent the solvability of Equation (5). For overcoming the remaining weaknesses of the parameter choice expressed in Equation (5), sequential versions of the discrepancy principle can be applied that approximate α_{discr}, and we refer to [4–6] for more details. Such an approach is used for performing the numerical case studies in Section 6.

Our focus is on oversmoothing penalties in the Tikhonov functional T_α^δ, where $x^\dagger \notin \mathcal{D}(B) = X_1$ such that $T_\alpha^\delta(x^\dagger) = \infty$. In this case, the regularizing property $T_\alpha^\delta(x_\alpha^\delta) \leq T_\alpha^\delta(x^\dagger)$ does not yield any information. This property, however, is the basic tool for obtaining error estimates and convergence assertions for the Tikhonov-regularized solutions in the standard case, where $T_\alpha^\delta(x^\dagger) < \infty$. We refer as an example to Chapter 10 of the monograph [7], which also deals with nonlinear operator equations, but adjusts the penalty functional to level zero. To derive error estimates in the oversmoothing case, the regularizing property must be replaced by inequalities of the form $T_\alpha^\delta(x_\alpha^\delta) \leq T_\alpha^\delta(x_{aux})$, where $x_{aux} \in \mathcal{D}$ is an appropriately chosen auxiliary element.

The seminal paper on Tikhonov regularization in Hilbert scales that includes the oversmoothing case was written by F. Natterer in 1984 (cf. [8]) and was restricted to linear operator equations. Error estimates in the X-norm and convergence rates were proven under two-sided inequalities that characterize the degree of ill-posedness $a > 0$ of the problem. We follow this approach and adapt it to the case of nonlinear problems throughout the subsequent sections and assume the inequality chain

$$c_a\|x - x^\dagger\|_{-a} \leq \|F(x) - F(x^\dagger)\|_Y \leq C_a\|x - x^\dagger\|_{-a} \quad \text{for all} \quad x \in \mathcal{D} \tag{6}$$

and constants $0 < c_a \leq C_a < \infty$. The left-hand inequality in Equation (6) represents a conditional stability estimate and is substantial for obtaining stable regularized solutions, whereas the right-hand inequality in Equation (6) contributes to the determination of the nonlinearity structure of the forward operator F. Convergence and rate results for the Tikhonov regularization expressed in Equation (4) with oversmoothing penalties under the inequality chain expressed in Equation (6) were recently

presented in [3,6,9] and complemented by case studies in [10]. The present paper continues this series of articles by addressing open questions with respect to the discrepancy principle for choosing the regularization parameter α and its comparison to a priori parameter choices. In this context, one of the examples from [10] is reused for performing new numerical experiments in order to obtain additional assertions that cannot be taken from analytical investigations.

The paper is organized as follows: We summarize in Section 2 basic properties of regularized solutions under assumptions that are typical for oversmoothing penalties and in Section 3 assertions concerning the convergence. In Section 4 we show that the error estimates derived in [6] for obtaining low order convergence rates are also applicable to obtain the order optimal Hölder convergence rates under the associated Hölder-type source conditions. Section 5 recalls a nonlinear inverse problem from an exponential growth model and an appropriate Hilbert scale, which can both be used for performing numerical experiments in the subsequent section. In that section (Section 6), the obtained numerical results are presented and interpreted based on a series of tables and figures.

2. Assumptions and Properties of Regularized Solutions

In this section, we formulate the standing assumptions concerning the forward operator F, the Tikhonov functional T_α^δ, and the solution $x^\dagger$ of Equation (1) in order to ensure the existence and stability of regularized solutions x_α^δ for all regularization parameters $\alpha > 0$ and noisy data y^δ.

Assumption 1.

(a) *The operator $F : \mathcal{D}(F) \subseteq X \to Y$ mapping between the real Hilbert spaces X and Y is weakly sequentially continuous, and its domain $\mathcal{D}(F)$ with $0 \in \mathcal{D}(F)$ is a convex and closed subset of X.*

(b) *The generator $B : \mathcal{D}(B) \subset X \to X$ of the Hilbert scale is a densely defined, unbounded, and self-adjoint linear operator that satisfies the inequality expressed in Equation (3).*

(c) *The solution $x^\dagger \in \mathcal{D}(F)$ of Equation (1) is supposed to be an interior point of the domain $\mathcal{D}(F)$.*

(d) *To characterize the case of oversmoothing penalties, we assume that*

$$x^\dagger \notin \mathcal{D}(B) = X_1. \tag{7}$$

(e) *There is a number $a > 0$, and there are constants $0 < c_a \le C_a < \infty$ such that the two-sided estimates expressed in Equation (6) hold.*

As a specific impact of Item (d) of Assumption 1 on approximate solutions to $x^\dagger$, we have the following proposition that is of interest for the behavior of regularized solutions in the case of oversmoothing penalties.

Proposition 1. *Let a sequence $\{x_n\} \subset \mathcal{D}(B) \subset X$ converge weakly in X to $x^\dagger \notin \mathcal{D}(B)$ as $n \to \infty$. We then have $\lim_{n\to\infty} \|Bx_n\|_X = \infty$.*

Proof. In order to construct a contradiction, let us assume that the sequence $\{x_n\} \subset \mathcal{D}(B)$ (or some of its subsequences) is bounded in X_1, i.e., $\|Bx_n\|_X \le K$ for all $n \in \mathbb{N}$. Thus, a subsequence of $\{Bx_n\}$ converges weakly in X to some element $z \in X$, because bounded sets are weakly pre-compact in the Hilbert space X. Since the operator B is densely defined and self-adjoint, it is closed, i.e., the graph $\{(x, Bx) : x \in \mathcal{D}(B)\}$ is closed and, due to the convexity of this set, a weakly closed subset of $X \times Y$. Hence, the operator B is weakly closed, which implies that $x^\dagger \in \mathcal{D}(B)$ and $Bx^\dagger = z$. This, however, contradicts the assumed property $x^\dagger \notin \mathcal{D}(B)$ and proves the proposition. $\square$

Remark 1. As a consequence of Proposition 1, we have for any sequence of regularized solutions $\{x_n = x_{\alpha_n}^{\delta_n}\}$, which is norm-convergent (and thus also weak-convergent) to $x^\dagger \notin \mathcal{D}(B)$ for $\delta_n \to 0$ as $n \to \infty$, such that it blows up to infinity with respect to the X_1-norm. In other words, we have the limit condition $\lim_{n\to\infty} \|Bx_{\alpha_n}^{\delta_n}\|_X = \|Bx^\dagger\|_X = \infty$.

Based on Lemma 1, we can formulate in Proposition 2 the existence of minimizers to the extremal problem expressed in Equation (4).

