[1]

NONLINEAR TIKHONOV REGULARIZATION IN HILBERT SCALES WITH OVERSMOOTHING PENALTY: INSPECTING BALANCING PRINCIPLES

B. HOFMANN, C. HOFMANN, P. MATHÉ, AND R. PLATO

Abstract. The analysis of Tikhonov regularization for nonlinear ill-posed equations with smoothness promoting penalties is an important topic in inverse problem theory. With focus on Hilbert scale models, the case of oversmoothing penalties, i.e., when the penalty takes an infinite value at the true solution gained increasing interest. The considered nonlinearity structure is as in the study B. Hofmann and P. Mathé. *Tikhonov regularization with oversmoothing penalty for non-linear ill-posed problems in Hilbert scales*. Inverse Problems, 2018. Such analysis can address two fundamental questions. When is it possible to achieve order optimal reconstruction? How to select the regularization parameter? The present study complements previous ones by two main facets. First, an error decomposition into a smoothness dependent and a (smoothness independent) noise propagation term is derived, covering a large range of smoothness conditions. Secondly, parameter selection by balancing principles is presented. A detailed discussion, covering some history and variations of the parameter choice by balancing shows under which conditions such balancing principles yield order optimal reconstruction. A numerical case study, based on some exponential growth model, provides additional insights.

1. Introduction

In the past years, a new facet has found interest in the theory of inverse problems. When considering variational (Tikhonov-type) regularization for the stable approximate solution of *ill-posed* operator equations

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

in Hilbert spaces the treatment of *oversmoothing penalties* gained attention. The current state of the art concerning that facet shows gaps. The present study aims at closing some of them. Specifically, we analyze variants of the balancing principle as parameter choice for the regularization parameter. Moreover, we would like to illustrate the theory by numerical case studies.

[1]Dedicated to Zuhair Nashed, our esteemed colleague and outstanding Professor, doyen of operator theory and regularization theory for inverse and ill-posed problems

Date: December 22, 2020.

The forward operator $F : \mathcal{D}(F) \subseteq X \to Y$, which is preferably assumed to be *nonlinear*, maps between the separable infinite dimensional Hilbert spaces X and Y and possesses the convex and closed subset $\mathcal{D}(F)$ of X as domain of definition. In the sequel, let $x^\dagger \in \mathcal{D}(F)$ be a solution to equation (1) for given exact right-hand side $y = F(x^\dagger) \in Y$. We throughout assume that $x^\dagger \in \mathrm{int}(\mathcal{D}(F))$, which means that the solution belongs to the interior of the domain $\mathcal{D}(F)$ of the operator F. Given the noise level $\delta \geq 0$, we consider the deterministic noise model

$$\|y - y^\delta\|_Y \leq \delta,$$

which means that instead of y, only perturbed data $y^\delta = y + \delta\,\xi \in Y$ with $\|\xi\|_Y \leq 1$ are available.

The equation (1) is ill-posed at least locally at $x^\dagger$, and finding stable solutions requires some regularization. We apply Tikhonov regularization with quadratic misfit and penalty functionals in a *Hilbert scale* setting.

1.1. **Hilbert scales with respect to an unbounded operator.**

The Hilbert scale is generated by some densely defined, unbounded and self-adjoint linear operator $B\colon \mathcal{D}(B) \subset X \to X$ with domain $\mathcal{D}(B)$. This operator B is assumed to be strictly positive such that we have for some $\underline{m} > 0$

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

The Hilbert scale $\{X_\tau\}_{\tau\in\mathbb{R}}$, generated by B, is characterized by the formulas $X_\tau = \mathcal{D}(B^\tau)$ for $\tau > 0$, $X_\tau = X$ for $\tau \leq 0$ and $\|x\|_\tau := \|B^\tau x\|_X$ for $\tau \in \mathbb{R}$. We do not need in our setting the topological completion of the spaces $X_\tau = X$, for $\tau < 0$, with respect to the norm $\|\cdot\|_\tau$.

1.2. **Tikhonov regularization with smoothness promoting penalty.**

The operator B is used for Tikhonov regularization in the corresponding functional

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

where $\|F(x) - y^\delta\|^2$ characterizes the misfit or fidelity term, and $\overline{x} \in X$ is an initial guess occurring in the penalty functional $\|B(x - \overline{x})\|_X^2$. Throughout this paper we suppose that $\overline{x} \in \mathcal{D}$. Given a regularization parameter $\alpha > 0$ the corresponding regularized solutions x_α^δ to $x^\dagger$ are obtained as the minimizers of the Tikhonov functional T_α^δ on the set $\mathcal{D}$. By definition of the Hilbert scale we have that $\|B(x - \overline{x})\|_X^2 = \|x - \overline{x}\|_1^2$ and consequently $x_\alpha^\delta \in \mathcal{D}(B) = X_1$ for all data $y^\delta \in Y$ and $\alpha > 0$. In order to ensure *existence* and *stability* of the regularized solutions x_α^δ for all $\alpha > 0$ (cf. [8, § 3], [25, Section 3.2] and [26, Section 4.1.1]), we additionally suppose that the forward operator F is weakly sequentially continuous.

Our focus will be on oversmoothing penalties, when $x^\dagger \notin \mathcal{D}(B)$ and hence $T_\alpha^\delta(x^\dagger) = \infty$. In this case, the regularizing property $T_\alpha^\delta(x_\alpha^\delta) \leq T_\alpha^\delta(x^\dagger)$, which often is a basic tool for obtaining error estimates does not yield consequences.

1.3. **State of the art.** The seminal study for Tikhonov regularization, including the case of oversmoothing penalty, and for linear ill-posed operator equations, was published by Natterer in 1984, cf. [20]. For linear bounded operator $F = A : X \to Y$, Natterer used for this study a two-sided inequality chain

$$c_a \|x\|_{-a} \leq \|Ax\|_Y \leq C_a \|x\|_{-a} \quad \text{for all} \quad x \in X,$$

with constants $0 < c_a \leq C_a < \infty$ and a *degree of ill-posedness* $a > 0$. Here, we adapt this to the nonlinear mapping $F : \mathcal{D}(F) \subset X \to Y$ as

$$c_a \|x - x^\dagger\|_{-a} \leq \|F(x) - F(x^\dagger)\|_Y \leq C_a \|x - x^\dagger\|_{-a} \qquad \text{for all} \quad x \in \mathcal{D}, \tag{3}$$

and rely upon (3) as an intrinsic *nonlinearity condition* for the mapping F under consideration.

This specific nonlinearity condition was first used within the present context in [9]. It was shown that a discrepancy principle yields optimal order convergence under the power type smoothness assumption that $x^\dagger \in X_p,\ 0 < p < 1$. One major tool was to use certain *auxiliary elements* $\hat{x}_\alpha$ of proximal type, which are minimizers of the functional

$$\widehat{T}_\alpha(x) := \|x - x^\dagger\|_{-a}^2 + \alpha \|x - \overline{x}\|_1^2, \quad x \in X, \tag{4}$$

and hence belong to $X_1 = \mathcal{D}(B)$. This study was complemented in [2] within the proceedings *Inverse Problems and Related Topics: Shanghai, China, October 12–14, 2018*, Springer, 2020, by case studies showing intrinsic problems when using oversmoothing penalties. The same issue contains results on *a priori* parameter choice $\alpha_* \sim \delta^{(2a+2)/(a+p)}$, i.e., when the smoothness p is assumed to be known, see [10].

However, the first decomposition of the error into a smoothness dependent increasing term (as a function of α tending to zero as $\alpha \to +0$), and a smoothness independent decreasing term proportional to $\frac{\delta}{\alpha^{a/(2a+2)}}$ was developed in [13]. As a consequence, there is something special about this study that norm convergence of regularized solutions to the exact solution can be shown for a wide region of a priori parameter choices and for a specific discrepancy principle without to make any specific smoothness assumption on $x^\dagger$. It is highlighted there that such error decomposition also allows for low order convergence rates under low order smoothness assumptions on $x^\dagger$. But it was observed in [7] that the error decomposition from [13] extends to power type smoothness $x^\dagger \in X_p,\ 0 < p < 1$, and hence yields optimal rates of convergence under the a priori parameter choice.

1.4. **Goal of the present study.** The present paper complements the series of articles mentioned before.

On the one hand side, it extends the error decomposition from [13, 7] to more general smoothness assumptions. On the other hand, we present new results to the *balancing principle* for choosing the regularization parameter in Tikhonov regularization for nonlinear problems with oversmoothing penalties in a Hilbert scale setting. This work was essentially motivated by the recent paper [23]. Pricop-Jeckstadt has also analyzed the balancing principle for nonlinear problems in Hilbert scales, but only for non-oversmoothing penalties. In this sense, we try to close a gap in the theory.

The material is organized as follows. In Section 2 we highlight the decomposition of the error under smoothness in terms of source conditions. This provides us with the required structure in order to found the balancing principles in Section 3. We give a brief account of the history of such principles, and formulate several specifications for the setup under consideration. The results presented in this part are general and may be of independent interest. Finally, in Section 4 we discuss the exponential growth model, both theoretically, and as subject for a numerical case study.

2. General error estimate for Tikhonov regularization in Hilbert scales with oversmoothing penalty

The basis for an analytical treatment of the balancing principle is formed by general error estimates for Tikhonov regularization in Hilbert scales with oversmoothing penalty. With the inequality chain (3) and for $x^\dagger \in \operatorname{int}(\mathcal{D}(F))$ such estimates have been developed recently in [9], [13], and [7] by using *auxiliary elements* as minimizers of (4). Introducing the injective linear operator $G := B^{-(2a+2)}$, the corresponding minimizers $\hat{x}_\alpha$ can be expressed explicitly as

$$\hat{x}_\alpha = \overline{x} + G(G + \alpha\,\mathrm{I})^{-1}(x^\dagger - \overline{x}) = x^\dagger - \alpha(G + \alpha\,\mathrm{I})^{-1}(x^\dagger - \overline{x}).$$

2.1. **Smoothness in terms of source conditions.** General error estimates were obtained under general source conditions, given in terms of index functions ψ.

Definition 1. A continuous and non-decreasing function $\varphi : (0,\infty) \to (0,\infty)$ with $\lim_{t\to+0}\varphi(t) = 0$ is called *index function*. We call this index function *sub-linear* if there is some $t_0 > 0$ such that the quotient $t/\varphi(t)$ is non-decreasing for $0 < t \le t_0$.

In these terms a general source condition for the unknown solution $x^\dagger$ is given in the form of

$$x^\dagger - \overline{x} = \psi(G)\,w, \quad w \in X. \tag{5}$$

Here the linear operator $\psi(G)$ is obtained from the operator G by spectral calculus.

2.2. **Error decomposition.** The balancing principle relies on an error bound in a specific form, and the corresponding fundamental error bound is given next.

Theorem 1. *Let* $x^\dagger \in \mathrm{int}(\mathcal{D}(F))$ *and let hold the inequality chain (3) for given the degree of ill-posedness* $a > 0$. *Moreover, let* ψ *be an index function such that* ψ^{2a+2} *is sub-linear. If* $x^\dagger$ *satisfies a source condition (5) for that* ψ, *then we have for some constant* $c_1 > 0$ *depending on* w *the general error estimate*

$$\| x_\alpha^\delta - x^\dagger \|_X \le c_1 \psi(\alpha) + \frac{\delta}{\lambda(\alpha)}, \tag{6}$$

where $\lambda(\alpha) = \frac{1}{K_2}\alpha^{a/(2a+2)}$, *and* $K_2 = \max\{\, 1, \frac{2}{c_a} \,\}$.

The proof follows the arguments of Proposition 3.4 in [13], and we briefly sketch these.

