

PARAMETER CHOICE IN BANACH SPACE REGULARIZATION UNDER VARIATIONAL INEQUALITIES

BERND HOFMANN AND PETER MATHÉ

ABSTRACT. The authors study parameter choice strategies for Tikhonov regularization of nonlinear ill-posed problems in Banach spaces. The effectiveness of any parameter choice for obtaining convergence rates depends on the interplay of the solution smoothness and the nonlinearity structure, and it can be expressed concisely in terms of variational inequalities. Such inequalities are link conditions between the penalty term, the norm misfit and the corresponding error measure. The parameter choices under consideration include an a priori choice, the discrepancy principle as well as the Lepskii principle. For the convenience of the reader the authors review in an appendix a few instances where the validity of a variational inequality can be established.

1. INTRODUCTION

In the past years there was a significant progress with respect to the error analysis including convergence rates results for regularized solutions to inverse problems in Banach spaces. Such problems can be formulated as ill-posed operator equations

$$(1.1) \quad F(x) = y$$

with an (in general nonlinear) forward operator $F : \mathcal{D}(F) \subseteq X \rightarrow Y$, with domain $\mathcal{D}(F)$, and mapping between the Banach spaces X and Y with norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively. Equations of this type frequently occur in natural sciences, engineering, imaging, and finance (see e.g. [19] and [20, Chapter 1]). We denote by X^* and Y^* the corresponding dual spaces and by $\langle \cdot, \cdot \rangle_{X^* \times X}$ the dual pairing between X and X^* . In this paper, for constructing stable approximate solutions to (1.1) our focus is on the Tikhonov type regularization based on noisy data $y^\delta \in Y$ of the exact right-hand side $y \in F(\mathcal{D}(F))$ under the deterministic noise model

$$(1.2) \quad \|y^\delta - y\|_Y \leq \delta.$$

Date: April 11, 2012.

2010 Mathematics Subject Classification. 65J20, 47J06, 47A52, 49J40.

Precisely, we use for regularization parameters $\alpha > 0$ regularized solutions $x_\alpha^\delta \in \mathcal{D}(F)$, which are minimizers of

$$(1.3) \quad \mathcal{T}_\alpha^\delta(x) := \frac{1}{p} \|F(x) - y^\delta\|_Y^p + \alpha \Omega(x), \quad \text{subject to } x \in \mathcal{D}(F) \subseteq X,$$

with a *convex* penalty functional $\Omega : X \rightarrow [0, \infty]$ and some positive exponent $1 < p < \infty$. We suppose in the sequel that the standard assumptions on $F, \mathcal{D}(F)$, and Ω , made for the Tikhonov regularization in [10] and in the recent monographs [19, 20] are fulfilled. In particular, we assume that Ω is *stabilizing*, which means that for all $c \geq 0$ the sublevel sets

$$\mathcal{M}^\Omega(c) := \{x \in \mathcal{D}(F) : \Omega(x) \leq c\}$$

are sequentially pre-compact in a topology τ_X weaker than the norm topology of the Banach space X . In this case minimizers $x_\alpha^\delta \in \mathcal{D}(F)$ of $\mathcal{T}_\alpha^\delta$ exist for all $\alpha > 0$, and we refer to Section 2 for more details.

The objective in the following study is to control a prescribed non-negative error functional, say $E(x_\alpha^\delta, x^\dagger)$, measuring the deviation of the regularized solution x_α^δ from an Ω -minimizing solution, i.e., from a solution x^\dagger to (1.1) for which we have

$$\Omega(x^\dagger) = \min\{\Omega(x) : x \in \mathcal{D}(F), F(x) = y\}.$$

Typical examples of error measures would be the norm misfit $E(x, x^\dagger) = \|x - x^\dagger\|_X$ or a power $E(x, x^\dagger) = \|x - x^\dagger\|_X^q$ of that with exponents $1 < q < \infty$. Within the present context the Bregman distance