Lemma 1. *The non-negative penalty functional $\|B \cdot\|_X^2 : \mathcal{D}(B) \subset X \to \mathbb{R}$ as part of the Tikhonov functional T_α^δ is a proper convex and a lower, semi-continuous, and stabilizing functional.*

Proof. The obviously convex penalty functional is proper, since it attains finite values for all $x \in \mathcal{D}(B) = X_1 \neq \emptyset$. It is also a stabilizing functional because, as a consequence of Equation (3), the sub-level sets $\{x \in \mathcal{D}(B) : \|Bx\|_X^2 \le c\}$ are weakly sequentially pre-compact subsets in X for all constants $c \ge 0$. Namely, all such non-empty sub-level sets are bounded in X and hence weakly pre-compact. For showing that the functional $\|B \cdot\|_X^2 = \|\cdot\|_1^2$ is lower semi-continuous, by taking into account Proposition 1 and its proof, it is enough to show that a sequence $\{x_n\} \subset \mathcal{D}(B)$ with $\|x_n\|_1 \le K < \infty$ for all $n \in \mathbb{N}$ that converges weakly in X to $\hat{x} \in X$ implies that $\hat{x} \in \mathcal{D}(B)$ and that this sequence also converges weakly in X_1 to $\hat{x}$. The lower semi-continuity of the norm functional $\|\cdot\|_1$ then yields $\|B\hat{x}\|_X^2 \le \liminf_{n\to\infty} \|Bx_n\|_X^2$. Now note that a subsequence $\{x_n\}$ bounded in X_1 has a subsequence that converges weakly in X_1 to some element $z \in X_1$. Since the operator B is weakly closed, we then have $B\hat{x} = z$. Since z is uniform for all subsequences, this completes the proof. □

Proposition 2. *For all $\alpha > 0$ and $y^\delta \in Y$, there is a regularized solution $x_\alpha^\delta \in \mathcal{D}$, solving the extremal problem expressed in Equation (4).*

Proof. Proposition 4.1 from [11], which coincides with our proposition, is immediately applicable, since the Assumptions 3.11 and 3.22 from [11] are satisfied due to Assumption 1 and Lemma 1 above. □

In addition to the existence assertion of Proposition 2, we also have, under the assumptions stated above, the stability of regularized solutions, which means that small changes in the data y^δ yield only small changes in x_α^δ. For a detailed description of this fact, see Proposition 4.2 from [11] that applies here under Assumption 1.

Remark 2. From Assumption 1, we have that there are no solutions $x^* \in \mathcal{D} = \mathcal{D}(F) \cap \mathcal{D}(B)$, satisfying with $F(x^*) = y$ the operator expressed in Equation (1), because this would contradict, with $F(x^\dagger) = F(x^*)$ and $\|x^\dagger - x^*\|_X > 0$, the left-hand inequality of Equation (6). Besides $x^\dagger$, however, other solutions with $x^* \notin \mathcal{D}(B)$ may exist. The regularized solutions x_α^δ, for fixed $\alpha > 0$ and $y^\delta \in Y$, need not be uniquely determined, since, though possessing a convex part $\|Bx\|^2$, the Tikhonov functional $T_\alpha^\delta(x)$ is not necessarily convex.

3. Convergence of Regularized Solutions in the Case of Oversmoothing Penalties

In this section, we discuss assertions about the X-norm convergence of regularized solutions with the Tikhonov functional T_α^δ introduced in Equation (4). First we recall the following lemma (from [6], Proposition 3.4).

Lemma 2. *Under Assumption 1, we have, for regularized solutions $x_\alpha^\delta \in \mathcal{D}$ solving the extremal problem expressed in Equation (4), a function $\varphi : [0, \infty) \to [0, \infty)$ satisfying the limit condition $\lim_{\alpha\to0} \varphi(\alpha) = 0$ and a constant $\overline{K} > 0$ such that the error estimate*

$$\|x_\alpha^\delta - x^\dagger\|_X \le \varphi(\alpha) + \overline{K} \frac{\delta}{\alpha^{a/(2a+2)}} \tag{8}$$

is valid for all $\delta > 0$ and for sufficiently small $\alpha > 0$.

From Lemma 2, we directly obtain the following proposition (cf. [6], Theorem 4.1):

Theorem 1. *For any a priori parameter choice* $\alpha_* = \alpha(\delta)$ *and any a posteriori parameter choice* $\alpha_* = \alpha(\delta, y^\delta)$, *the regularized solutions* $x^\delta_{\alpha_*}$ *converge under Assumption 1 to the solution* $x^\dagger$ *of the operator expressed in Equation (1) for* $\delta \to 0$, *i.e.,*

$$\lim_{\delta \to 0} \|x^\delta_{\alpha_*} - x^\dagger\|_X = 0, \tag{9}$$

whenever

$$\alpha_* \to 0 \quad and \quad \frac{\delta^2}{\alpha_*^{a/(a+1)}} \to 0 \quad as \quad \delta \to 0. \tag{10}$$

Remark 3. By inspection of the corresponding proofs in [6], it becomes clear that the validity of Lemma 2 and consequently of Theorem 1 is not restricted to the case of oversmoothing penalties, but it holds if Items (a), (b), (c), and (e) of Assumption 1 are fulfilled. This means that the solution $x^\dagger \in \mathcal{D}(F)$ can possess arbitrary smoothness.

Example 1. In this example, we consider with respect to Theorem 1 the a priori parameter choice

$$\alpha_* = \alpha(\delta) \sim \delta^\kappa \tag{11}$$

for varying exponents $\kappa > 0$. As the following proposition, as a consequence of Theorem 1, indicates, there is a wide range of exponents κ yielding convergence.

Proposition 3. *For the a priori choice expressed in Equation (11) of the regularization parameter* $\alpha > 0$, *the condition expressed in Equation (10) in Theorem 1 holds if and only if* $0 < \kappa < 2 + \frac{2}{a}$.

However, the proof of the underlying Theorem 4.1 in [6] shows that the general verification of the basic estimate expressed in Equation (8), developed with a focus on the case of oversmoothing penalties, requires both the left-hand inequality as well as the right-hand inequality in the nonlinearity condition expressed in Equation (6) and, moreover, that $x^\dagger$ is an interior point of $\mathcal{D}(F)$. More discussions in that direction can be made if we distinguish the following three κ-intervals:

(i): $0 < \kappa < 2$ with

$$\alpha_* \to 0 \quad and \quad \frac{\delta^2}{\alpha_*} \to 0 \quad as \quad \delta \to 0. \tag{12}$$

(ii): $\kappa = 2$ with two constants $\underline{c}$ and $\overline{c}$ such that

$$\alpha_* \to 0 \quad and \quad 0 < \underline{c} \le \frac{\delta^2}{\alpha_*} \le \overline{c} < \infty \quad for\ all \quad \delta > 0. \tag{13}$$

and
(iii): $2 < \kappa < 2 + \frac{2}{a}$ with

$$\alpha_* \to 0 \quad and \quad \frac{\delta^2}{\alpha_*} \to \infty \quad as \quad \delta \to 0. \tag{14}$$

If we have $x^\dagger \in X_1$ in contrast to Item (d) of Assumption 1, then, for the convergence of regularized solutions to $x^\dagger$ in Case (i), the nonlinearity condition expressed in Equation (6) is not needed at all provided that Items (a) and (b) of Assumption 1 are fulfilled. Also Item (c) is not necessary there. However to derive Equation (9), $x^\dagger$ must be the uniquely determined penalty-minimizing solution to Equation (1) (cf. [11], Sect. 4.1.2 or alternatively [12,13]). Note that, for $x^\dagger \in X_1$ and parameter choices according to (i), conditions of the type expressed in Equation (6) are only relevant for proving convergence rates.