Proof. A substantial ingredient of the proof is the fact that due to [19] sub-linear index functions are qualifications for the classical Tikhonov regularization approach with norm square penalty, and the verification of formula (25) in [7], which in turn is based on the bounds (21)–(23) ibid. As can be seen from there it is enough to show that under the above assumptions, and for $0 \le \theta \le \frac{2a+1}{2a+2}$, we have that

$$\alpha^{1-\theta} \| G^\theta \left(G + \alpha \,\mathrm{I}\right)^{-1} (x^\dagger - \overline{x}) \|_X \le C\psi(\alpha), \quad \alpha > 0. \tag{7}$$

To this end we start from the observation that under the source condition (5) we have

$$\| G^\theta \left(G + \alpha \,\mathrm{I}\right)^{-1} (x^\dagger - \overline{x}) \|_X \le \| w \|_X \| G^\theta \left(G + \alpha \,\mathrm{I}\right)^{-1} \psi(G) \|_{X \to X}.$$

By introducing the residual function for (classical) Tikhonov regularization $r_\alpha(t) := \alpha/(t + \alpha)$, for $t, \alpha > 0$, it suffices to bound

$$\frac{1}{\alpha} \| r_\alpha(G) G^\theta \psi(G) \|_{X \to X}.$$

The function $t \mapsto t^\theta \psi(t)$ plainly constitutes an index function. We shall establish that it is sub-linear. To this end we write

$$\left[\frac{t}{t^\theta \psi(t)} \right]^{2a+2} = \frac{t^{(1-\theta)(2a+2)}}{\psi^{2a+2}(t)} = t^{(1-\theta)(2a+2)-1} \frac{t}{\psi^{2a+2}(t)}, \quad 0 < t \le t_0.$$

Under the made sub-linearity assumption we find that $t \mapsto t^\theta \psi(t)$ is sub-linear provided that $(1-\theta)(2a+2) - 1 \ge 0$, which corresponds to $\theta \le \frac{2a+1}{2a+2}$. Thus in this range the function $t^\theta \psi(t)$ is a qualification for Tikhonov regularization, and we conclude that

$$\frac{1}{\alpha} \| r_\alpha(G) G^\theta \psi(G) \|_{X \to X} \le C \frac{1}{\alpha} \alpha^\theta \psi(\alpha) = C \alpha^{\theta - 1} \psi(\alpha), \quad \alpha > 0,$$

which yields (7). □

We note that the error estimate (6) does not correspond to the natural estimate $\|x_\alpha^\delta - x^\dagger\|_X \le \|x_\alpha - x^\dagger\|_X + \|x_\alpha^\delta - x_\alpha\|_X$.

We highlight the above result for the prototypical examples, as studied previously.

Example 1 (power-type smoothness)**.** Theorem 1 applies for *Hölder-type source conditions* of the form

$$x^\dagger - \overline{x} = B^{-p}w = G^{p/(2a+2)} \quad \text{for} \quad 0 < p \le 1, \tag{8}$$

which characterize for this type the case of oversmoothing penalties. Indeed, the corresponding function ψ is $\psi(t) = t^{\frac{p}{2a+2}}$, such that $\psi^{2a+2}(t) = t^p$ is sub-linear whenever $0 < p \le 1$. Then the a priori parameter choice $\alpha_* := \delta^{\frac{2a+2}{a+p}}$ yields under the source condition (8) a Hölder-type convergence rate of the form

$$\|x_{\alpha_*}^\delta - x^\dagger\|_X = \mathcal{O}(\delta^{\frac{p}{a+p}}) \quad \text{as} \quad \delta \to 0. \tag{9}$$

Such power-type rates correspond to *moderately ill-posed problems.*

We note that the same convergence rate (9) can also be obtained for α_*, determined by the discrepancy principle in the sense of formula (51) below, cf. [9].

Example 2 (low order smoothness)**.** Theorem 1 also applies for *low order source conditions* with ψ, for which ψ^{2a+2} is always sub-linear. Most prominently, we assume that there is some exponent $\mu > 0$, such that $\psi(t) = K\left(\log^{-\mu}(1/t)\right)$ for $0 < t \le t_0 = e^{-1}$, and continuously extended as constant for $t_0 < t \le \|G\|_{X\to X}$. Moreover, the a priori parameter choice resulting in $\alpha_* := \delta$ yields the rate of convergence

$$\|x_{\alpha_*}^\delta - x^\dagger\|_X = \mathcal{O}(\log^{-\mu}(1/\delta)) \quad \text{as} \quad \delta \to 0. \tag{10}$$

Such logarithmic rates correspond to *severely ill-posed problems* (cf. the study [14]).

Example 3 (no explicit smoothness)**.** Theorem 1 provides us with an upper bound of the error, once a smoothness condition of the form (5) is available such that ψ^{2a+2} is sub-linear. Now we argue as follows. For the operator B^{-1} there is an index function φ such that $x^\dagger - \overline{x} = \varphi(G)v,\ v \in X$, see [18] for compact B^{-1}, and [11] for the general case. In the compact case it is shown in [18, Cor. 2] that this index function may be chosen concave, and hence sub-linear. Then letting $\psi := \varphi^{1/(2a+2)}$ we can apply Theorem 1 for this index function ψ. In particular we conclude the following. If $\alpha_* := \alpha_*(y^\delta, \delta)$ is any parameter choice such that $\alpha_* \to 0$, and also $\delta/\alpha_*^{a/(2a+2)} \to 0$ as $\delta \to 0$, then $\|x_{\alpha_*}^\delta - x^\dagger\|_X \to 0$. For the general case of B, which includes B^{-1} non-compact, the latter result may be found in [13, Thm. 4.1] by an alternative proof based on the Banach–Steinhaus theorem.

3. Balancing principles

Vast majority of regularization theory for ill-posed equations is concerned with asymptotic properties of regularization, as this is convergence, and if so, rates of convergence. For given operator $F : \mathcal{D}(F) \subseteq X \to Y$ between Hilbert spaces X and Y, let $Y \ni y^\delta \mapsto x_\alpha^\delta := R_\alpha(y^\delta) \in X$ be any regularization scheme for the stable approximate determination of $x^\dagger \in \mathcal{D}(F)$ from data $y^\delta \in Y$ such that $\| F(x^\dagger) - y^\delta \|_Y \leq \delta$. Its error at $x^\dagger$ is then considered uniformly for admissible data as

$$e(x^\dagger, R_\alpha, \delta) := \sup_{y^\delta \colon \| F(x^\dagger) - y^\delta \|_Y \leq \delta} \| x_\alpha^\delta - x^\dagger \|_X .$$

We call a parameter choice $\alpha = \alpha(y^\delta, \delta)$ *convergent* if $e(x^\dagger, R_{\alpha(y^\delta,\delta)}, \delta) \to 0$ whenever $\delta \to 0$. In most cases, convergence of regularization parameter choices is analyzed *uniformly on some class* $\mathcal{M} \subset X$, i.e., it is studied when we have $\sup_{x^\dagger \in \mathcal{M}} e(x^\dagger, R_{\alpha(y^\delta,\delta)}, \delta) \to 0$.

In contrast, there are studies which highlight a different paradigm: What is the quality of a particular regularization and parameter choice at any given data y^δ? This is relevant, since in practice we are given just one instance of data y^δ. Then convergence is not an issue, rather the aim is to do the best possible for any such instance. Balancing principles may be used for this purpose.

Lepskiĭ's balancing principle is most prominent for the latter paradigm. It arose in a series of papers, starting from [16], and it gained special attention in statistics within the subject of 'oracle inequalities' for the purpose of model selection since then.

Within classical regularization theory it was first studied in [19]; fundamental discussions are given in [21, 17]. A variation of this statistically motivated approach was followed starting from [24], where the above mentioned paradigm was called quasi-optimality. The original presentation of this idea dates back to [15], called point-wise pseudo-optimal. It is shown ibid. Theorem 7, and for Tikhonov regularization, that the choice α_*, obtained as solutions of the extremal problem

$$\| \alpha \frac{d x_\alpha^\delta}{d\alpha} \|_X + \frac{\delta}{\sqrt{\alpha}} \longrightarrow \min$$

is point-wise pseudo-optimal. Here we follow the approach from [24, 5] by using the concept of *quasi-optimality*.

3.1. **Quasi-optimality.** In what follows, we assume that for some fixed searched-for element $x^\dagger \in X$ and $x_\alpha^\delta = R_\alpha(y^\delta) \in X$ $(\alpha > 0,\ 0 < \delta \leq \delta_0)$ obtained from some noisy data $y^\delta \in Y$ and by using some regularization scheme R_α not further specified, the estimate

$$\| x_\alpha^\delta - x^\dagger \|_X \leq \varphi(\alpha) + \frac{\delta}{\lambda(\alpha)} \qquad (\alpha > 0) \tag{11}$$

holds, where $\varphi,\ \lambda : (0,\infty) \to (0,\infty)$ are both index functions. The function $\alpha \mapsto \lambda(\alpha)$ is assumed to be known. This approach is generic and we shall not make use of any specific properties of the operator F, nor of any specific conditions on the noisy data y^δ.

Definition 2. Suppose that the error bound (11) holds true. A parameter choice strategy $\alpha_* = \alpha_*(y^\delta, \delta)$ $(0 < \delta \le \delta_0)$ is called *quasi-optimal*, if there is a constant $c_2 > 0$ such that for $0 < \delta \le \delta_0$ an estimate of the following kind is satisfied:

$$\| x_{\alpha_*}^\delta - x^\dagger \|_X \le c_2 \inf_{\alpha>0} \left(\varphi(\alpha) + \frac{\delta}{\lambda(\alpha)} \right). \tag{12}$$

Note that the constant c_2 in Definition 2, called the *error constant* below, may depend on the function φ. In addition, note that Definition 2 also includes a posteriori parameter choices since in applications x_α^δ rely on data $y^\delta \in Y$.

We highlight a general feature of quasi-optimal parameter choice.

Proposition 1. *Suppose that the error bound (11) holds true.*

(1) *If for some parameter choice rule $\alpha_+ = \alpha_+(y^\delta, \delta)$ we can guarantee a rate*

$$\varphi(\alpha_+) + \frac{\delta}{\lambda(\alpha_+)} = \mathcal{O}(\varrho(\delta)), \quad \text{as} \quad \delta \to 0,$$

for some index function ϱ, then any quasi-optimal parameter choice $\alpha_* = \alpha_*(y^\delta, \delta)$ yields the convergence rate

$$\| x_{\alpha_*}^\delta - x^\dagger \|_X = \mathcal{O}(\varrho(\delta)) \quad \text{as} \quad \delta \to 0.$$

(2) *Any quasi-optimal parameter choice $\alpha_* = \alpha_*(y^\delta, \delta)$ yields convergence $\| x_{\alpha_*}^\delta - x^\dagger \|_X \to 0$ as $\delta \to 0$.*

Proof. The first part is obvious. For the second part, one may consider the right-hand side of (12) with any parameter choice $\alpha_+ = \alpha_+(y^\delta, \delta)$ such that both $\alpha_+ \to 0$ and $\delta/\lambda(\alpha_+) \to 0$ as $\delta \to 0$. □

We stress once again, the focus of quasi-optimality of a parameter choice strategy is not convergence, rather it emphasizes an *oracle property*: If (by some oracle) we are given a parameter choice rule $\alpha_+ := \alpha_+(y^\delta, \delta)$ which realizes

$$\varphi(\alpha_+) + \frac{\delta}{\lambda(\alpha_+)} \le C \inf_{\alpha>0} \left(\varphi(\alpha) + \frac{\delta}{\lambda(\alpha)} \right), \tag{13}$$

then the quasi-optimal rule is (up to the constant c_2) as good as the oracle choice.

In the sub-sequent sections, we shall describe variants of the balancing principle and we shall show that these are quasi-optimal.