$$(1.4) \quad E(x, x^\dagger) = D_{\xi^\dagger}^\Omega(x, x^\dagger) := \Omega(x) - \Omega(x^\dagger) - \langle \xi^\dagger, x - x^\dagger \rangle_{X^* \times X}.$$

is often used, where we assume that $x \in \mathcal{D}(\Omega) := \{x \in X : \Omega(x) < \infty\}$ and that x^\dagger belongs to the Bregman domain

$$x^\dagger \in \mathcal{D}_B(\Omega) := \{\tilde{x} \in \mathcal{D}(\Omega) : \partial\Omega(\tilde{x}) \neq \emptyset\}.$$

Above, we denote by $\xi^\dagger \in \partial\Omega(x^\dagger) \subseteq X^*$ the subdifferential of the convex functional Ω at the point x^\dagger . The Bregman distance was introduced into the regularization theory by the study [5] in 2004, and henceforth this concept was adopted, refined and developed by many authors (cf., e.g., [9, 10, 14, 17, 18]).

The goal of the present paper is to study convergence rates of $E(x_\alpha^\delta, x^\dagger)$ as $\delta \rightarrow 0$ for several choices of the regularization parameter $\alpha = \alpha(\delta, y^\delta)$. The quality of any parameter choice (in terms of rates of convergence) will depend on the interplay of the following four relevant ingredients, as these are

- (i) the smoothness of the solution x^\dagger ,
- (ii) the structure of the forward operator F , and its domain $\mathcal{D}(F)$,
- (iii) properties of the functional Ω ,
- (iv) and the character of the error measure $E(\cdot, \cdot)$.

In this context, conditions are necessary that link the four factors. For E from (1.4) such conditions were presented in a rather general form in [10] as *variational inequalities*. Here we refer to the following variant (cf. [6, 7, 8]), which uses the concept of index functions. We call a function $\varphi : (0, \infty) \rightarrow (0, \infty)$ an index function if it is continuous, strictly increasing, and satisfies the limit condition $\lim_{t \rightarrow +0} \varphi(t) = 0$, see e.g. [11, 16].

Assumption VI (variational inequality). *We assume to have a constant $0 < \beta \leq 1$, and a concave index function φ such that*

$$(1.5) \quad \beta E(x, x^\dagger) \leq \Omega(x) - \Omega(x^\dagger) + \varphi(\|F(x) - F(x^\dagger)\|_Y) \quad \text{for all } x \in \mathcal{M}.$$

Remark 1. *The set \mathcal{M} of the validity of Assumption VI must be large enough such that it contains x^\dagger and all regularized solutions x_α^δ under consideration for $0 < \delta \leq \delta_{\max}$. This is for example the case if $\mathcal{M} = \mathcal{M}^\Omega(\Omega(x^\dagger) + c)$ for some $c > 0$.*

Moreover, there are good reasons to restrict in (1.5) to concave index functions. Namely, for index functions φ with $\lim_{t \rightarrow +0} \frac{\varphi(t)}{t} = 0$, including the family of strictly convex index functions, the variational inequality degenerates in the sense that $\Omega(x^\dagger) \leq \Omega(x)$ for all $x \in \mathcal{M}$ (see [7, Proposition 12.10], and for a special case [12, Proposition 4.3]). If $0 < \lim_{t \rightarrow +0} \frac{\varphi(t)}{t} < \infty$ then the situation is equivalent to the case $\varphi(t) = ct$, $c > 0$, in (1.5) (see [7, Proposition 12.11]), and for an index function φ with $\lim_{t \rightarrow +0} \frac{\varphi(t)}{t} \nearrow +\infty$ we can find a concave majorant index function that can be used in (1.5).

The outline is as follows. We present the general methodology of our approach in Section 2. Then we draw some consequences of the variational inequality (1.5) in form of inequalities in Section 3. Parts of these inequalities have been underestimated or even overlooked in past work. However, they will be essentially used in Section 4 to derive error bounds for several parameter choices. In an appendix we shall indicate how (1.5) may be derived for some linear and nonlinear problems.