　　　If regularization parameters are chosen such that Equation (12) is violated as in Cases (ii) and (iii), then even for $x^\dagger \in X_1$ inequalities from the condition expressed in Equation (6) are important. Precisely, Case (iii) seems to require both inequalities of Equation (6) for deriving convergence of regularized solutions to $x^\dagger$. The parameter choice according to Case (ii) with $\alpha_* = \alpha(\delta) \sim \delta^2$ represents for $x^\dagger \in X_1$ the typical conditional stability estimate situation introduced in the seminal paper [14]. There, only the left-hand inequality of condition expressed in Equation (6) is needed for convergence, which then is a consequence of convergence rate results (cf. [15–17] and references therein). However, to derive Equation (9), $x^\dagger$ must be the uniquely determined solution to Equation (1). In the oversmoothing case $x^\dagger \notin X_1$, both inequalities in Equation (6) seem to be indispensable for obtaining convergence; moreover, for all suggested choices of the regularization parameter α, the convergence proofs published by now, all using auxiliary elements, are essentially based on the fact that $x^\dagger$ is an interior point of the domain $\mathcal{D}(F)$. Determining the conditions under which convergence takes place if $\kappa \geq 2 + 2/a$ is chosen in Equation (11) is an open problem.

　　　Now we turn to convergence assertions, provided that the regularization parameter $\alpha > 0$ is selected according to the discrepancy principle expressed in Equation (5) with prescribed constant $C > 1$. The main ideas of the proof are outlined along the lines of [6], Theorem 4.9, where a sequential discrepancy has been considered.

Theorem 2. *Under Assumption 1, let there be, for a sequence $\{\delta_n\}$ of positive noise levels with $\lim_{n\to\infty} \delta_n = 0$ and all admissible noisy data $y^{\delta_n} \in Y$ obeying $\|y^{\delta_n} - y\|_Y \leq \delta_n$, regularization parameters $\alpha_n := \alpha_{discr}(\delta_n, y^{\delta_n}) > 0$, satisfying the discrepancy principle*

$$\|F(x_{\alpha_n}^{\delta_n}) - y^{\delta_n}\|_Y = C \, \delta_n \tag{15}$$

for a prescribed constant $C > 1$. We then have

$$\lim_{n\to\infty} \alpha_n = 0 \tag{16}$$

and convergence as

$$\lim_{n\to\infty} \|x_{\alpha_n}^{\delta_n} - x^\dagger\|_X = 0. \tag{17}$$

Proof. First, we show that Equation (16) always takes place for oversmoothing penalties with $x^\dagger \notin \mathcal{D}(B) = X_1$. To find a contradiction, we assume that $\liminf_{n\to\infty} \alpha_n > 0$. Since $0 \in \mathcal{D}$ as a consequence of Item (a) of Assumption 1, we have $T_{\alpha_n}^{\delta_n}(x_{\alpha_n}^{\delta_n}) \leq T_{\alpha_n}^{\delta_n}(0)$ and thus $\alpha_n \|x_{\alpha_n}^{\delta_n}\|_1^2 \leq \|F(0) - y^{\delta_n}\|_Y^2 \leq 2(\|F(0) - y\|_Y^2 + \delta_n^2)$, which means that $\|x_{\alpha_n}^{\delta_n}\|_1 = \|Bx_{\alpha_n}^{\delta_n}\|_X$ and, by Equation (3), $\|x_{\alpha_n}^{\delta_n}\|_X$ are uniformly bounded from above for all $n \in \mathbb{N}$. We then have for a subsequence the weak convergences in X as $x_{\alpha_{n_k}}^{\delta_{n_k}} \rightharpoonup \tilde{x} \in X$ and $Bx_{\alpha_n}^{\delta_n} \rightharpoonup \tilde{\tilde{x}} \in X$ as $k \to \infty$. Since the operator B is weakly sequentially closed, we therefore obtain $\tilde{x} \in \mathcal{D}(B)$ and for F weakly sequentially continuous (cf. Item (a) of Assumption 1) also $F(x_{\alpha_{n_k}}^{\delta_{n_k}}) \rightharpoonup F(\tilde{x})$. Now, by Equation (15), we easily derive that $\|F(x_{\alpha_{n_k}}^{\delta_{n_k}}) - F(x^\dagger)\|_Y \to 0$ as $k \to \infty$ and consequently $F(\tilde{x}) = F(x^\dagger)$. Thus, the left-hand inequality of Equation (6) (cf. Item (e) of Assumption 1) yields $\tilde{x} = x^\dagger$, which contradicts the assumption $x^\dagger \notin \mathcal{D}(B)$ and proves the property expressed in Equation (16) of the regularization parameter choice.
　　　Secondly, we prove the convergence property expressed in Equation (17). From [6], Lemma 3.2, we have that

$$\|F(x_{\alpha_n}^{\delta_n}) - y^{\delta_n}\|_Y \leq \psi(\alpha_n) \, \alpha_n^{a/(2a+2)} + \delta_n$$

for some function $\psi : [0, \infty) \to [0, \infty)$ satisfying the limit condition $\lim_{\alpha \to 0} \psi(\alpha) = 0$, for sufficiently small $\alpha_n > 0$ and arbitrary $\delta_n > 0$. In combination with $\|F(x_{\alpha_n}^{\delta_n}) - y^{\delta_n}\|_Y = C\,\delta_n$, this yields under the condition expressed in Equation (16), implying $\lim_{n \to \infty} \psi(\alpha_n) = 0$ the estimate

$$\frac{\delta_n}{\alpha_n^{a/(2a+2)}} \le \frac{\psi(\alpha_n)}{C - 1} \to 0 \quad \text{as} \quad n \to \infty.$$

Now Theorem 1 applies. This completes the proof. $\square$

Remark 4. In the case $x^\dagger \in X_1$ of non-oversmoothing penalties, the limit condition expressed in Equation (16) represents a canonical situation for regularized solutions, whereas the non-existence of α_{discr} from Equation (5) and the violation of Equation (16) only occur in exceptional cases. For the sequential variant of the discrepancy principle, the exceptional case $\liminf_{n \to \infty} a_n > 0$ is discussed in [4] in the context of the exact penalization veto introduced there.

4. An Alternative Approach to Prove Hölder Convergence Rates in the Case of Oversmoothing Penalties for an A Priori Parameter Choice of the Regularization Parameter

In this section, we consider order optimal convergence rate results in the case of oversmoothing penalties for an a priori parameter choice of the regularization parameter $\alpha > 0$. Such results have been proven in the paper [9] under the condition $x^\dagger \in X_p = \mathcal{D}(B^p) = \mathcal{R}(B^{-p})$ for $0 < p < 1$, which is a Hölder-type source condition. In that paper, the proof is formulated for the penalty $\|B(x - \bar{x})\|_X^2$ with reference element $\bar{x} \in \mathcal{D}$. This proof has been repeated in the appendix of the paper [10] in the simplified version with $\bar{x} = 0$ and penalty term $\|Bx\|_X^2$, which is also utilized in the present work.