3.2. **The balancing principles: setup and formulation.** We constrain to the following setup. First, we shall assume that the noise amplification term is of the form

$$\lambda(\alpha) = \frac{\alpha^{\varkappa}}{K_2}, \tag{14}$$

where $K_2 > 0$ and $\varkappa > 0$. For standard regularization schemes for selfadjoint and non-selfadjoint linear problems in Hilbert spaces, we have $\varkappa = 1$ and $\varkappa = \frac{1}{2}$, respectively. In the situation of Section 2, representation (14) holds for $\varkappa = \frac{a}{2a+2}$ and $K_2 = \max\{\,1, \frac{2}{c_a}\,\}$.

The following result will be utilized at several occasions.

Lemma 1. *Let $\alpha_+ = \alpha_+(y^\delta, \delta) > 0$ be any parameter choice satisfying*

$$\varphi(\alpha_+) \le c_3 \frac{\delta}{\lambda(\alpha_+)}, \qquad \frac{\delta}{\lambda(\varrho\alpha_+)} \le c_4 \varphi(\varrho\alpha_+), \tag{15}$$

where c_3, $c_4 > 0$ and $\varrho \ge 1$ denote some finite constants chosen independently of δ. Then α_+ satisfies the oracle estimate (13), *with a constant that may by chosen as $C = \varrho^{\varkappa}(1 + \max\{c_3, c_4\})$.*

Proof. Consider the case $0 < \alpha \le \varrho\alpha_+$ first. From $\varphi(\alpha_+) \le c_3 \frac{\delta}{\lambda(\alpha_+)}$ we then obtain

$$\begin{aligned} \varphi(\alpha_+) + \frac{\delta}{\lambda(\alpha_+)} &\le (c_3 + 1)\frac{\delta}{\lambda(\alpha_+)} \le \varrho^{\varkappa}(c_3 + 1)\frac{\delta}{\lambda(\alpha)} \\ &\le \varrho^{\varkappa}(c_3 + 1)\Big(\varphi(\alpha) + \frac{\delta}{\lambda(\alpha)}\Big). \end{aligned}$$

Next we consider the case $\alpha \ge \varrho\alpha_+$. Then the estimate $\frac{\delta}{\lambda(\alpha_+)} \le c_4 \varrho^{\varkappa} \varphi(\varrho\alpha_+)$ yields

$$\varphi(\alpha_+) + \frac{\delta}{\lambda(\alpha_+)} \le (1 + c_4\varrho^{\varkappa})\varphi(\varrho\alpha_+) \le (1 + c_4\varrho^{\varkappa})\Big(\varphi(\alpha) + \frac{\delta}{\lambda(\alpha)}\Big).$$

This completes the proof of the lemma. □

For the numerical realization of the balancing principle considered below, we utilize the following finite set of regularization parameters:

$$\Delta_\delta = \{\, \alpha_0 < \alpha_1 < \cdots < \alpha_N \,\}, \tag{16}$$

where each element of Δ_δ as well as $N \ge 0$ may depend on the noise level δ. We further assume that the elements of Δ_δ form a finite geometric sequence, i.e.,

$$\alpha_j = q^j \alpha_0, \quad j = 0, 1, \ldots, N, \quad \text{with } q > 1, \tag{17}$$

where the spacing parameter q is assumed to be independent of δ. We confine the search for the regularization parameter to the set Δ_δ. Given a tuning parameter $0 < \sigma \le 1$, we consider the set

$$M_\delta := \Big\{\, \alpha \in \Delta_\delta : \varphi(\alpha) \le \sigma \frac{\delta}{\lambda(\alpha)} \,\Big\}. \tag{18}$$

The case $\sigma > 1$ does not provide any improvement and thus is excluded from the considerations below. For α_0 chosen sufficiently small and α_N sufficiently large in a way such that the set M_δ is not empty and in addition satisfies $M_\delta \neq \Delta_\delta$, then the maximum value $\alpha_+ = \max M_\delta$ enjoys quasi-optimality. This immediately follows from Lemma 1, applied with $\varrho = q$, $c_3 = \sigma$, and $c_4 = \frac{1}{\sigma}$. However, such a parameter choice strategy is not implementable since the function φ is not available, in general. Thus we look for feasible sets which contain M_δ and are as close as possible to M_δ.

For $c_5, c_6, c_7 > 0$ fixed, below we assume that

$$0 < \alpha_0 \le c_5 \delta^{1/\varkappa}, \qquad c_6 \le \alpha_N \le c_7, \tag{19}$$

which guarantees that a sufficiently large interval is covered by the set Δ_δ introduced in (16).

Below we discuss several such balancing principles. These differ, e.g., in the number of comparisons executed at each step, and it is seen from the discussion in Section 3.3 that by increasing the number of comparisons we can decrease the error constant c_2.

In each of the subsequent versions, the balancing is controlled by a parameter $\tau_{\mathrm{L}} > 0$, therefore called *balancing constant*, and it is assumed to satisfy

$$\tau_{\mathrm{L}} > 1 + q^{-\varkappa}, \tag{20}$$

where $q > 1$ is the spacing parameter from (17).

First version. Here we consider

(21)
$$H_\delta := \Big\{ \alpha_k \in \Delta_\delta : \| x_{\alpha_j}^\delta - x_{\alpha_{j-1}}^\delta \|_X \le \tau_{\mathrm{L}} \frac{\delta}{\lambda(\alpha_{j-1})} \text{ for any } 1 \le j \le k \Big\}.$$

In order to find the maximum value $\alpha_* = \max H_\delta$, we shall start from $k = 0$, and increase k until $\| x_{\alpha_{k+1}}^\delta - x_{\alpha_k}^\delta \|_X > \tau_{\mathrm{L}} \frac{\delta}{\lambda(\alpha_k)}$ is satisfied for the first time, and take $\alpha_* = \alpha_k$ then. In the exceptional case that there is no such $k \le N-1$, it terminates with $k = N$. There is no need to compute the candidate approximations $x_{\alpha_{k+1}}^\delta, x_{\alpha_{k+2}}^\delta, \ldots, x_{\alpha_N}^\delta$ for this version. In the present paper, our focus will be on this version.

Standard version. The standard version of the balancing principle is related to the set

(22)
$$\widetilde{H}_\delta := \Big\{ \alpha_k \in \Delta_\delta : \| x_{\alpha_k}^\delta - x_{\alpha_j}^\delta \|_X \le \tau_{\mathrm{L}} \frac{\delta}{\lambda(\alpha_j)} \text{ for any } 0 \le j < k \Big\},$$

and it uses the maximum value $\alpha_* = \max \widetilde{H}_\delta$ as regularizing parameter, cf., e.g., [17, 21, 23]. In order to find the maximum value $\alpha_* = \max H_\delta$, one may start from $k = N$, and decrease k until the condition considered in (22) is satisfied for the first time.

This version of the balancing principle requires more comparisons than the first version introduced above, but on the other hand it allows to reduce the error constant. More details on the latter issue are given in Section 3.3.

A third version. Finally we consider a variant of the standard version given through

(23)

$$\widehat{H}_\delta := \Big\{ \alpha_k \in \Delta_\delta : \| x^\delta_{\alpha_i} - x^\delta_{\alpha_j} \|_X \le \tau_{\mathrm{L}} \frac{\delta}{\lambda(\alpha_i)} \text{ for any } 0 \le i < j \le k \Big\},$$

and consider the maximum value $\alpha_* = \max \widehat{H}_\delta$ as regularizing parameter. This variant is considered in [22] in a special framework. In order to find the maximum value $\alpha_* = \max \widehat{H}_\delta$, we shall start from $k = 0$, and increase k until, for the first time, the condition $\| x^\delta_{\alpha_{k+1}} - x^\delta_{\alpha_j} \|_X > \tau_{\mathrm{L}} \frac{\delta}{\lambda(\alpha_j)}$ is satisfied for some index $0 < j \le k$, and take $\alpha_* = \alpha_k$ then. In the exceptional case that there is no such index $k \le N - 1$, the algorithm terminates with $k = N$. Only the candidate approximations $x^\delta_{\alpha_0}, x^\delta_{\alpha_1}, \dots, x^\delta_{\alpha_k}$ have to be computed in the course of this procedure.

Typically one expects $\widetilde{H}_\delta = \widehat{H}_\delta$, and one can show that this identity in fact holds, e.g., for Lavrentiev's method for solving linear, symmetric, positive semidefinite ill-problems. However, in general $\widehat{H}_\delta \subset \widetilde{H}_\delta$ can be guaranteed only, and the set $\widetilde{H}_\delta$ man have gaps in Δ_δ. Under such general circumstances, the standard version of the balancing procedure requires the computation of all elements $x^\delta_{\alpha_0}, x^\delta_{\alpha_1}, \dots, x^\delta_{\alpha_N}$ and consequently has a larger cost than the third version.

Remark 1. In its original form, as introduced by Lepskiĭ, the principle is based on $\sigma = 1$ for the set M_δ from (18). In this case the standard technique only allows choices $\tau_{\mathrm{L}} \ge 4$. Numerical experiments show that sometimes smaller balancing constants τ_{L} produce better results. We note that in the present paper, we verify quasi-optimality of the balancing principle for a range of balancing constants τ_{L} which is bounded from below by the number given in (20). Note that condition (20) even permits τ_{L} close to 1 provided that q is chosen large. The latter case, however, corresponds to a maybe undesirable coarse grid Δ_δ. In addition, it leads to a large error constant, as is shown below, cf. Proposition 2 and the discussion following that proposition.

Finally we mention that within the present context, the analog of Leonov's proposal would read as

$$\| x^\delta_{\alpha_i} - x^\delta_{\alpha_{i-1}} \|_X + \frac{\delta}{\lambda(\alpha_{i-1})} \longrightarrow \min .$$

We shall establish the quasi-optimality for the first variant with some details. The corresponding proofs for the other variants are similar, and hence omitted. The quasi-optimality of Leonov's approach is not clear for the present context of nonlinear ill-posed problems.

We start with the following observation.

Lemma 2. *Suppose that the error bound* (11) *holds true, where* $\lambda(\alpha)$ *is of the form* (14)*. In addition, let* (16), (17) *and* (20) *be satisfied. Let the tuning parameter* $0 < \sigma \le 1$ *used in the definition of the set* M_δ*, cf.* (18)*, be chosen sufficiently small such that* $(\sigma+1)(1+q^{-\varkappa}) \le \tau_L$ *holds. Then we have* $M_\delta \subset H_\delta$*.*

Proof. Let $\alpha = \alpha_k \in M_\delta$ and $1 \le j \le k$. Then α_j, $\alpha_{j-1} \in M_\delta$, and thus

$$\begin{aligned}
\| x_{\alpha_j}^\delta - x_{\alpha_{j-1}}^\delta \|_X &\le \| x_{\alpha_j}^\delta - x^\dagger \|_X + \| x_{\alpha_{j-1}}^\delta - x^\dagger \|_X \\
&\le \varphi(\alpha_j) + \frac{\delta}{\lambda(\alpha_j)} + \varphi(\alpha_{j-1}) + \frac{\delta}{\lambda(\alpha_{j-1})} \\
&\le (\sigma+1)\Big(\frac{1}{\lambda(\alpha_j)} + \frac{1}{\lambda(\alpha_{j-1})}\Big)\delta = (\sigma+1)(1+q^{-\varkappa})\frac{\delta}{\lambda(\alpha_{j-1})} \\
&\le \tau_{\mathrm{L}} \frac{\delta}{\lambda(\alpha_{j-1})}.
\end{aligned}$$

Thus, each $\alpha \in M_\delta$ obeys the estimate in (21), and the proof is complete. □

Lemma 2 and the considerations at the end of Section 3.1 give rise to the following a posteriori choice of the parameter $\alpha = \alpha_*$:

$$\alpha_* = \max H_\delta. \tag{24}$$

Theorem 2. *Suppose that the error bound* (11) *holds true, where* $\lambda(\alpha)$ *is of the form* (14)*. Let* (16), (17), (19) *and* (20) *be satisfied. Then the balancing principle* (24) *is quasi-optimal.*