2. METHODOLOGY AND A FUNDAMENTAL ERROR BOUND

The existence and behavior of Tikhonov minimizers x_α^δ was analyzed in several studies (cf., e.g., [10, 19, 20]). Under natural assumptions, stated there, Ω -minimizing solutions

$$x^\dagger \in \mathcal{D} := \mathcal{D}(F) \cap \mathcal{D}(\Omega) \neq \emptyset$$

exist whenever (1.1) has a solution which belongs to \mathcal{D} . Also, minimizers x_α^δ to the Tikhonov functional (1.3) exist for all data $y^\delta \in Y$ and regularization parameters $\alpha > 0$, and these are stable with respect to perturbations in the data for fixed α . Particularly relevant for our purpose is the following: For

any parameter choice $\alpha_* = \alpha_*(y^\delta, \delta)$ satisfying

$$(2.1) \quad \alpha_* \rightarrow 0 \quad \text{and} \quad \frac{\delta^p}{\alpha_*} \rightarrow 0 \quad \text{as} \quad \delta \rightarrow 0$$

we have convergence for both

$$(2.2) \quad \Omega(x_{\alpha_*}^\delta) \rightarrow \Omega(x^\dagger) \quad \text{and} \quad \|F(x_{\alpha_*}^\delta) - F(x^\dagger)\|_Y \rightarrow 0 \quad \text{as} \quad \delta \rightarrow 0.$$

Hence, all regularized solutions $x_{\alpha_*}^\delta$ for sufficiently small $\delta > 0$ belong to $\mathcal{M}^\Omega(\Omega(x^\dagger) + c)$, for some $c > 0$, and moreover if $\delta_n \rightarrow 0$ then the regularized solutions $x_{\alpha_*(\delta_n)}^{\delta_n}$ converge to x^\dagger in the weaker topology τ_X of X . This is a weak convergence in the sense of subsequences if the Ω -minimizing solution x^\dagger is not unique. For more details see, for example, [20, Section 4.1.2].

This gives rise to the following methodology: In view of the convergence as stated in (2.2) and the variational inequality (1.5) the following region of stability is of interest.

Definition (stability region). *Given $\delta > 0$ and a concave index function φ we let*

$$\mathcal{F}_{K,C}(\delta) := \left\{ x \in \mathcal{D} : \Omega(x) - \Omega(x^\dagger) \leq K\varphi(C\delta), \|F(x) - F(x^\dagger)\|_Y \leq C\delta \right\},$$

be a stability region for the Ω -minimizing solution $x^\dagger \in \mathcal{D}$ of (1.1) with constants $K > 0$, $C \geq 1$.

Notice that $\mathcal{F}_{K,C}(\delta) \subset \mathcal{M}^\Omega(\Omega(x^\dagger) + K\varphi(C\delta))$, such that minimizers which are pushed towards $\mathcal{F}_{K,C}(\delta)$ belong to specified sublevel sets, in agreement with the outline in the beginning of this section. Here the constants C and K do not depend on δ , however on the exponent $p > 1$, and on additional parameters used for the specific parameter choice.

This methodology immediately allows for the following very elementary but fundamental error bound.

Proposition 1. *Let x^\dagger obey Assumption VI for some set \mathcal{M} . If the approximate solution $x \in \mathcal{M}$ belongs to a stability region $\mathcal{F}_{K,C}(\delta)$ for some $K > 0$, $C \geq 1$ and $\delta > 0$ then*

$$(2.3) \quad E(x, x^\dagger) \leq C \frac{K+1}{\beta} \varphi(\delta).$$

Above, we used the fact that $\varphi(C\delta) \leq C\varphi(\delta)$ is valid for all concave index functions φ and all $\delta > 0$, $C \geq 1$.