In the following, we present the sketch of an alternative proof for the order optimal Hölder convergence rates under the Hölder-type source condition $x^\dagger \in X_p$ for $0 < p < 1$. This alternative approach is based on error estimates that have been verified in [6] for showing convergence of the regularized solutions x_α^δ to $x^\dagger$ and for proving low order (e.g., logarithmic) convergence rates under corresponding low order source conditions. By one novel idea outlined below, the results from [6] can be extended to prove Hölder convergence rates, too.

For the subsequent investigations, we complement Assumption 1 with an assumption that specifies the smoothness of the solution $x^\dagger$:

Assumption 2. *There are $0 < p < 1$ and $w \in X$ such that*

$$x^\dagger = B^{-p}w \in X_p. \tag{18}$$

Theorem 3. *Under Assumptions 1 and 2, we have for the a priori parameter choice*

$$\alpha_* = \alpha(\delta) \sim \delta^{\frac{2(a+1)}{a+p}} \tag{19}$$

the convergence rate

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

Proof. We give only a sketch of a proof for this theorem, presupposing the results of the recent paper [6]. Precisely, we outline only the points, where we amend and complement the results of [6] in order to extend [6], Theorem 5.3, to the case of appropriate power-type functions φ.

Auxiliary elements $z_\alpha \in \mathcal{D}(B)$, which are for all $\alpha > 0$ the uniquely determined minimizers of the artificial Tikhonov functional $T_{\alpha,a} := \|x - x^\dagger\|_{-a}^2 + \alpha\|Bx\|_X^2$, represent, in combination with the moment inequality, the essential tool for the proof. By introducing the self-adjoint and positive

semi-definite bounded linear operator $G := B^{-(2a+2)} : X \to X$, we can verify these elements in an explicit manner as

$$z_\alpha = G(G + \alpha I)^{-1} x^\dagger = x^\dagger - \alpha (G + \alpha I)^{-1} x^\dagger,$$

which implies that, for all $\alpha > 0$,

$$z_\alpha - x^\dagger = -\alpha (G + \alpha I)^{-1} x^\dagger,$$

$$B^{-a}(z_\alpha - x^\dagger) = -G^{a/(2a+2)} [\alpha (G + \alpha I)^{-1} x^\dagger],$$

and

$$B(z_\alpha - x^\dagger) = G^{(2a+1)/(2a+2)} [\alpha (G + \alpha I)^{-1} x^\dagger].$$

According to [6], Lemma 3.1, we then have the functions $f_1(\alpha) = o(1)$, $f_2(\alpha) = o(1)$ and $f_3(\alpha) = o(1)$ as $\alpha \to 0$ introduced there, which can be found in our notation from the representations $\|z_\alpha - x^\dagger\|_X = f_1(\alpha)$, $\|B^{-a}(z_\alpha - x^\dagger)\|_X = f_2(\alpha) \alpha^{a/(2a+2)}$, and $\|B(z_\alpha - x^\dagger)\|_X = f_3(\alpha) \alpha^{-1/(2a+2)}$. Under the source condition expressed in Equation (18), which attains the form $x^\dagger = G^{p/(2a+2)} w$ with some source element $w \in X$, we derive in detail the formulas

$$f_1(\alpha) = \|G^{p/(2a+2)} [\alpha (G + \alpha I)^{-1} w]\|_X = \mathcal{O}\left(\alpha^{\frac{p}{2a+2}}\right) \quad \text{as} \quad \alpha \to 0, \tag{21}$$

$$f_2(\alpha) = \alpha^{-a/(2a+2)} \|G^{(a+p)/(2a+2)} [\alpha (G + \alpha I)^{-1} w]\|_X = \mathcal{O}\left(\alpha^{\frac{p}{2a+2}}\right) \quad \text{as} \quad \alpha \to 0, \tag{22}$$

and

$$f_3(\alpha) = \alpha^{-(2a+1)/(2a+2)} \|G^{(2a+1+p)/(2a+2)} [\alpha (G + \alpha I)^{-1} w]\|_X = \mathcal{O}\left(\alpha^{\frac{p}{2a+2}}\right) \quad \text{as} \quad \alpha \to 0. \tag{23}$$

The asymptotics $\mathcal{O}(\alpha^{p/(2a+2)})$ in Equations (21)–(23) is a consequence of the properties $\|G(G + \alpha I)^{-1}\| \leq 1$, $\|(G + \alpha I)^{-1}\| \leq 1/\alpha$, which yield by exploiting the moment inequality (cf. [7], Formula (2.49))

$$\|G^\theta (G + \alpha I)^{-1}\| \leq \|G(G + \alpha I)^{-1}\|^\theta \|(G + \alpha I)^{-1}\|^{1-\theta} \leq \alpha^{\theta-1}, \quad \text{for} \quad \alpha > 0 \quad \text{and} \quad 0 \leq \theta \leq 1 \tag{24}$$

for the self-adjoint and positive semi-definite operator G. Here, $\|\cdot\|$ denotes the operator norm in the space of bounded linear operators mapping in X. In Equation (21), the inequality expressed in Equation (24) is applied with $\theta = p/(2a + 2)$, it is applied with $\theta = (a + p)/(2a + 2)$ in Equation (22), and it is applied with $\theta = (2a + p + 1)/(2a + 2)$ in Equation (23), taking into account that all three θ-values are smaller than one. These are the new ideas of the present proof.

Because the function $f_9(\alpha)$ in [6], Formula (3.12), is found by linear combination and maximum-building of the functions $f_1(\alpha)$, $f_2(\alpha)$, and $f_3(\alpha)$, we derive here $f_9(\alpha) \sim \alpha^{p/(2a+2)}$ along the lines of Section 3 in [6] and consequently the error estimate

$$\|x_\alpha^\delta - x^\dagger\|_X \leq \underline{K} \alpha^{p/(2a+2)} + \overline{K} \frac{\delta}{\alpha^{a/(2a+2)}} \tag{25}$$

with constants $\underline{K}, \overline{K} > 0$, which is valid for all $\delta > 0$ and a sufficiently small $\alpha > 0$. Such a restriction to a sufficiently small $\alpha > 0$ is due to the fact that z_α has to belong to $\mathcal{D}(F)$ in order to apply the inequality chain expressed in Equation (6), but this is the case for small α, since $x^\dagger$ is assumed to be an interior point of $\mathcal{D}(F)$. Under the a priori parameter choice expressed in Equation (19), we immediately obtain the convergence rate expressed in Equation (20) from the error estimate expressed in Equation (25). This completes the sketch of the proof of the theorem. □

Remark 5. Obviously, the a priori parameter choice expressed in Equation (19) satisfies the sufficient condition expressed in Equation (10) for the convergence of regularized solutions from Theorem 1.

More precisely, taking into account Example 1, the choice expressed in Equation (19) has the form of Equation (11) with $\kappa = \frac{2(a+1)}{a+p}$, which for $0 < p < 1$ yields $2 < \kappa < 2 + \frac{2}{a}$ and belongs to Case (iii), where the quotient $\frac{\delta^2}{\alpha_*}$ tends toward infinity as $\delta \to 0$. We mention that the choice expressed in Equation (19) coincides with the choice in [8] suggested by Natterer, who proved the order optimal convergence rate expressed in Equation (20) for linear ill-posed operator equations. For the nonlinear operator expressed in Equation (1), in [3], the convergence rate expressed in Equation (20) has also been proven for the a posteriori parameter choice $\alpha_{discr} = \alpha(\delta, y^\delta)$ from Equation (5). However, by now, there are no analytical results about the α_{discr}-asymptotics of the discrepancy principle as δ tends toward zero. The numerical experiments in the subsequent sections will provide some hints that the hypothesis $\alpha_{discr} \sim \delta^{\frac{2(a+1)}{a+p}}$ does not have to be rejected.