Proof. The proof will distinguish three cases, and as a preparation we first prove the following assertion. For any $\alpha \in \Delta_\delta$ with $\alpha \le \alpha_*$, we have

$$\| x_{\alpha_*}^\delta - x^\dagger \|_X \le \varphi(\alpha) + c_8 \frac{\delta}{\lambda(\alpha)}, \tag{25}$$

where $c_8 = 1 + \frac{\tau_{\mathrm{L}}}{1-q^{-\varkappa}}$. In fact, there are indices $0 \le k \le N$ and $0 \le m \le N-k$, such that $\alpha = \alpha_k$ and $\alpha_* = \alpha_{k+m}$. We can bound

$$\begin{aligned}
\| x_{\alpha_*}^\delta - x^\dagger \|_X &\le \| x_{\alpha}^\delta - x^\dagger \|_X + \sum_{j=0}^{m-1} \| x_{\alpha_{k+j+1}}^\delta - x_{\alpha_{k+j}}^\delta \|_X \\
&\le \varphi(\alpha) + \frac{\delta}{\lambda(\alpha)} + \tau_{\mathrm{L}}\delta \sum_{j=0}^{m-1} \frac{1}{\lambda(\alpha_{k+j})} \\
&= \varphi(\alpha) + \frac{\delta}{\lambda(\alpha)} + \frac{\tau_{\mathrm{L}}\delta}{\lambda(\alpha)} \sum_{j=0}^{m-1} q^{-j\varkappa} \le \varphi(\alpha) + \Big(1 + \frac{\tau_{\mathrm{L}}}{1-q^{-\varkappa}}\Big)\frac{\delta}{\lambda(\alpha)}.
\end{aligned}$$

This proves (25) with constant c_8 as given. We turn to the main part of the proof. Assume that $\sigma > 0$ is chosen as in Lemma 2. Clearly, $M_\delta \subset \Delta_\delta$.

Case (i): ($M_\delta \neq \varnothing$ and $M_\delta \neq \Delta_\delta$)

The property $M_\delta \neq \varnothing$ allows to consider

$$\alpha_+ := \max M_\delta. \tag{26}$$

From the definition of M_δ and the assumption $M_\delta \neq \Delta_\delta$, we obtain $(\varphi\lambda)(\alpha_+) \leq \sigma\delta \leq (\varphi\lambda)(q\alpha_+)$. Thus, by Lemma 1, the parameter α_+ satisfies an oracle inequality of the form (13). In addition, by Lemma 2 we have that $M_\delta \subset H_\delta$, such that the inequality $\alpha_+ \leq \alpha_*$ holds. Quasi-optimality of α_* under the current situation now immediately follows from the error estimate (25) applied with $\alpha = \alpha_+$.

Case (ii): ($M_\delta = \Delta_\delta$)

For α_+ given by (26), this in fact means $\alpha_+ = \alpha_* = \alpha_N$ and thus

$$(\varphi\lambda)(\alpha_*) \leq \sigma\delta. \tag{27}$$

For the lower bound of $(\varphi\lambda)(\alpha_*)$, we make use of $\alpha_* \geq c_6$ which implies that $\varphi(\alpha_*) \geq \varphi(c_6)$ as well as $\lambda(\alpha_*) \geq \lambda(c_6) = c_6^\varkappa/K_2$. We therefore arrive at

$$\delta \leq c_4(\varphi\lambda)(\alpha_*) \quad \text{with } c_4 = \frac{\delta_0 K_2}{\varphi(c_6)c_6^\varkappa}. \tag{28}$$

This implies that α_* satisfies an estimate of the form (15) and thus is quasi-optimal. This completes the considerations of the case (ii).

Case (iii): ($M_\delta = \varnothing$)

In this case we may consider $\alpha_+ := \alpha_0$. This by (19) means $\alpha_+ = \alpha_0 \leq c_5\delta^{1/\varkappa}$, and thus $\lambda(\alpha_+) \leq (c_5^\varkappa/K_2)\delta$ and $\varphi(\alpha_+) \leq \varphi(c_5\delta_0^{1/\varkappa})$. Then, by the definition of M_δ, we have

$$\sigma\delta \leq (\varphi\lambda)(\alpha_+) \leq \frac{c_5^\varkappa \varphi(c_5\delta_0^{1/\varkappa})}{K_2}\delta. \tag{29}$$

This implies that α_+ satisfies an estimate of the form (15) and thus also the oracle inequality (13). As in case (i), employing estimate (25), we deduce an estimate of the form (12) for α_* for this particular case.

The proof of the theorem is thus completed. □

Remark 2. We stress that the case (i) considered in the above proof is prototypical. For, if the maximum noise level δ_0 is sufficiently small, the cases (ii) and (iii) cannot occur. This follows from the estimates (27) and (29), which lead to contradictions then, respectively. However, larger levels δ_0 give rise for the cases (ii) and (iii), respectively.

Remark 3. Lemma 2 and Theorem 2 also hold for the other two balancing principles given by the sets $\widetilde{H}_\delta$ and $\widehat{H}_\delta$, respectively. More precisely, the same range of balancing constants $\tau_{\rm L}$, cf. (20), and tuning parameters σ may be used. In the wording of Lemma 2, only H_δ has to be replaced by $\widehat{H}_\delta$ and $\widetilde{H}_\delta$, respectively. In (25) in the proof of Theorem 2, the constant c_8 may be reduced to $c_8 = 1 + \tau_{\rm L}$, which in fact has an impact on the corresponding error constant c_2, cf. the discussion in Section 3.3 below.

3.3. **Discussion.** We shall discuss several aspects concerning the balancing principles.

Comparison of the three considered variants of the balancing principle. We continue with a comparison of the considered variants of the balancing principle. Since we have

$$M_\delta \subset \widehat{H}_\delta \subset H_\delta, \qquad M_\delta \subset \widehat{H}_\delta \subset \widetilde{H}_\delta,$$

the latter balancing principle, which is related to the set $\widehat{H}_\delta$, seems to be superior to the other versions. In fact, the set $\widehat{H}_\delta$ it closer to the oracle set M_δ than the other two sets H_δ and $\widetilde{H}_\delta$. In addition, the latter version related to the set $\widehat{H}_\delta$ requires less computational complexity, since the number of x_α^δ to be computed does not exceed the related number for the other versions. Note that for the classical balancing principle related to the set $\widetilde{H}_\delta$, one always has to compute x_α^δ for each $\alpha \in \Delta_\delta$.

Oracle property of the parameter choices. It may be of interest to consider quasi-optimality-type estimates without assuming (19), in particular, the minimal value α_0, and the maximal α_N are mis-specified. The following result is obtained as a corollary of Theorem 2 and its proof.

Corollary 1. *For any of the three considered variants of the balancing principle, we have*

$$\| x_{\alpha_*}^\delta - x^\dagger \|_X \le c_2 \inf_{\alpha_0 \le \alpha \le \alpha_N} \Big(\varphi(\alpha) + \frac{\delta}{\lambda(\alpha)} \Big),$$

where $c_2 > 0$ denotes some finite constant.

Proof. We consider the first balancing principle only. The proofs for the other two balancing principles are quite similar, and are left to the reader. Below, for different situations, we verify estimates of the form

$$\| x_{\alpha_*}^\delta - x^\dagger \|_X \le c_2 \inf_{\alpha \in \mathcal{I}} \Big(\varphi(\alpha) + \frac{\delta}{\lambda(\alpha)} \Big), \tag{30}$$

with appropriate intervals $\mathcal{I}$ and constants c_2, respectively. This in fact follows by a careful inspection of the proofs of Lemma 1 and Theorem 2. In the following considerations, σ denotes a constant satisfying the conditions of Lemma 2, and for the meaning of the constant c_8, we refer to (25).

For case (i) considered in Theorem 2, i.e., $M_\delta \neq \varnothing$ and $M_\delta \neq \Delta_\delta$, we have (30) with $\mathcal{I} = (0, \infty)$ and $c_2 = q^{\varkappa} \frac{\sigma + c_8}{\sigma}$.

For case (ii) in that theorem, i.e., $M_\delta = \Delta_\delta$, we have $\alpha_* = \alpha_+ = \alpha_N$ and $\varphi(\alpha_*) \leq \sigma \frac{\delta}{\lambda(\alpha_*)}$. The first part of the proof of Lemma 1, applied with $\varrho = 1$, then gives (30) with $\mathcal{I} = (0, \alpha_N]$ and $c_2 = \sigma + c_8$.

Finally, for case (iii) considered in the theorem, i.e., $M_\delta = \varnothing$, we have $\alpha_+ = \alpha_0$ and $\varphi(\alpha_+) \geq \sigma \frac{\delta}{\lambda(\alpha_+)}$. The second part of the proof of Lemma 1 then gives (30) with $\mathcal{I} = [\alpha_0, \infty)$ and $c_2 = \frac{1}{\sigma}(\sigma + c_8)$. A combination of those three cases finally gives the statement of the corollary. $\square$

The assertion of Corollary 1 may be considered as *oracle type*: If the range of parameters $[\alpha_0, \alpha_N]$ is not specified correctly, then the chosen parameter α_* is, up to the constant c_2, at least as good as the best value within the specified range.

Controlling the error constant. The following proposition specifies the error constant c_2 for each of the considered balancing principles.

Proposition 2. *Let the maximum noise level δ_0 be sufficiently small, and in addition, let* (17), (19) *and* (20) *be satisfied. Then the error constant may be chosen as*

$$c_2 = q^{\varkappa} \frac{\sigma + c_8}{\sigma}, \tag{31}$$

where $\sigma \leq \min\{\frac{\tau_L}{1+q^{-\varkappa}} - 1, 1\}$. In addition, $c_8 := 1 + \frac{\tau_L}{1-q^{-\varkappa}}$ for the balancing principle (21), and $c_8 := 1 + \tau_L$ for the versions from (22) and (23), respectively.

Proof. Under the given assumptions on δ_0, case (i) in the proof of Theorem 2 applies, i.e., we have $M_\delta \neq \varnothing$ and $M_\delta \neq \Delta_\delta$ there. For α_+ as in (26), an application of the estimate in (25) for $\alpha = \alpha_+$, and a careful inspection of the proof of Lemma 1 gives

$$\begin{aligned} \|x_{\alpha_*}^\delta - x^\dagger\|_X &\leq \max\left\{\left(1 + \frac{q^{\varkappa} c_8}{\sigma}\right)\varphi(\alpha),\ q^{\varkappa}(\sigma + c_8)\frac{\delta}{\lambda(\alpha)}\right\} \\ &\leq q^{\varkappa}\frac{\sigma + c_8}{\sigma}\left(\varphi(\alpha) + \frac{\delta}{\lambda(\alpha)}\right) \qquad (\alpha > 0). \end{aligned} \tag{32}$$

For the balancing principle (21) this was shown to hold for $c_8 = 1 + \frac{\tau_L}{1-q^{-\varkappa}}$. For the balancing principles from (22) and (23) the reasoning in Theorem 2 simplifies, and the bound (25) holds with $c_8 := 1 + \tau_L$. This completes the sketch of this proof. $\square$

We note that the special form of σ considered in Proposition 2 is caused by the requirement made in Lemma 2.