The concept of stability region only controls the *excess penalty* $\Omega(x) - \Omega(x^\dagger)$, but not its modulus $|\Omega(x) - \Omega(x^\dagger)|$. In the proofs given below, we shall obtain the following strengthening. The parameter choices will push the approximate solutions towards the *convergence region*, given similarly to the Definition of the stability region as

Definition (convergence region).. For $\delta > 0$ and a concave index function φ let

$$\mathcal{F}_{K,C}^{\text{conv}}(\delta) := \left\{ x \in \mathcal{D} : \left| \Omega(x) - \Omega(x^\dagger) \right| \leq K\varphi(C\delta), \|F(x) - F(x^\dagger)\|_Y \leq C\delta \right\},$$

be a convergence region for the Ω -minimizing solution $x^\dagger \in \mathcal{D}$ of (1.1) with constants $K > 0$, $C \geq 1$.

Plainly, the inclusion $\mathcal{F}_{K,C}^{\text{conv}}(\delta) \subset \mathcal{F}_{K,C}(\delta)$ holds. But the *additional* requirement provides us with a rate of convergence for $\Omega(x_\alpha^\delta) \rightarrow \Omega(x^\dagger)$, a surplus which we kindly appreciate.

Proposition 2. Let x^\dagger obey Assumption VI for some set \mathcal{M} . If the approximate solution $x \in \mathcal{M}$ belongs to a convergence region $\mathcal{F}_{K,C}^{\text{conv}}(\delta)$ for some $K > 0$, $C \geq 1$ and $\delta > 0$ then, in addition to the assertion from Proposition 1, we have

$$\left| \Omega(x) - \Omega(x^\dagger) \right| \leq KC\varphi(\delta) \quad \text{and} \quad \|F(x) - F(x^\dagger)\|_Y \leq C\delta.$$

The above bounds quantify the convergence assertions from (2.2) under Assumption VI.

We shall exhibit this methodology for a natural *a priori* parameter choice as well as for the discrepancy principle, and a variant of the Lepskiĭ (balancing) principle.

3. PRELIMINARY ESTIMATES BASED ON THE VARIATIONAL INEQUALITY

Before discussing parameter choice in detail we shall draw some first conclusions from the validity of the variational inequality. Here we neglect the specific structure of the error functional $E(\cdot, \cdot)$, and we only use its non-negativity. Let x_α^δ be any minimizer of the Tikhonov functional $\mathcal{T}_\alpha^\delta$ from (1.3). The first observation is the following.

Lemma 1. Under Assumption VI we have for $\alpha > 0$ and $x_\alpha^\delta \in \mathcal{M}$ that

$$\begin{aligned} \Omega(x^\dagger) - \Omega(x_\alpha^\delta) &\leq \varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y), \text{ and} \\ \Omega(x_\alpha^\delta) - \Omega(x^\dagger) &\leq \frac{\delta^p}{p\alpha}. \end{aligned}$$

Proof. The first assertion is an immediate consequence of (1.5) taking into account that $\beta > 0$ and $E(x_\alpha^\delta, x^\dagger) \geq 0$. For the second we use the minimizing property to see that

$$\frac{\|F(x_\alpha^\delta) - y^\delta\|_Y^p}{p} + \alpha\Omega(x_\alpha^\delta) \leq \frac{\|F(x^\dagger) - y^\delta\|_Y^p}{p} + \alpha\Omega(x^\dagger) \leq \frac{\delta^p}{p} + \alpha\Omega(x^\dagger),$$

from which the assertion follows. \square

Another conclusion is less obvious.