5. Model Problem and Appropriate Hilbert Scale

In the following, we introduce an example for a nonlinear inverse operator expressed in Equation (1) together with an appropriate Hilbert scale, for which we will investigate the analytic results from the previous section numerically, following up on [10]. The well-known scale of Hilbert-type Sobolev spaces $H^p(0,1)$ with integer values of $p \geq 0$ consists of functions whose p-th derivative is still in $L^2(0,1)$. For positive indices p, the spaces can be defined by using an interpolation argument, and for general real parameters of $p \in \mathbb{R}$ the norms of $H^p(0,1)$ can be defined by using the Fourier transform $\hat{x}$ of the function x as

$$\|x\|^2_{H^p(0,1)} := \int_{\mathbb{R}} (1 + |\xi|^2)^p \, |\hat{x}(\xi)|^2 \, d\xi \tag{26}$$

(cf. [18]). The Sobolev scales do not constitute a Hilbert scale in the strict sense, but for each $0 < p^* < \infty$ there is an operator $B : L^2(0,1) \to L^2(0,1)$ such that $\{X_p\}_{0 \leq p \leq p^*}$ is a Hilbert scale (see [19]). In order to form a full Hilbert scale for arbitrary real values of p, boundary values conditions need to be imposed.

Hilbert scale. To generate a full Hilbert scale $\{X_\tau\}_{\tau \in \mathbb{R}}$, we exploit the simple integration operator

$$[Jh](t) := \int_0^t h(\tau)d\tau \quad (0 \leq t \leq 1) \tag{27}$$

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

$$B := (J^*J)^{-1/2}. \tag{28}$$

Using the Riemann-Liouville fractional integral operator J^p and its adjoint $(J^*)^p = (J^p)^*$ for $0 < p \leq 1$, we receive

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

(cf. [20–23]); hence, by [20], Lemma 8, the explicit structure

$$X_p = \begin{cases} H^p(0,1) & \text{for} \quad 0 < p < \frac{1}{2} \\ \{x \in H^{\frac{1}{2}}(0,1) : \int\limits_0^1 \frac{|x(t)|^2}{1-t} dt < \infty\} & \text{for} \quad p = \frac{1}{2} \\ \{x \in H^p(0,1) : x(1) = 0\} & \text{for} \quad \frac{1}{2} < p \leq 1 \end{cases} \tag{29}$$

Further boundary conditions have to be incorporated for higher Sobolev indices p.

Model problem. The exponential growth model of this example was discussed in early literature (cf., e.g., [24], Section 3.1). More details and properties can be found in [25] and in the appendix of [3]. To identify the time dependent growth rate $x(t)$ $(0 \leq t \leq T)$ of a population, we use observations

$y(t)$ $(0 \le t \le T)$ of the time-dependent size of the population with initial size $y(0) = y_0 > 0$, where the O.D.E. initial value problem

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

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

$$[F(x)](t) = y_0 \exp\left(\int_0^t x(\tau)d\tau\right) \qquad (0 \le t \le 1), \tag{30}$$

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

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

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

$$\|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\|_Y \le \hat{K}\,\|F(x) - F(x^\dagger)\|_Y\,\|x - x^\dagger\|_X \tag{31}$$

is valid. This in turn guarantees that a tangential cone condition

$$\|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\|_Y \le \eta\|F(x) - F(x^\dagger)\|_Y \tag{32}$$

holds with some $0 < \eta < 1$ in $\mathcal{D}(F) = \overline{\mathcal{B}_r(x^\dagger)}$ for a sufficiently small r > 0 (cf. [3], Example A.2), where $\overline{\mathcal{B}_r(x^\dagger)}$ denotes a closed ball around $x^\dagger$ with radius r. According to the construction of the Hilbert scale $\{X_\tau\}_{\tau \in \mathbb{R}}$ generated by the operator B in Equation (28), and due to $0 < \underline{c} \le F(x^\dagger) \le \overline{c} \le \infty$ as a consequence of Equation (30), we receive from [3], Proposition A.4 that the inequality chain expressed in Equation (6) holds with a = 1 in this example.

6. Numerical Case Studies

In this section, numerical evidence for the behavior of the regularized solutions x_α^δ of the model problem introduced in Section 5 is provided. In Section 6.1, numerical experiments with a focus on the discrepancy principle are conducted using exact solutions for low order Hölder-type smoothness $x^\dagger \in X_p$ with $0 < p < 1/2$, while the focus of the recent paper [10] was on results for $p = 1/2$ and larger values p. The essential point of Section 6.2 is the comparison of results obtained by the discrepancy principle with those calculated by a priori choices expressed in Equation (11) of the regularization parameter α.

6.1. Case Studies for Exact Solutions with Low Order Hölder Smoothness $p < \frac{1}{2}$

In our first series of experiments, we investigated the interplay between the value $p \in (0, \frac{1}{2})$, the decay rates of the regularization parameter α_{discr} with respect to the noise level δ for different values p as δ tends toward zero, and the corresponding rates of the error of regularized solutions x_α^δ. Therefore, we turn to exact solutions of the form $x^\dagger(t) = ct^{-\beta}$ $(0 < t \le 1)$ with $\beta \in (0, 1/2)$. These functions $x^\dagger$ do not belong to the Sobolev space $H^p(0,1)$ with fractional order p if $1/2 - \beta < p$ (see for example [26], p. 422). This allows us to study the behavior of the regularized solutions for exact solutions with low order Hölder-type solution smoothness. For the numerical simulations, we therefore assume that $1/2 - \beta$ is at least approximately the smoothness of the exact solution $x^\dagger$.

To confirm our theoretical findings, we solve Equation (4) after discretization using the trapezoidal rule for the integral, the MATLAB®-function `fmincon`. We would also like to point out the difficulties associated with the numerical treatment of functions of this particular type. Obviously the pole occurring at zero is source force for the low smoothness of the exact solution and needs to be captured

accordingly. After multiple different approaches, equidistant discretization with the first discretization point very close to zero was proven to be very successful. Typically, a discretization level $N = 200$ is used. To the simulated data $y = F(x^\dagger)$, we added random noise for which we prescribe the *relative error* $\bar{\delta}$ such that $\|y - y^\delta\| = \bar{\delta}\|y\|$; i.e., we have Equation (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 enforce the boundary condition $x(1) = 0$. The regularization parameter α in these series of experiments is chosen as $\alpha_{discr} = \alpha(\delta, y^\delta)$ using, with some prescribed multiplier $C > 1$, the discrepancy principle

$$\delta \le \|F(x^\delta_{\alpha_{discr}}) - y^\delta\|_Y \le C\delta, \tag{33}$$

which approximates α_{discr} from Equation (5). Unless otherwise noted, C = 1.3 was used. From the case studies in [10], we can conjecture, but have no stringent proof, that the α-rate of the discrepancy principle does not systematically deviate from the a priori rate expressed in Equation (19), which for $a = 1$ attains the form

$$\alpha(\delta) \sim \delta^{\frac{4}{p+1}}. \tag{34}$$

This α-rate already occurred in Natterer's paper [8] for linear problems, and occurs in the case of oversmoothing penalties. We should in our numerical experiments be able to observe the order optimal convergence rate, which is for $a = 1$

$$\|x^\delta_{\alpha(\delta)} - x^\dagger\|_X = \mathcal{O}(\delta^{\frac{p}{p+1}}) \qquad \text{as} \qquad \delta \to 0. \tag{35}$$

This convergence rate was proven for the a priori parameter choice expressed in Equation (19) as well as for the discrepancy principle expressed in Equation (5) in [9] and [3], respectively.