We next discuss the optimal choice of the parameters used in the balancing principle (24) to minimize the error constant c_2. First, we consider the spacing parameter $q > 1$ to be fixed. Thus, in order to minimize c_2 we need to consider $\frac{\sigma+c_8}{\sigma}$ only. The constant c_8, as a function of τ_L is monotone, such that the smallest possible value of τ_L minimizes c_8, and, taking into account the requirements in Lemma 2, we let $\tau_L(q^\varkappa) := (\sigma+1)(1+q^{-\varkappa})$. Now we need to distinguish the values for c_8 as indicated in Proposition 2. For the first balancing principle, based on (21), we find that

$$c_2 = q^\varkappa \frac{\sigma + c_8}{\sigma} = \frac{\sigma+1}{\sigma} \frac{2q^\varkappa}{1-q^{-\varkappa}} \geq \frac{4q^\varkappa}{1-q^{-\varkappa}},$$

the latter being achieved for $\sigma = 1$. This can further be optimized with respect to the spacing parameter q, and it is minimized for $q^\varkappa = q_{\mathrm{opt}}^\varkappa := 2$. With these specifications we find that

$$\tau_{L,\mathrm{opt}} = 3, \quad \text{and} \quad c_{2,\mathrm{opt}} = 16.$$

We next consider the size of the error constants of the other two balancing principles related with the sets given by (22) and (23), respectively. In either case, the error constant c_2 is again of the form (31), with $c_8 = 1 + \tau_{\mathrm{L}}$.

Thus, for $\tau_{\mathrm{L}} = (\sigma+1)(1+q^{-\varkappa})$ we have that

$$c_2 = q^\varkappa \frac{\sigma + c_8}{\sigma} = q^\varkappa \frac{\sigma+1}{\sigma}(2 + q^{-\varkappa}).$$

Again, this is minimized for $\sigma := 1$, and it gives

$$c_{2,\mathrm{opt}} = 2\,(2q^\varkappa + 1) \tag{33}$$

with corresponding $\tau_{L,\mathrm{opt}} = 2(1+q^{-\varkappa}) < 4$. This means that the error constant becomes smaller as the grid Δ_δ becomes finer, with $c_2 \to 6$ as $q \to 1$. For the best grid independent choice of the parameter τ_L, we find that $\tau_L = 4$, and hence we recover the original Lepskiĭ principle with constant $c_2 = 6q^\varkappa$.

Remark 4. The quasi-optimality results for all three methods considered in the present work can also be written in the frequently used form

$$\| x_{\alpha_*}^\delta - x^\dagger \|_X \leq c_2 \varphi(\alpha_+), \tag{34}$$

where the parameter $\alpha_+ > 0$ satisfies $\varphi(\alpha_+) = \frac{\delta}{\lambda(\alpha_+)}$, and the error constant c_2 is given by (31). This can be seen by considering estimate (32) in the proof of Proposition 2. Note that we have $\varphi(\alpha_+) \leq \inf_{\alpha>0}(\varphi(\alpha) + \frac{\delta}{\lambda(\alpha)}) \leq 2\varphi(\alpha_+)$, so quasi-optimality is in fact equivalent to (34) for some constant c_2.

For the standard balancing principle (22), utilized with the traditional balancing constant $\tau_{\mathrm{L}} = 4$, the error constant in estimate (34) takes the form $c_2 = 6q^\varkappa$, which is a well-known result, cf., e.g., [17, 21]. The above discussion shows that the error constant c_2 in (34) can be reduced to the form (33) by choosing τ_{L} somewhat smaller.

3.4. **Specific impact on oversmoothing penalties.** After the presentation of various facets of the general theory for balancing principles in the preceding paragraphs of this section, we return to the specific situation of oversmoothing penalties as outlined in Sections 1 and 2. To characterize the impact of the general theory on that situation, we recall the error estimate (6) with the specific index function $\lambda(\alpha) = \frac{1}{K_2}\alpha^{a/(2a+2)}$ and with the specific constant $K_2 = \max\{\,1, \frac{2}{c_a}\,\}$.

In Examples 1 and 2 we have explicitly described a priori parameter choice rules as well as the discrepancy principle as an a posteriori choice rule. For the nonlinear inverse problem (1) at hand, the corresponding convergence rates are given in (9) for the Hölder case, and (10) for the logarithmic case. Here, we complement those rate results by analog assertions for the balancing principles. The results are based on the quasi-optimality of the balancing principles under consideration, in connection with the first part of Proposition 1.

In case that no explicit smoothness for the solution of the nonlinear inverse problem (1) is available, we note that any quasi-optimal rule for Tikhonov regularization with oversmoothing penalty yields convergence. For B^{-1} compact, this can be seen by consulting Example 3 and the second part of Proposition 1.

We briefly summarize the impact of the theory of the first balancing principle (24), when applied to nonlinear Tikhonov regularization with oversmoothing penalty, by the following corollary. Note that the assertions of the corollary can be formulated in an analog manner for the other two balancing principles, and we refer to Remark 3 above.

Corollary 2. *Consider nonlinear Tikhonov regularization with oversmoothing penalty as introduced in Sections 1 and 2, with the regularization parameter α_* determined by the balancing principle* (24) *under the required conditions (19) and (20). For Hölder-type smoothness (8) as considered in Example 1, with $0 < p \le 1$, one has*

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

Similarly, for logarithmic source conditions as considered in Example 2, we obtain logarithmic rates

$$\|x_{\alpha_*}^\delta - x^\dagger\|_X = \mathcal{O}(\log^{-\mu}(1/\delta)) \quad \text{as} \quad \delta \to 0.$$

Proof. This is an immediate consequence of quasi-optimality of the balancing principle (24), and of the convergence rate results in Examples 1 and 2 in connection with the first part of Proposition 1. □

We explicitly highlight the following fact, intrinsic in the proof: If the parameter $\alpha_*(\delta)$, obtained by the a priori choice, cf. Examples 1–3, is in the interval $[\alpha_0(\delta), \alpha_N(\delta)]$, then the corresponding convergence rates for the parameter choice according to the balancing principle are valid. Otherwise Corollary 1 applies.

As already noticed, the study [23] by Pricop-Jeckstadt is close to our approach on balancing principles. However, it does not include the case of oversmoothing penalties, a gap which is closed here. Despite the fact that the nonlinearity requirements of [23] are slightly different, the main difference lies in the following fact: The proofs (for the non-oversmoothing case) in ibid. are based on an error decomposition into a noise amplification error, and a bias that occurs when the data are noise-free. This technique fails in the oversmoothing case, where instead a certain auxiliary element is used.

4. Exponential growth model: properties and numerical case study

For a case study we shall collect the theoretical properties of the exponential growth model, first presented in [4, Section 3.1]. More details about properties of the nonlinear forward map F as in (36), below, can be found in [6]. Then we highlight different behavior for the reconstruction in the oversmoothing and non-oversmoothing cases, respectively.

4.1. Properties.

For analytical and numerical studies we are going to exploit the exponential growth model

$$y'(t) = x(t)\,y(t) \quad (0 \le t \le 1), \qquad y(0) = 1, \tag{35}$$

considered in the Hilbert space $L^2(0,1)$. The inverse problem consists in the identification of the square-integrable time-dependent function $x(t)$ $(0 \le t \le 1)$ in (35) from noisy data $y^\delta \in L^2(0,1)$ of the solution $y(t)$ $(0 \le t \le 1)$ to the corresponding initial value O.D.E. problem. In this context, we suppose a deterministic noise model $\|y^\delta - y\|_{L^2(0,1)} \le \delta$ with noise level $\delta > 0$. This identification problem can be written in form of an operator equation (1) with the nonlinear forward operator

$$[F(x)](t) = \exp\left(\int_0^t x(\tau)d\tau\right) \qquad (0 \le t \le 1) \tag{36}$$

mapping in $L^2(0,1)$ with full domain $\mathcal{D}(F) = L^2(0,1)$.

Evidently, F is globally *injective*. One can also show on the one hand that F is *weakly sequentially continuous* and on the other hand that F is *Fréchet differentiable* everywhere. It possesses for all $x \in L^2(0,1)$ the Fréchet derivative $F'(x)$, explicitly given as

$$[F'(x)h](t) = [F(x)](t) \int_0^t h(\tau)d\tau \quad (0 \le t \le 1, \;\; h \in L^2(0,1)). \tag{37}$$

This Fréchet derivative is a compact linear mapping in $L^2(0,1)$, because it is a composition $F'(x) = M \circ J$ of the bounded linear multiplication operator M mapping in $L^2(0,1)$ defined as

$$[Mg](t) = [F(x)](t)\,g(t) \;(0 \le t \le 1),$$

and the compact linear integration operator J mapping in $L^2(0,1)$ defined as

$$[Jh](t) = \int_0^t h(\tau)d\tau \quad (0 \le t \le 1). \tag{38}$$

We mention that the continuous multiplier function $F(x)$ in M is bounded below and above by finite positive values due to

$$\exp(-\|x\|_{L^2(0,1)}) \le [F(x)](t) \le \exp(\|x\|_{L^2(0,1)}) \ (0 \le t \le 1), \tag{39}$$

for $x \in L^2(0,1)$, and hence $F(x) \in L^\infty(0,1)$.

Let us denote by

$$\mathcal{B}_r(x^\dagger) = \{z \in L^2(0,1) : \ \|z - x^\dagger\|_{L^2(0,1)} \le r\},$$

the closed ball with radius $r > 0$ and center $x^\dagger$.

The following lemma highlights, that F satisfies a nonlinearity condition of tangential cone-type.

Lemma 3. *For the nonlinear operator F from (36), the inequality*

$$\begin{aligned} \|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\|_{L^2(0,1)} \\ \le \|x - x^\dagger\|_{L^2(0,1)} \, \|F(x) - F(x^\dagger)\|_{L^2(0,1)} \end{aligned} \tag{40}$$

is valid for all $x, x^\dagger \in L^2(0,1)$. Consequently, we have for arbitrary but fixed $r > 0$ and $x^\dagger \in L^2(0,1)$ the inequality

$$\frac{1}{1+r} \, \|F'(x^\dagger)(x - x^\dagger)\|_{L^2(0,1)} \le \|F(x) - F(x^\dagger)\|_{L^2(0,1)} \tag{41}$$

for all $x \in \mathcal{B}_r(x^\dagger)$. Moreover, whenever $0 < r < 1$ we have that

$$\|F(x) - F(x^\dagger)\|_{L^2(0,1)} \le \frac{1}{1-r} \, \|F'(x^\dagger)(x - x^\dagger)\|_{L^2(0,1)} \tag{42}$$

for all $x \in \mathcal{B}_r(x^\dagger)$.