Lemma 2. *Under Assumption VI we have for $\alpha > 0$ and $x_\alpha^\delta \in \mathcal{M}$ that*

$$\frac{\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y^p}{p} \leq 2^p \frac{\delta^p}{p} + \alpha 2^{p-1} \varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y).$$

Proof. We bound, by using the minimizing properties of $x_\alpha^\delta \in \mathcal{M}$, as

$$\begin{aligned} 0 &\leq \Omega(x_\alpha^\delta) - \Omega(x^\dagger) + \varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y) \\ &\leq \frac{1}{\alpha} \left(\frac{\delta^p}{p} - \frac{1}{p} \|F(x_\alpha^\delta) - y^\delta\|_Y^p \right) + \varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y). \end{aligned}$$

We also have the following lower bound

$$\frac{\|F(x_\alpha^\delta) - y^\delta\|_Y^p}{p} \geq \frac{1}{2^{p-1}} \frac{\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y^p}{p} - \frac{\delta^p}{p}.$$

Inserting this we see that

$$0 \leq 2 \frac{\delta^p}{p} - \frac{1}{2^{p-1}} \frac{\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y^p}{p} + \alpha \varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y),$$

which completes the proof. \square

The bound in Lemma 2 can be used on two ways. First, given a specific value of the parameter $\alpha > 0$ we can bound the norm misfit from above. Secondly, assuming that the norm misfit is larger than δ we can bound the value of the parameter α from below. Both consequences will prove important. In this context we introduce the function

$$(3.1) \quad \Phi_p(t) := \frac{t^p}{\varphi(t)}, \quad t > 0,$$

which is an index function for $p > 1$, since φ is concave.

Corollary 1. *Let α_* be given from*

$$(3.2) \quad \alpha_* := \Phi_p(\delta).$$

Then we have for $\alpha \leq \alpha_$ and $x_\alpha^\delta \in \mathcal{M}$ that*

$$\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y \leq 2(2+p)^{1/(p-1)} \delta.$$

Proof. If $\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y > \delta$ then we use the bound from Lemma 2 and the value for α_* to obtain

$$\begin{aligned}
\frac{\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y^p}{p} &\leq 2^p \frac{\delta^p}{p} + \alpha 2^{p-1} \varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y) \\
&\leq 2^p \frac{\delta^p}{p} + \alpha_* 2^{p-1} \varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y) \\
&= 2^p \frac{\delta^p}{p} + \delta^p 2^{p-1} \frac{\varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y)}{\varphi(\delta)} \\
&\leq \frac{\delta^p}{p} \left(2^p + p 2^{p-1} \frac{\varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y)}{\varphi(\delta)} \right) \\
&\leq 2^{p-1} (2 + p) \frac{\delta^p}{p} \frac{\varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y)}{\varphi(\delta)} \\
&\leq 2^{p-1} (2 + p) \frac{\delta^p}{p} \frac{\delta^{-1} \|F(x_\alpha^\delta) - F(x^\dagger)\|_Y \varphi(\delta)}{\varphi(\delta)} \\
&= 2^{p-1} (2 + p) \frac{\delta^{p-1}}{p} \|F(x_\alpha^\delta) - F(x^\dagger)\|_Y.
\end{aligned}$$

Because of $2(2+p)^{1/(p-1)} > 1$ the bound given in the corollary is also valid for $\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y \leq \delta$. \square

Corollary 2. *Let $\tau > 1$. Suppose that the parameter $\alpha > 0$ is chosen such that $x_\alpha^\delta \in \mathcal{M}$ and the residual obeys $\|F(x_\alpha^\delta) - y^\delta\|_Y > \tau\delta$. Then we have*

$$(3.3) \quad \alpha \geq \frac{1}{p 2^{p-1}} \frac{\tau^p - 1}{\tau^p + 1} \Phi_p((\tau - 1)\delta).$$

Proof. Using the first assertion in Lemma 1 and that $x_\alpha^\delta \in \mathcal{M}$ is a minimizer of $\mathcal{T}_\alpha^\delta$ we have under the assumption made on α that

$$\frac{\tau^p \delta^p}{p} \leq \frac{\|F(x_\alpha^\delta) - y^\delta\|_Y^p}{p} \leq \frac{\delta^p}{p} + \alpha \left(\Omega(x^\dagger) - \Omega(x_\alpha^\delta) \right) \leq \frac{\delta^p}{p} + \alpha \varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y).$$