As the exact solutions $x^\dagger$ are known, we can compute the regularization errors $\|x^\delta_\alpha - x^\dagger\|_X$. Interpreting these errors as a function of δ justifies a regression for the convergence rates according to

$$\|x^\delta_\alpha - x^\dagger\|_X \le c_x \delta^{\kappa_x}. \tag{36}$$

The α-rates are then computed in a similar fashion using

$$\alpha = \alpha(\delta) = c_\alpha \delta^{\kappa_\alpha}. \tag{37}$$

Both exponents κ_x and κ_α and the corresponding multipliers c_x and c_α, all obtained by using a least squares regression based on samples for varying δ, are displayed for different values β in Table 1. As we know, the convergence rate expressed in Equation (35), we can estimate the smoothness p by the the formula $p_{est} := \frac{\kappa_x}{1-\kappa_x}$. The far right column of Table 1 displays the quotient $\frac{4}{p_{est}+1}$ estimating the exponent in Equation (34), which can be compared with the κ_α-values in the second to right column obtained by regression from a data sample. By comparing the far right column and the second to right column of Table 1, we can state that the asymptotics of α_{discr} as $\delta \to 0$ seems to be approximately the same as for the optimal a priori parameter choice expressed in Equation (34). Such observation was already made for larger values p in [10] for the same model problem.

Table 1. Numerically computed results for discrepancy principle expressed in Equation (33) and $x^\dagger(t) = ct^{-\beta}$ ($0 < t \le 1$) yielding by regression multipliers and exponents of regularization error expressed in Equation (36) and α-rates expressed in Equation (37) for various values $0 < \beta < 0.5$.

β	c_x	κ_x	$p_{est} = \frac{\kappa_x}{1-\kappa_x}$	c_α	κ_α	$\frac{4}{p_{est}+1}$
0.1	0.2014	0.2497	0.3328	111.5089	2.4738	3.0011
0.2	0.4504	0.2383	0.3129	43.9993	2.7924	3.0467
0.3	0.7428	0.1970	0.2453	19.2617	3.1172	3.2120
0.4	1.3294	0.1831	0.2241	2.8548	3.1619	3.2677
0.45	1.5829	0.1422	0.1657	5.1283	3.6665	3.4314

Figure 1 illustrates results from Table 1 for $x^\dagger(t) = ct^{-\beta}$ $(0 < t \leq 1)$ characterizing with varying $\beta \in (0, 1/2)$ different smoothness levels of the solution. Since we have an oversmoothing penalty for all such β, the κ_α-values lie between 2 and $2 + \frac{2}{a} = 4$ (cf. the κ-interval (iii) in Example 1). Additionally, the border lines for $\kappa_\alpha = 2$ and $\kappa_\alpha = 2 + \frac{2}{a}$ are displayed taking into account that [6] guarantees convergence of the regularized solution x_α^δ to the exact solution $x^\dagger$ as $\delta \to 0$ for a priori choices in the sense of Equation (19) whenever $2 \leq \kappa_\alpha < 2 + \frac{2}{a}$. It becomes evident that the α-rates resulting from the discrepancy principle also lie between those bounds.

Figures 2 and 3 give some more insight into the situation for the special case $\beta = 0.2$, which approximately corresponds with the smoothness $x^\dagger \in X_{0.31}$. In Figure 2 (left), the realized errors $\|x_\alpha^\delta - x^\dagger\|_X$ are visualized for a discrete set of noise levels and compared with the associated regression line in a double-logarithmic scale. It becomes evident that the approximation using Hölder rates is highly accurate. The right image of Figure 2 visualizes the behavior of $\frac{\delta^2}{\alpha_{discr}}$ for various noise levels on a logarithmic scale. The tendency that $\frac{\delta^2}{\alpha_{discr}} \to \infty$ as $\delta \to 0$ seems to be convincing. Figure 3 (left) displays the realized α_{discr}-values for this particular situation together with the best approximating regression line according to Equation (37). We see again a very good fit for this type of approximation. The right subfigure shows the exact and regularized solution for $\delta = 10^{-3.5}$. The excellent fit of the regularized solution confirms our confidence in the numerical implementation, especially considering the problems associated with this type of exact solution.
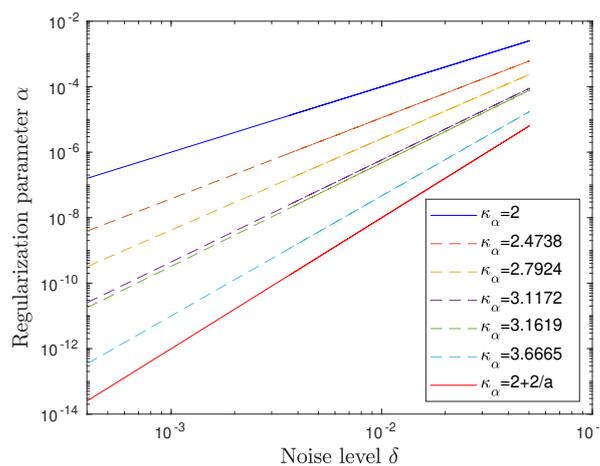