Proof. By setting $\theta(t) := [J(x - x^\dagger)](t)$ $(0 \le t \le 1)$ we have

$$[F(x) - F(x^\dagger)](t) = [F(x^\dagger)](t) \, (\exp(\theta(t)) - 1)$$

and

$$[F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)](t) = [F(x^\dagger)](t) \, (\exp(\theta(t)) - 1 - \theta(t)).$$

Then the general estimate

$$|\exp(\theta) - 1 - \theta| \le |\theta| \, |\exp(\theta) - 1|,$$

which is valid for all $-\infty < \theta < +\infty$, leads to

$$|[F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)](t)| \le |\theta(t)| \, |[F(x) - F(x^\dagger)](t)|,$$

for all $0 \le t \le 1$. By using the Cauchy–Schwarz inequality this implies

$$|[F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)](t)| \le \|x - x^\dagger\|_{L^2(0,1)},$$

again, for $0 \le t \le 1$. This yields the inequality (40). Both inequalities (41) and (42) are immediate consequences by applying the triangle inequality. The proof of the lemma is complete. $\square$

We are going to establish that the forward operator F from (36) obeys the nonlinearity condition (3). To this end we use the Hilbert scale model as introduced in Section 1.1. Here, the Hilbert scale $\{X_\tau\}_{\tau\in\mathbb{R}}$ with $X_0 = X = L^2(0,1)$, $X_\tau = \mathcal{D}(B^\tau)$ for $\tau > 0$ and $X_\tau = X$ for $\tau < 0$, is generated by the unbounded, self-adjoint, and positive definite linear operator

$$B := (J^*J)^{-1/2} \tag{43}$$

induced by the integration operator J from (38). The domain $\mathcal{D}(B)$ of B is dense in $L^2(0,1)$ and its range $\mathcal{R}(B)$ coincides with $L^2(0,1)$. For each $\tau \in \mathbb{R}$ one can define the norm

$$\|x\|_\tau := \|B^\tau x\|_{L^2(0,1)} \quad \text{defined for all} \quad x \in X_\tau.$$

The powers of B are linked to the powers of J. By analyzing the Riemann–Liouville fractional integral operator J^p for levels p from the interval $(0,1]$ we have that

$$X_p = \mathcal{D}(B^p) = \mathcal{R}((J^*J)^{p/2}) \quad \text{for} \quad 0 < p \le 1.$$

Due to [3, Lemma 8] this gives the explicit representation

$$X_p = \begin{cases} H^p(0,1) & \text{for} \quad 0 < p < \frac{1}{2} \\ \{x \in H^{\frac{1}{2}}(0,1) : \int\limits_0^1 \frac{|x(t)|^2}{1-t} dt < \infty\} & \text{for} \quad p = \frac{1}{2} \\ \{x \in H^p(0,1) : \; x(1) = 0\} & \text{for} \quad \frac{1}{2} < p \le 1 \end{cases}, \tag{44}$$

where $H^p(0,1)$ denotes the corresponding fractional hilbertian Sobolev space. Note that for $0 < p < \frac{1}{2}$ the spaces X_p and the hilbertian Sobolev spaces $H^p(0,1)$ coincide, whereas for $p > \frac{1}{2}$ an additional homogeneous boundary condition occurs at the right end of the interval.

Using the Hilbert scale introduced above we have collected now all ingredients for verifying an inequality chain of type (3) with a degree $a = 1$ of ill-posedness.

We start with

$$\|Jh\|_{L^2(0,1)} = \|(J^*J)^{1/2} h\|_{L^2(0,1)} = \|B^{-1}h\|_{L^2(0,1)} = \|h\|_{-1}$$

valid for all $h \in L^2(0,1)$, and we aim at applying Lemma 3. In this context, we set on the one hand $k_0 := \exp(-\|x^\dagger\|_{L^2(0,1)})$, $K_0 := \exp(\|x^\dagger\|_{L^2(0,1)})$ and on the other hand $c_1 := \frac{k_0}{1+r}$ and $C_1 := \frac{K_0}{1-r}$. Then we have $0 < k_0 \le [F(x^\dagger)](t) \le K_0 < \infty$ from (39). By using formula (37) we obtain, for all $x \in L^2(0,1)$ and $r > 0$, the estimate

$$c_1 \|x - x^\dagger\|_{-1} = \frac{k_0}{1+r} \|J(x - x^\dagger)\|_{L^2(0,1)} \le \frac{1}{1+r} \|F'(x^\dagger)(x - x^\dagger)\|_{L^2(0,1)}, \tag{45}$$

whenever $x \in \mathcal{B}_r(x^\dagger)$. In the same manner one deduces that for all $x \in L^2(0,1)$, and $0 < r < 1$, the right-side estimate

$$\frac{1}{1-r}\,\|F'(x^\dagger)(x-x^\dagger)\|_{L^2(0,1)} \le \frac{K_0}{1-r}\,\|J(x-x^\dagger)\|_{L^2(0,1)} = C_1\|x-x^\dagger\|_{-1}$$

holds true whenever $x \in \mathcal{B}_r(x^\dagger)$. By Lemma 3 this yields the inequality chain

$$c_1\,\|x-x^\dagger\|_{-1} \le \|F(x)-F(x^\dagger)\|_{L^2(0,1)} \le C_1\,\|x-x^\dagger\|_{-1} \quad \text{for} \quad x \in \mathcal{B}_r(x^\dagger),$$

with $0 < r < 1$. This proves the assertion of the following proposition.

Proposition 3. *For $X = Y = L^2(0,1)$ we consider the nonlinear operator F from (36). Its domain $\mathcal{D}(F)$ is restricted to a closed ball $\mathcal{B}_r(x^\dagger)$ around some element $x^\dagger \in X$, and with radius radius $r < 1$.*

Within the Hilbert scale generated by the operator B from (43), induced by the integration operator J from (38), the operator F obeys the nonlinearity condition (3) with $a = 1$. The positive constants c_1 and C_1 depend on $x^\dagger$ and r.

Next we discuss assertions on missing stability and well-posedness for the operator equation (1).

Proposition 4. *For arbitrary solution $x^\dagger \in L^2(0,1)$ the operator equation (1) with $F : \mathcal{D}(F) = X = L^2(0,1) \to Y = L^2(0,1)$ from (36) is locally ill-posed (cf.* [12, Definition 3]*). It is not stably solvable at y (cf.* [12, Definition 1]*) for arbitrary right-hand element $y \in \mathcal{R}(F) := \{z \in L^2(0,1) :\ z = F(\xi),\ \xi \in L^2(0,1)\}$ from the range of F.*

Proof. One can easily show that, as a consequence of the compactness of J in $L^2(0,1)$, the operator $F : \mathcal{D}(F) = X = L^2(0,1) \to Y = L^2(0,1)$ from (36) is *strongly continuous* in the sense of Definition 26.1 (c) from [28]. Thus weakly convergent sequences $x_n \rightharpoonup x_0$ in $L^2(0,1)$ imply norm convergent image sequences such that $\lim_{n\to\infty} \|F(x_n) - F(x_0)\|_{L^2(0,1)} = 0$. For any orthonormal system $\{e_n\}_{n=1}^\infty$ in $L^2(0,1)$ and any radius $r > 0$ we then have $x_n = x^\dagger + re_n \rightharpoonup x^\dagger,\ x_n \in \mathcal{B}_r(x^\dagger)$, and $\lim_{n\to\infty} \|F(x_n) - F(x_0)\|_{L^2(0,1)} = 0$. On the one hand this proves the local ill-posedness at $x^\dagger$. On the other hand, since F is injective with the inverse operator $F^{-1} : \mathcal{R}(F) \subset L^2(0,1) \to L^2(0,1)$, the mapping F^{-1} cannot be continuous at $y = F(x^\dagger)$, which contradicts the stable solvability of the operator equation at y. The proof is complete. □

Note that for injective forward operators stable solvability and continuity of the inverse operator coincide. Moreover note that in the Hilbert space $L^2(0,1)$ the strong continuity of F implies that F is compact (cf. [28, Proposition 26.2]).

Remark 5. Due to the local ill-posedness of F defined by formula (36) and mapping from $X_0 = X = L^2(0,1)$ to $Y = L^2(0,1)$ with the associated norm topologies, we have that for arbitrarily small $r > 0$ there is no constant $c_0 > 0$ such that

$$c_0 \, \|x - x^\dagger\|_{L^2(0,1)} \le \|F(x) - F(x^\dagger)\|_{L^2(0,1)} \;\text{ for all }\; x \in \mathcal{B}_r(x^\dagger).$$

However, if the weaker X_{-1}-norm $\|\cdot\|_{-1} = \|B^{-1}\cdot\|_{L^2(0,1)} = \|J\cdot\|_{L^2(0,1)}$ applies for the pre-image space of F, one can see from the estimates (41) and (45) that for all $r > 0$ there exists a constant $c_1 > 0$ depending on $x^\dagger$ and r such that

$$c_1 \, \|x - x^\dagger\|_{-1} \le \|F(x) - F(x^\dagger)\|_{L^2(0,1)} \;\text{ for all }\; x \in \mathcal{B}_r(x^\dagger),$$

which proves that (1) is *locally well-posed* everywhere for that norm pairing. For the convergence theory of Tikhonov regularization in Hilbert scales in case of oversmoothing penalties, however, this requires an a priori restriction of the domain $\mathcal{D}(F)$ to bounded sets (balls), because (3) is originally needed in [13] with respect to that example with the full domain $\mathcal{D}(F) = L^2(0,1)$.

On the other hand, there exists no global constant $c_1 > 0$ depending only on $x^\dagger$ such that $c_1 \, \|x - x^\dagger\|_{-1} \le \|F(x) - F(x^\dagger)\|_{L^2(0,1)}$ for each $x \in L^2(0,1)$. This follows, for any $x^\dagger$, by considering, e.g., the functions

$$x(t) = x_n(t) \equiv -n \quad \text{for} \quad n = 1, 2, \ldots \,.$$

In fact, we then have $[Jx_n](t) = -nt$ and $[F(x_n)](t) = \exp(-nt)$, and thus

$$\|Fx_n\|^2_{L^2(0,1)} = \frac{1}{2n}(1 - e^{-2n}) \to 0,$$

but

$$\|x_n\|_{-1} = \|Jx_n\|_{L^2(0,1)} = \frac{n}{\sqrt{3}} \to \infty,$$

as $n \to \infty$.

The restriction to a small ball with radius $r < 1$ for the right-hand inequality of (3), as caused by the condition (42), is not problematic for the theory (cf. [9, 13]). This part is only used by auxiliary elements that are close to $x^\dagger$ for sufficiently small regularization parameters $\alpha > 0$ whenever $x^\dagger$ is supposed to be an interior point of $\mathcal{D}(F)$, which is trivial for $\mathcal{D}(F) = X$.

The character of ill-posedness of the operator equation (1) with F from (36) can be illustrated at the exact solution

$$x^\dagger(t) \equiv 1 \quad (0 \le t \le 1), \tag{46}$$

which will be used in the numerical study below. In this case we have $[F(x^\dagger)](t) = \exp(t)$ $(0 \le t \le 1)$. If we perturb this exact right-hand side by a continuously differentiable noise function $\eta(t)$ $(0 \le t \le 1)$

with $\eta(0) = 0$, then the pre-image of $F(x^\dagger) + \eta$ attains the explicit form

$$x_\eta(t) = \frac{\exp(t) + \eta'(t)}{\exp(t) + \eta(t)} \quad (0 \le t \le 1).$$

In particular, for $\eta(t) = \delta \sin(nt)$ $(0 \le t \le 1)$ with multiplier $\delta > 0$ and $\|\eta\|_{L^2(0,1)} \le \delta$ we have $x_\eta = x_n$ defined as

$$x_n(t) = \frac{\exp(t) + n\,\delta\,\cos(nt)}{\exp(t) + \delta\,\sin(nt)} \quad (0 \le t \le 1), \tag{47}$$

as well as $\|F(x_n) - F(x^\dagger)\|_{L^2(0,1)} \le \delta$ for all $n \in \mathbb{N}$ and $\delta > 0$. The next proposition emphasizes for F from (36) that in spite of very small image deviations $\|F(x) - F(x^\dagger)\|_{L^2(0,1)}$ the corresponding error norm $\|x - x^\dagger\|_{L^2(0,1)}$ can be arbitrarily large.