Thus

$$\frac{\delta^p}{p} \leq \frac{1}{\tau^p - 1} \alpha \varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y).$$

We plug this into the bound in Lemma 2, and we temporarily abbreviate $t_\alpha := \|F(x_\alpha^\delta) - F(x^\dagger)\|_Y$. We thus obtain that

$$\begin{aligned}
\frac{t_\alpha^p}{p} &\leq 2^p \frac{\delta^p}{p} + \alpha 2^{p-1} \varphi(t_\alpha) \leq 2^p \frac{1}{\tau^p - 1} \alpha \varphi(t_\alpha) + \alpha 2^{p-1} \varphi(t_\alpha) \\
&= \left(\frac{2^p}{\tau^p - 1} + 2^{p-1} \right) \alpha \varphi(t_\alpha) = 2^{p-1} \frac{\tau^p + 1}{\tau^p - 1} \alpha \varphi(t_\alpha).
\end{aligned}$$

Since $\tau\delta \leq t_\alpha + \delta$, we arrive, using the function Φ_p from (3.1), at

$$\Phi_p((\tau - 1)\delta) \leq \Phi_p(t_\alpha) \leq p 2^{p-1} \frac{\tau^p + 1}{\tau^p - 1} \alpha,$$

and the proof is complete. \square

4. PARAMETER CHOICE

The objective of this study is the error analysis of several parameter choice strategies, commonly used in regularization theory. This concerns a priori strategies, i.e., when $\alpha_* = \alpha_*(\delta)$ does not depend on the given data y^δ , as well as a posteriori strategies, when $\alpha_* = \alpha_*(y^\delta, \delta)$.

4.1. A natural a priori parameter choice. Several a priori parameter choices can be found in earlier studies (cf. [4, 6]). Here we present an intuitive parameter choice, which was obtained in [8] by means of tools from convex analysis. Our approach, however, is elementary and directly based on Assumption VI. In addition we show that this parameter choice pushes the approximate solution $x_{\alpha_*}^\delta$ into a specific set $\mathcal{F}_{K,C}^{\text{conv}}(\delta)$. We recall the index function Φ_p from (3.1).

Theorem 1. *Suppose that x^\dagger obeys Assumption VI for some concave index function φ . Let $\alpha_* = \alpha_*(\delta) = \Phi_p(\delta)$ be chosen a priori.*

- (i) *If $x_{\alpha_*}^\delta \in \mathcal{M}$ then $x_{\alpha_*}^\delta \in \mathcal{F}_{K,C}^{\text{conv}}(\delta)$ with $K = 1$ and $C = 2(2+p)^{1/(p-1)}$.*
- (ii) *If $x_{\alpha_*}^\delta \in \mathcal{M}$ for all $0 < \delta \leq \delta_{\max}$ and some $\delta_{\max} > 0$, then this a priori parameter choice yields the convergence rates*

$$E(x_{\alpha_*}^\delta, x^\dagger) = \mathcal{O}(\varphi(\delta)), \quad \|F(x_{\alpha_*}^\delta) - F(x^\dagger)\|_Y = \mathcal{O}(\delta), \quad \text{and}$$

$$|\Omega(x_{\alpha_*}^\delta) - \Omega(x^\dagger)| = \mathcal{O}(\varphi(\delta)) \quad \text{as } \delta \rightarrow 0.$$

Proof. Corollary 1 provides us with a bound of the norm misfit $\|F(x_{\alpha_*}^\delta) - F(x^\dagger)\|_Y$. In view of the first assertion of Lemma 1 this also bounds $\Omega(x^\dagger) - \Omega(x_{\alpha_*}^\delta)$, appropriately. Furthermore, from the second assertion of Lemma 1 we have that $\Omega(x_{\alpha_*}^\delta) - \Omega(x^\dagger) \leq \delta^p/(p\alpha_*) \leq \varphi(\delta)$, by the choice of α_* . The convergence rates in Item (ii) are a consequence of Propositions 1 & 2. \square