Figure 1. Regression lines for decay rates of α_{discr} for $\delta \to 0$ and different values β from Table 1 on a double-logarithmic scale.
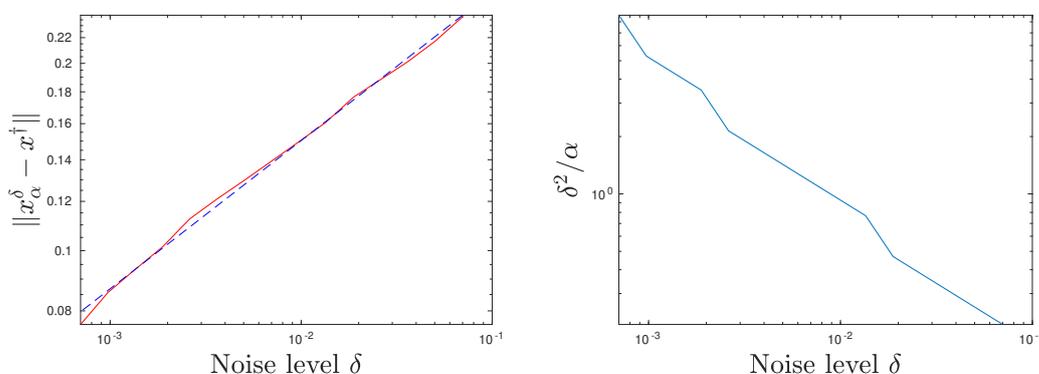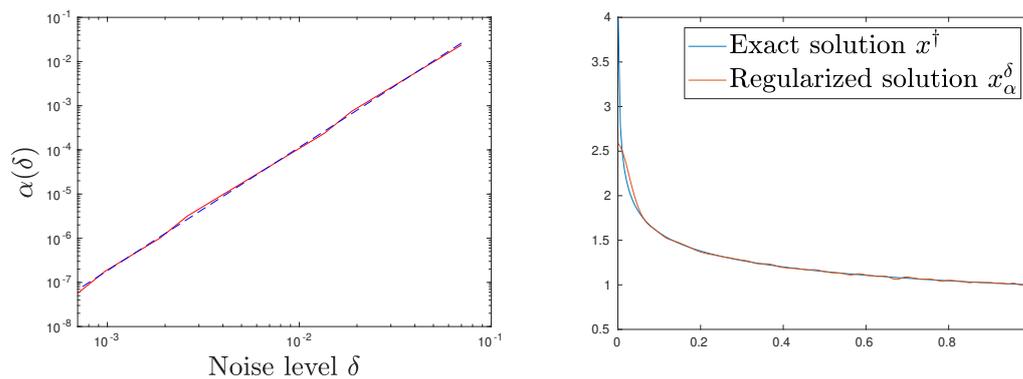

Figure 2. Approximation error $\|x_\alpha^\delta - x^\dagger\|_X$ in red and approximate rate in blue/dashed (**left**) and $\frac{\delta^2}{\alpha_{discr}}$ (**right**), both depending on various δ on a log-log scale.

Figure 3. $\alpha_{discr}(\delta, y^\delta)$ in red for various δ and best approximating regression line in blue/dashed on a log-log scale (**left**). Regularized (red) compared with exact solution (blue) for $\delta = 10^{-3.5}$ (**right**).

6.2. A Comparison with Results from A Priori Parameter Choices

The question of whether the a posteriori choices using the discrepancy principle or appropriate a priori choices according to Equation (19) yield better results is of interest. The influence of the constant c_α when using a priori choice according to Equation (37) remains especially unclear. To numerically investigate this, we remain in the setting of Section 6.1; i.e., we consider $x^\dagger(t) = ct^{-\beta}$ $(0 < t \le 1)$ as an exact solution. Figure 4 illuminates this situation for $\beta = 0.2$. The error $\|x_\alpha^\delta - x^\dagger\|_X$ is plotted for various constants c_α, where we use $\alpha_* = \alpha(\delta) = c_\alpha \delta^{\frac{4}{1+p}}$. The error curve shows a clear minimum, which is connected with smaller values $\|x_\alpha^\delta - x^\dagger\|_X$ compared with those obtained by exploiting the discrepancy principle with $C = 1.4$ and $C = 1.6$. It is completely unclear how to find suitable multipliers c_α in practice, whereas the discrepancy principle can always be applied as a robust parameter choice rule for practical applications.
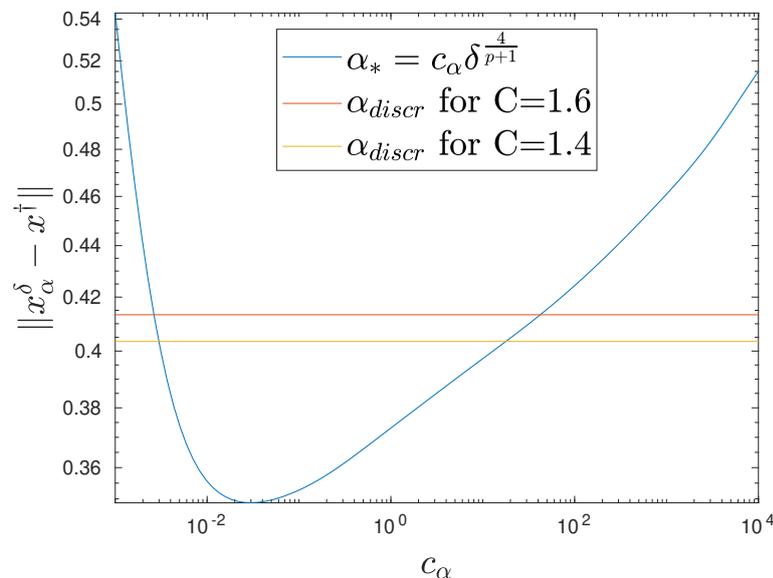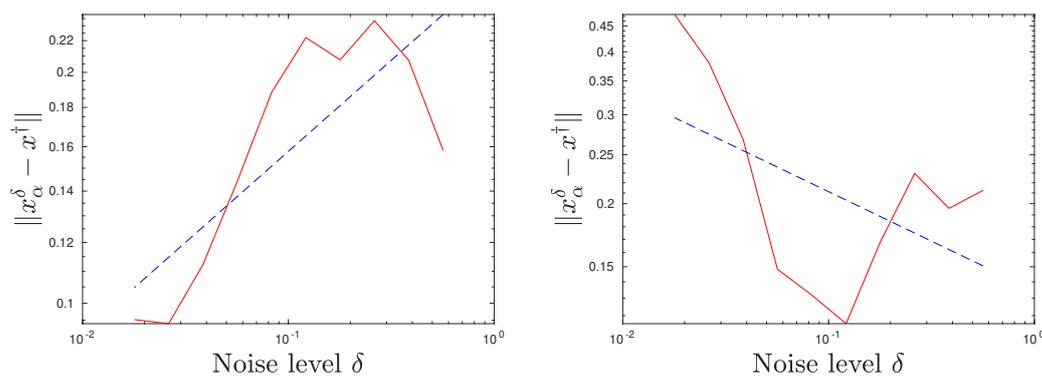

Figure 4. Regularization error $\|x_\alpha^\delta - x^\dagger\|_X$ using the a priori rate expressed in Equation (19) implemented in the sense of Equation (37) depending on various constants c_α. Comparison with the error occurring for the discrepancy principle with $C = 1.4$ (orange) or $C = 1.6$ (red) and with noise level $\delta = 10^{-2.5}$.

We complete our numerical experiments on a priori choices of the regularization parameter with Table 2 and Figure 5, where we list and illustrate the best regression exponents κ_x according to the error norm estimate expressed in Equation (36) for different exponents κ_α in the a priori parameter choice expressed in Equation (37). In this case study, we used the exact solution $x(t) = 1$ ($0 \leq t \leq 1$) with the higher smoothness $p = 0.5$. For the a priori parameter choice expressed in Equation (37) with varying exponents κ_α, the factor $c_\alpha = 1$ has been fixed. The discretization level $N = 1000$ was considered.