Proposition 5. *For arbitrarily small $\delta > 0$ the pre-image set $F^{-1}(\mathcal{B}_\delta(F(x^\dagger)))$ for $x^\dagger$ from (46) is not bounded in $L^2(0,1)$, i.e. there exist sequences $\{x_n\}_{n=1}^\infty \subset F^{-1}(\mathcal{B}_\delta(F(x^\dagger)))$ with $\lim_{n\to\infty} \|x_n\|_{L^2(0,1)} = +\infty$.*

Proof. To this end we shall use the explicit sequence $\{x_n\}_{n=1}^\infty$ from (47), for which the estimate

$$\|x_n\|^2_{L^2(0,1)} = \int_0^1 \frac{(\exp(t) + n\,\delta\,\cos(nt))^2}{(\exp(t) + \delta\,\sin(nt))^2} dt = \int_0^1 \frac{(1 + n\,\delta\,\frac{\cos(nt)}{\exp(t)})^2}{(1 + \delta\,\frac{\sin(nt)}{\exp(t)})^2} dt$$

$$\ge \frac{n^2\delta^2 \int_0^1 (\cos(nt))^2 dt}{e^2(1+\delta)^2} = \frac{n^2\delta^2}{e^2(1+\delta)^2}\left(\frac{1}{2} + \frac{\sin(n)\cos(n)}{2n}\right)$$

holds true. Consequently, we have $\|x_n\|_{L^2(0,1)} \ge \frac{n\delta}{2e(1+\delta)}$ for sufficiently large $n \in \mathbb{N}$, such that $\lim_{n\to\infty} \|x_n\|_{L^2(0,1)} = +\infty$. □

4.2. **Numerical case study.** The following numerical case study operates in the setting of Section 4.1 and complements the theoretical results. Therefore we minimize the Tikhonov functional of type (2) with $\bar{x} = 0$ and forward operator (36) derived from the exponential growth model. Recall that the inequality chain (3) holds with degree of ill-posedness $a = 1$. Further, set $X = Y = L^2(0,1)$. To obtain the X_1-norm in the penalty, use $\|\cdot\|_1 = \|\cdot\|_{H^1(0,1)}$ and additionally enforce the boundary condition $x(1) = 0$ in accordance with the construction of the Hilbert scale (44). In all experiments we use the exact solution $x^\dagger(t) = 1$; $(0 < t \le 1)$. As this particular exact solution is smooth, but violates the boundary condition $x^\dagger(1) = 0$, we have $x^\dagger \in X_p$ for all $0 < p < \frac{1}{2}$. This means, a Hölder-type source condition as outlined in Example 1 holds. A discretization level of $N = 1000$ in the time domain is used. The noise is then constructed by sampling one realization of a vector $\xi = (\xi_1, \ldots, \xi_{1000})$ of 1000 i.i.d. standard Gaussian random variables. This is then normalized to have $\|\xi\|_{L^2(0,1)} = 1$, and $\delta\xi$ added to the exact data y. For noise level δ this yields $\|y - y^\delta\|_Y = \delta$. The minimization problem itself is solved using the MATLAB®-routine `fmincon`. Integrals are discretized using the trapezoid rule.

Further, a modification of the first variant (21) of the balancing principle is implemented. Precisely, we modify H_δ introduced in formula (21) as

(48)
$$H_\delta^{mod} := \left\{ \alpha_k \in \Delta_\delta : \|x^\delta_{\alpha_j} - x^\delta_{\alpha_{j-1}}\|_X \le C_{BP} \frac{\delta}{\alpha_{j-1}^{a/(2a+2)}} \text{ for} 1 \le j \le k \right\}.$$

This is necessary, because the constant K_2, which is required to be known in (14), is not available. So, we use the constant C_{BP} as a replacement for $\tau_L K_2$, instead. Then we set $\alpha_* = \alpha_{BP} := \max H_\delta^{mod}$.

Since $x^\dagger$ is known we can compute the regularization errors $\|x^\delta_\alpha - x^\dagger\|_X$. This can be interpreted as a function of δ and justifies a regression for the model function

$$\|x^\delta_\alpha - x^\dagger\|_X \le c_x \delta^{\kappa_x}. \tag{49}$$

Similarly we estimate the asymptotic behavior of the selected regularization parameter through the ansatz

$$\alpha \sim c_\alpha \delta^{\kappa_\alpha}. \tag{50}$$

Both exponents κ_x and κ_α as well as the corresponding multipliers c_x and c_α are obtained using a least squares regression based on samples for varying δ.

In a first case study we consider several constants C_{BP} used in the balancing principle, the results of which are displayed in Table 1. Recall the results of Theorem 1 as well as Example 1: if a Hölder-type source condition holds, we expect $\kappa_x = \frac{p}{a+p}$. As $a = 1$ and $p \approx \frac{1}{2}$, this corresponds very well with the numerical observations in the third column of Table 1. Further recall the a priori parameter choice $\alpha_* = \delta^{\frac{2a+2}{a+p}}$, which here reads as $\alpha_* = \delta^{\frac{8}{3}}$. The fifth column of Table 1 shows the resulting $\alpha-$rates for the balancing principle. We therefore conclude that the resulting rates coincide with this particular a priori choice.

C_{BP}	c_x	κ_x	c_α	κ_α
0.02	0.5275	0.3373	3.3750	3.0000
0.05	0.5241	0.3337	2.2241	2.8613
0.1	0.7188	0.3426	5.8352	2.5925

TABLE 1. Exponential growth model with $x^\dagger(t) \equiv 1$; $(0 < t \le 1)$. Numerically computed results for the balancing principle (21) yielding multipliers and exponents of regularization error (49) and α-rates (50) for various C_{BP}.

Figure 1 shows the realized regularization parameters α_{BP} (left) ob-

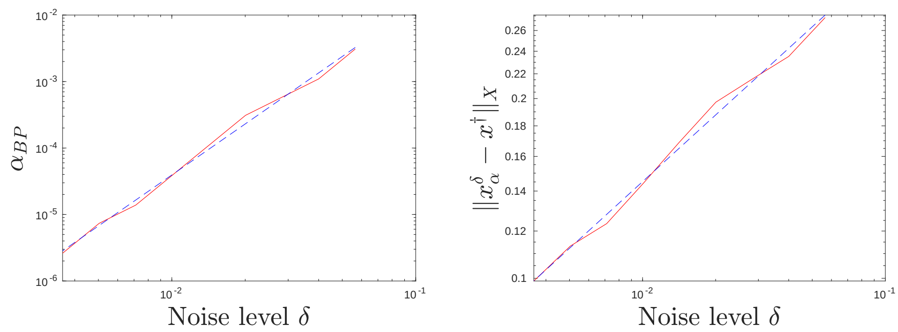

FIGURE 1. Exponential growth model with $x^\dagger(t) \equiv 1$; $(0 < t \leq 1)$ and parameter choice using the balancing principle (21) with $C_{BP} = 0.1$. α_{BP} in red for various δ and best approximating regression line in blue/dashed on a log-log scale (left) and approximation error $\|x_\alpha^\delta - x^\dagger\|_X$ in red and approximate rate in blue/dashed (right).

tained by the balancing principle and the corresponding regularization errors (right) as well as their respective approximations. We observe an excellent fit which confirms our confidence in this approach, and in the implementation.

Next we fix the noise level δ and compare various parameter choice rules. Besides the balancing principle, we consider a discrepancy principle, where the parameter is chosen a posteriori such that

$$\|F(x_\alpha^\delta) - y^\delta\|_Y = C_{DP}\,\delta \tag{51}$$

for given noise level δ and suitable constant $C_{DP} > 0$. Recall, that this parameter choice rule also yields the order optimal convergence rate. Additionally we study the heuristic parameter choice originally developed by Tikhonov, Glasko and Leonov[1] [15, 27] violating the Bakushinskij veto established in [1]. Hence we consider a sequence of regularization parameters

$$\alpha_k = \alpha_0 q^k : j = 0, 1, \ldots, M \tag{52}$$

for $q > 0$, some suitable α_0 and appropriate M. Here, $x_{\alpha_k}^\delta$ denotes the regularized solutions to the functional (2) with regularization parameter α_k. Using the series of parameters (52) the suitable regularization parameter according to this parameter choice rule is then chosen by minimizing

$$\|x_{\alpha_{k+1}}^\delta - x_{\alpha_k}^\delta\|_X \to \min \quad 1 \leq k \leq M-1$$

[1]The authors in the respective publications call this parameter choice quasioptimality. In order to avoid ambiguity with respect to Definition 2 we avoid this terminology here, but denote the resulting regularization parameter as α_{QO}.

with respect to α_k. Moreover consider $\alpha_{\rm opt}$ which minimizes the error $\|x_\alpha^\delta - x^\dagger\|_X$, i.e.,

$$\alpha_{\rm opt} := \min_\alpha \|x_\alpha^\delta - x^\dagger\|_X \tag{53}$$

assuming the exact solution is known.

All of the above parameter choices are visualized for $\delta = 0.0179$ in

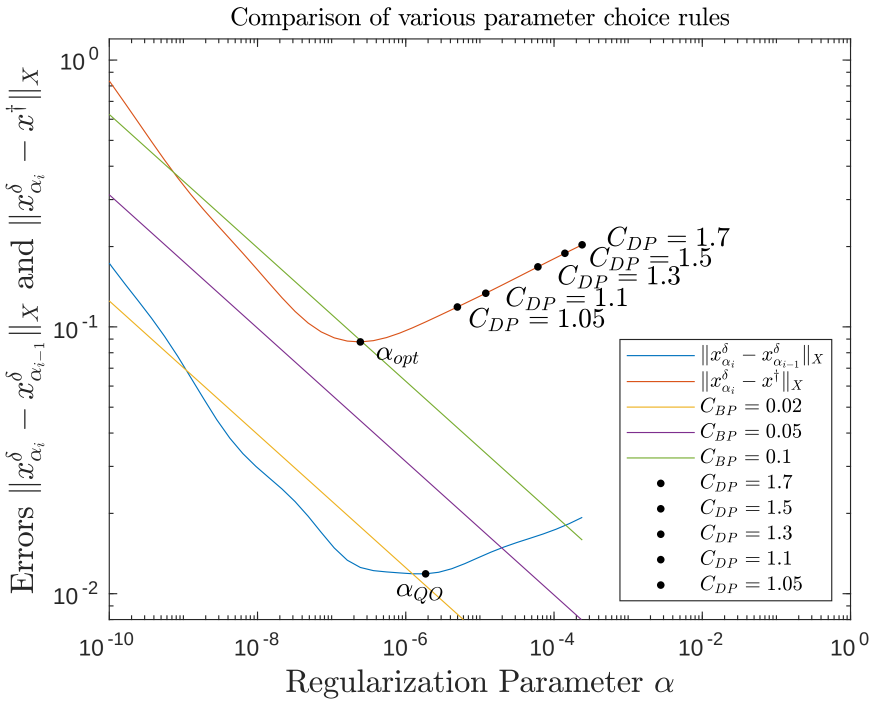

FIGURE 2. Exponential growth model with $x^\dagger(t) \equiv 1$; $(0 < t \leq 1)$ and $\delta = 0.0179$. Visualization of $\|x^\delta_{\alpha_{k+1}} - x^\delta_{\alpha_k}\|_X$ and $\|x^\delta_\alpha - x^\dagger\|_X$ as well as parameter choice using the balancing principle, discrepancy principle, quasi-optimality and $\alpha_{\rm opt}$ for various C_{BP} and C_{DP}.

Figure 2. Regularization error $\|x^\delta_\alpha - x^\dagger\|_X$ and the term $\|x^\delta_{\alpha_i} - x^\delta_{\alpha_{i-1}}\|_X$ are plotted for various regularization parameters on a log-log scale. Their respective minima are marked as parameter choices $\alpha_{\rm opt}$ and α_{QO}. The colored graphs correspond to the right hand term of the balancing principle (48) for different choices of C_{BP}. The parameter choice rule can then be interpreted in the following way: the balancing principle chooses the largest regularization parameter from the admissible set, such that the left hand term in (48) is less or equal to the right hand side. Visually speaking this means choosing the regularization parameter at the intersect or just below the intersect of the blue and colored lines, again depending on C_{BP}. Larger C_{BP} leads to larger regularization parameters and vice versa. The resulting parameters for

parameter choice using the discrepancy principle and the respective regularization error are also marked for various C_{DP}.

In this situation, with fixed noise level, we observe that the heuristic parameter choice α_{QO} performs surprisingly well. Parameter choice using the discrepancy principle follows our intuition: for smaller constants C_{DP} the regularization error decreases and vice versa. The success of the balancing principle highly depends on the choice C_{BP}. Although theoretical results on the choice of this constant exist it is difficult to chose this accordingly in practice.

Finally we highlight the differences between oversmoothing to non-oversmoothing penalties. We therefore remain in the same setting and consider regularized solutions $x^\dagger(t) \equiv 1$ $(0 < t \le 1)$ with $x^\dagger \in X_p$ $(0 < p < \frac{1}{2})$ (oversmoothing case) and $\hat{x}^\dagger(t) = -(t-\frac{1}{2})^2 + \frac{1}{4}$ $(0 < t \le 1)$ with $\hat{x}^\dagger \in X_1$ (non-oversmoothing case). We again minimize the Tikhonov functional (2) for various parameter choice rules. Understand that $\hat{x}^\dagger \in X_p$ for some $p > 1$ and therefore the penalty is not oversmoothing. The exact and regularized solution for various parameter choice are visualized in Figure 3. The left column considers $x^\dagger$, the right column $\hat{x}^\dagger$. Parameter choice $\alpha \approx 9.52e{-}09$ is chosen a priori and too small in both instances. Regularized solutions are displayed in the first row. We therefore see highly oscillating regularized solutions and an insufficient noise suppression. In the second row, $\alpha \approx 2.44e{-}07$ is the optimal parameter choice in the sense of (53) for the oversmoothing situation. Similarly, $\alpha \approx 2.12e{-}05$ (third row) is the optimal parameter choice for the non oversmoothing situation. We see that the first parameter choice leads to highly oscillating regularized solutions in the non oversmoothing case. Further, we obverse a phenomenon inherent to regularization with oversmoothing regularization (left): when comparing the parameter choices $\alpha \approx 2.44e{-}07$ and $\alpha \approx 2.12e{-}05$ it becomes evident that the regularized solution for the first parameter choice oscillates mildly, while the latter appears much smoother. This occurs, as regularized solution have to adhere to the boundary condition $x^\dagger(1) = 0$. Therefore a trade off between noise suppression and boundary condition occurs. This is in agreement with results in [9, 13]. The fourth column shows the regularized solutions for too large regularization parameters $\alpha \approx 1.60e{-}04$. Noise is effectively suppressed, but in both instances the regularized solutions are too smooth.

References

[1] A. B. Bakushinskij. Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion. *U.S.S.R. Comput. Math. Math. Phys.*, 24(4):181–182, 1984. Translation from *Zh. Vychisl. Mat Mat. Fiz.*, 24(8), 1258–1259, 1984 (Russian).

[2] D. Gerth, B. Hofmann, and C. Hofmann. Case studies and a pitfall for nonlinear variational regularization under conditional stability. In J. Cheng, S. Lu, and M. Yamamoto, editors, *Inverse Problems and Related Topics: Shanghai,*

China, October 12–14, 2018, Springer Proceedings in Mathematics & Statistics, Vol. 310, Chapter 9, pages 177–203. Springer Nature, Singapore, 2020.

[3] 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.

[4] C. W. Groetsch. *Inverse Problems in the Mathematical Sciences*. Vieweg Mathematics for Scientists and Engineers. Friedr. Vieweg & Sohn, Braunschweig, 1993.

[5] U. Hämarik and T. Raus. About the balancing principle for choice of the regularization parameter. *Numer. Funct. Anal. Opt.*, 30(9–10):951–970, 2009.

[6] B. Hofmann. A local stability analysis of nonlinear inverse problems. In D. Delaunay et al., editor, *Inverse Problems in Engineering - Theory and Practice*, pages 313–320. The American Society of Mechanical Engineers, New York, 1998.

[7] B. Hofmann and C. Hofmann. The impact of the discrepancy principle on the Tikhonov-regularized solutions with oversmoothing penalties. *mathematics - www.mdpi.com/journal/mathematics*, 8(3):331 (16pp), 2020.

[8] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer. A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Problems*, 23(3):987–1010, 2007.

[9] B. Hofmann and P. Mathé. Tikhonov regularization with oversmoothing penalty for non-linear ill-posed problems in Hilbert scales. *Inverse Problems*, 34(1):015007 (14pp), 2018.

[10] B. Hofmann and P. Mathé. A priori parameter choice in Tikhonov regularization with oversmoothing penalty for non-linear ill-posed problems. In J. Cheng, S. Lu, and M. Yamamoto, editors, *Inverse Problems and Related Topics: Shanghai, China, October 12–14, 2018*, Springer Proceedings in Mathematics & Statistics, Vol. 310, Chapter 8, pages 169–176. Springer Nature, Singapore, 2020.

[11] B. Hofmann, P. Mathé, and H. von Weizsäcker. Regularization in Hilbert space under unbounded operators and general source conditions. *Inverse Problems*, 25(11):115013 (15pp), 2009.

[12] B. Hofmann and R. Plato. On ill-posedness concepts, stable solvability and saturation. *J. Inverse Ill-Posed Probl.*, 26(2):287–297, 2018.

[13] B. Hofmann and R. Plato. Convergence results and low order rates for nonlinear Tikhonov regularization with oversmoothing penalty term. *Electronic Transactions on Numerical Analysis*, 53:313–328, 2020.

[14] T. Hohage. Regularization of exponentially ill-posed problems. *Numer. Funct. Anal. Optim.*, 21(3–4):439–464, 2000.

[15] A. S. Leonov. On the accuracy of Tikhonov regularizing algorithms and the quasi-optimal choice of regularization parameter. *Dokl. Akad. Nauk SSSR*, 321(3):460–465, 1991.

[16] O. V. Lepskiĭ. A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatnost. i Primenen.*, 35(3):459–470, 1990.

[17] P. Mathé. The Lepskiĭ principle revisited. *Inverse Problems*, 22(3):L11–L15, 2006.

[18] P. Mathé and B. Hofmann. How general are general source conditions? *Inverse Problems*, 24(1):015009 (5pp), 2008.

[19] P. Mathé and S. V. Pereverzev. Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19(3):789–803, 2003.

[20] F. Natterer. Error bounds for Tikhonov regularization in Hilbert scales. *Appl. Anal.*, 18(1-2):29–37, 1984.

[21] S. Pereverzev and E. Schock. On the adaptive selection of the parameter in regularization of ill-posed problems. *SIAM J. Numer. Anal.*, 43(5):2060–2076, 2005.

[22] R. Plato. The regularizing properties of multistep methods for first kind Volterra integral equations with smooth kernels. *Comput. Methods Appl. Math.*, 17(1):139–159, 2017.

[23] M. Pricop-Jeckstadt. Nonlinear Tikhonov regularization in Hilbert scales with balancing principle tuning parameter in statistical inverse problems. *Inverse Probl. Sci. Eng.*, 27(2):205–236, 2019.

[24] T. Raus and U. Hämarik. On the quasioptimal regularization parameter choices for solving ill-posed problems. *J. Inv. Ill-Posed Problems*, 15:419–439, 2007.

[25] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, New York, 2009.

[26] T. Schuster, B. Kaltenbacher, B. Hofmann, and K.S. Kazimierski. *Regularization Methods in Banach Spaces*. Walter de Gruyter, Berlin/Boston, 2012.

[27] A. N. Tikhonov and V. B. Glasko. The approximate solution of Fredholm integral equations of the first kind. *U.S.S.R. Comput. Math. Math. Phys.*, 4(3):236–247, 1964. Translation from *Zh. Vychisl. Mat Mat. Fiz.*, 4(3), 564–571, 1964 (Russian).

[28] E. Zeidler. *Nonlinear Functional Analysis and Its Applications - Nonlinear Monotone Operators*, volume II/B. Springer-Verlag, New York, 1990. Translated from the German.

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

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

Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstrasse 39, 10117 Berlin, Germany

Department Mathematik, Emmy-Noether-Campus, Universität Siegen, Walter-Flex-Str. 3, 57068 Siegen, Germany

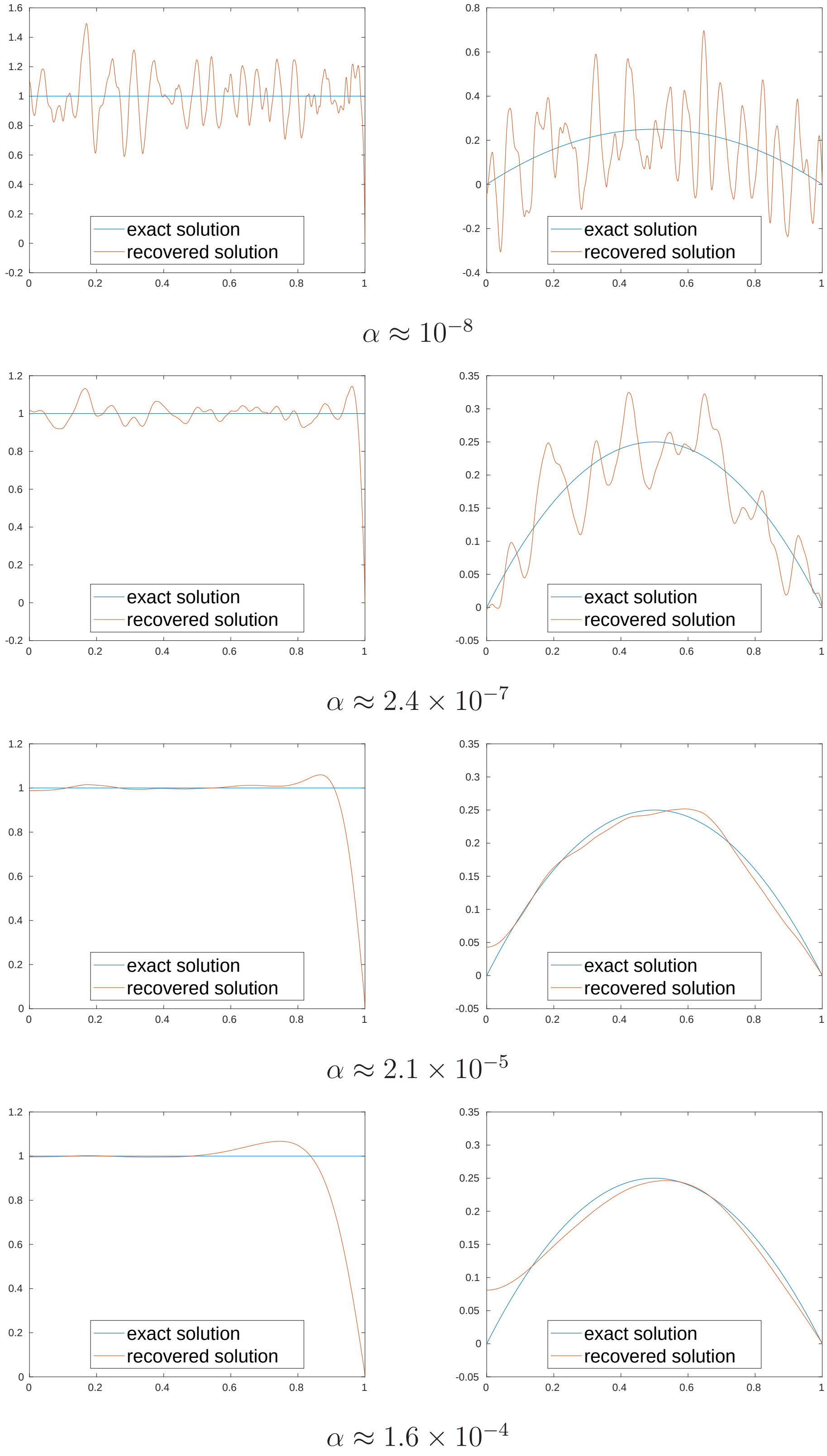

FIGURE 3. Exponential growth model with $x^\dagger(t) \equiv 1$; $(0 < t \le 1)$ (left) and $\hat{x}^\dagger(t) = -(t - \frac{1}{2})^2 + \frac{1}{4}$; $(0 < t \le 1)$, $\delta = 0.0179$. Regularized and exact solutions for various regularization parameters.