Table 2. Error rates κ_x by regression for $x^\dagger \equiv 1$ and varying a priori exponents κ_α.

a Priori α-Rate κ_α	2	2.5	2.66	3	3.5
convergence rate κ_x	0.2763	0.3304	0.3479	0.3686	0.3114

As expected, Table 2 indicates that maximal error rates occur if κ_α is close to the optimal value $\frac{4}{1+p}$. These rates also correspond with the order optimal rates according to Equation (20). For smaller exponents κ_α, the error rates are falling, and for large exponents $\kappa_\alpha \geq 2 + \frac{2}{a} = 4$, the convergence seems to degenerate. This is visualized in Figure 5: For $\kappa_\alpha = 3.5$, convergence still takes place, whereas for $\kappa_\alpha = 5.5$ convergence cannot be observed anymore.

Figure 5. Regularization error $\|x_\alpha^\delta - x^\dagger\|_X$ and regression line for different noise levels δ on a log-log scale. $x^\dagger(t) = 1$, a priori parameter choice according to Equation (37) with $\kappa_\alpha = 3.5$ (**left**) and $\kappa_\alpha = 5.5$ (**right**).

Remark 6. As an alternative a posteriori approach for choosing the regularization parameter α, one could also consider the balancing (Lepskiǐ) principle (cf., e.g., [27,28]). In [29], this principle is adapted to the Hilbert scale setting, but not with respect to oversmoothing penalties. In future work, we can discuss this missing facet and perform numerical experiments for the balancing principle in the case of oversmoothing penalties.

Author Contributions: Formal analysis, B. H.; Investigation, B.H. and C.H.; Software, C.H.; Supervision, B.H.; Visualization, C.H. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hofmann, B.; Scherzer, O. Factors influencing the ill-posedness of nonlinear problems. *Inverse Probl.* **1994**, *10*, 1277–1297. [CrossRef]
2. Morozov, V.A. *Methods for Solving Incorrectly Posed Problems*; Springer: New York, NY, USA, 1984.

3.   Hofmann, B.; Mathé, P. Tikhonov regularization with oversmoothing penalty for non-linear ill-posed problems in Hilbert scales. *Inverse Probl.* **2018**, *34*, 015007. [CrossRef]

4.   Anzengruber, S.W.; Hofmann, B.; Mathé, P. Regularization properties of the sequential discrepancy principle for Tikhonov regularization in Banach spaces. *Appl. Anal.* **2014**, *93*, 1382–1400. [CrossRef]

5.   Anzengruber, S.W.; Ramlau, R. Morozov's discrepancy principle for Tikhonov-type functionals with nonlinear operators. *Inverse Probl.* **2010**, *26*, 025001. [CrossRef]

6.   Hofmann, B.; Plato, R. Convergence results and low order rates for nonlinear Tikhonov regularization with oversmoothing penalty term. *Electron. Trans. Numer. Anal.* **2020**, *93*.

7.   Engl, H.W.; Hanke, M.; Neubauer, A. *Regularization of Inverse Problems*; Volume 375 of Mathematics and Its Applications; Kluwer Academic Publishers Group: Dordrecht, The Netherlands, 1996.

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

9.   Hofmann, B.; Mathé, P. A priori parameter choice in Tikhonov regularization with oversmoothing penalty for non-linear ill-posed problems. In *Springer Proceedings in Mathematics & Statistics*; Cheng, J., Lu, S., Yamamoto, M., Eds.; Springer: Singapore, 2020; pp. 169–176.

10. Gerth, D.; Hofmann, B.; Hofmann, C. Case studies and a pitfall for nonlinear variational regularization under conditional stability. In *Springer Proceedings in Mathematics & Statistics*; Cheng, J., Lu, S., Yamamoto, M., Eds.; Springer: Singapore, 2020; pp. 177–203.

11. Schuster, T.; Kaltenbacher, B.; Hofmann, B.; Kazimierski, K.S. *Regularization Methods in Banach Spaces*; Volume 10 of Radon Series on Computational and Applied Mathematics; Walter de Gruyter: Berlin, Germany; Boston, MA, USA, 2012.

12. Scherzer, O.; Grasmair, M.; Grossauer, H.; Haltmeier, M.; Lenzen, F. *Variational Methods in Imaging*; Volume 167 of Applied Mathematical Sciences; Springer: New York, NY, USA, 2009.

13. Tikhonov, A.N.; Leonov, A.S.; Yagola, A.G. *Nonlinear Ill-Posed Problems*; Chapman & Hall: London, UK; New York, NY, USA, 1998; Volume 1.

14. Cheng, J.; Yamamoto, M. One new strategy for a priori choice of regularizing parameters in Tikhonov's regularization. *Inverse Probl.* **2000**, *16*, L31–L38. [CrossRef]

15. Egger, H.; Hofmann, B. Tikhonov regularization in Hilbert scales under conditional stability assumptions. *Inverse Probl.* **2018**, *34*, 115015. [CrossRef]

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

17. Tautenhahn, U. On a general regularization scheme for nonlinear ill-posed problems II: Regularization in Hilbert scales. *Inverse Probl.* **1998**, *14*, 1607–1616. [CrossRef]

18. Adams, R.A.; Fournier, J.F.J. *Sobolev Spaces*; Elsevier/Academic Press: Amsterdam, The Netherlands, 2003.

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

20. Gorenflo, R.; Yamamoto, M. Operator-theoretic treatment of linear Abel integral equations of first kind. *Jpn. J. Ind. Appl. Math.* **1999**, *16*, 137–161. [CrossRef]

21. Gorenflo, R.; Luchko, Y.; Yamamoto, M. Time-fractional diffusion equation in the fractional Sobolev spaces. *Fract. Calc. Appl. Anal.* **2015**, *18*, 799–820. [CrossRef]

22. Hofmann, B.; Kaltenbacher, B.; Resmerita, E. Lavrentiev's regularization method in Hilbert spaces revisited. *Inverse Probl.* **2016**, *10*, 741–764. [CrossRef]

23. Plato, R.; Hofmann, B.; Mathé, P. Optimal rates for Lavrentiev regularization with adjoint source conditions. *Math. Comput.* **2018**, *87*, 785–801. [CrossRef]

24. Groetsch, C.W. Inverse Problems in the Mathematical Sciences. In *Vieweg Mathematics for Scientists and Engineers*; Vieweg+Teubner Verlag: Wiesbaden, Germany, 1993.

25. Hofmann, B. A local stability analysis of nonlinear inverse problems. In *Inverse Problems in Engineering—Theory and Practice*; The American Society of Mechanical Engineers: New York, NY, USA, 1998; pp. 313–320.

26. Fleischer, G.; Hofmann, B. On inversion rates for the autoconvolution equation. *Inverse Probl.* **1996**, *12*, 419–435. [CrossRef]

27. Mathé, P. The Lepskiĭ principle revisited. *Inverse Probl.* **2006**, *22*, L11–L15. [CrossRef]

28. Lu, S.; Pereverzev, S.V.; Ramlau, R. An analysis of Tikhonov regularization for nonlinear ill-posed problems under a general smoothness assumption. *Inverse Probl.* **2007**, *23*, 217–230. [CrossRef]

29. Pricop-Jeckstadt, M. Nonlinear Tikhonov regularization in Hilbert scales with balancing principle tuning parameter in statistical inverse problems. *Inverse Probl. Sci. Eng.* **2019**, *27*, 205–236. [CrossRef]