Remark 2. *An inspection of the proofs in Section 3 shows that the first two convergence rates in Theorem 1 use only the implication*

$$x_{\alpha_*}^\delta \in \mathcal{M} \implies x_{\alpha_*}^\delta \in \mathcal{F}_{K,C}(\delta), \quad \text{for sufficiently small } \delta > 0.$$

Only for the third Ω -rate the membership $x_{\alpha_}^\delta \in \mathcal{F}_{K,C}^{\text{conv}}(\delta)$ is required. The a priori parameter choice from (3.2) satisfies the condition (2.1).*

Moreover, it corresponds to the well-known a priori parameter choice for linear problems in Hilbert spaces, and this shows how φ is related to smoothness. Indeed, taking into account that $E(x, x^\dagger) = \|x - x^\dagger\|_X^2$ measures the squared error, we obtain a rate $\|x_{\alpha_*}^\delta - x^\dagger\|_X = \mathcal{O}(\sqrt{\varphi(\delta)})$. Presuming that this is the optimal rate, we ‘guess’ the relation $\sqrt{\varphi(t)} \equiv \psi(\Theta_\psi^{-1}(t))$ by setting $\Theta_\psi(t) := \sqrt{t}\psi(t)$, $t > 0$, where ψ should be the smoothness in the source condition for x^\dagger , meaning that $x^\dagger = \psi(A^*A)v$, $\|v\|_X \leq 1$. Taking this for

granted we see that

$$\sqrt{\alpha_*} \sim \frac{\delta}{\sqrt{\varphi(\delta)}} = \frac{\delta}{\psi(\Theta_\psi^{-1}(\delta))} = \frac{\Theta_\psi(\Theta_\psi^{-1}(\delta))}{\psi(\Theta_\psi^{-1}(\delta))} = \sqrt{\Theta_\psi^{-1}(\delta)},$$

and hence $\Theta_\psi(\alpha_*) = \delta$, which is the ‘ordinary’ a priori parameter choice in linear problems in Hilbert space under given source condition $x^\dagger = \psi(A^*A)v$, $\|v\|_X \leq 1$ (cf. [16]).

4.2. A posteriori parameter choice. For the a posteriori parameter choice we restrict the selection of the regularization parameter to a discrete exponential grid. Precisely, we select $0 < q < 1$, choose a largest parameter α_0 and consider the set

$$(4.1) \quad \Delta_q := \{\alpha_j : \alpha_j := q^j \alpha_0, \quad j = 1, 2, \dots\}.$$

Above, the parameter q determines the roughness of searching for the optimal parameter. If q is close to one than we scan for the optimal parameter accurately, however, many trials may be necessary to find the best candidate. If q is small than we roughly scan for the parameter, at a dispense of loosing accuracy.

4.2.1. Discrepancy principle. Previous use of the discrepancy principle for nonlinear problems in Banach space was restrictive; a stronger version was used. Precisely, for two parameters $1 < \tau_1 < \tau_2 < \infty$ the chosen parameter α_* was assumed to fulfill $\tau_1 \delta \leq \|F(x_{\alpha_*}^\delta) - y^\delta\|_Y \leq \tau_2 \delta$ (cf. [1, 7]). We shall call this the *strong discrepancy principle*. It is not clear that this is always possible, and it was mentioned in [20, Chapt. 4] that for nonlinear operators F there may be a duality gap due to the non-convexity of the functional $\mathcal{T}_\alpha^\delta$ which prevents the use of this strong discrepancy principle. Here we establish the use of the in general applicable classical discrepancy principle for which the variational inequality in Assumption VI is strong enough to ensure convergence rates.

Theorem 2. *Let $\tau > 1$ be given. Let $\alpha_* \in \Delta_q$, $\alpha_* < \alpha_1$ (no immediate stop) be chosen, according to the discrepancy principle, as the largest parameter within Δ_q for which*

$$\|F(x_\alpha^\delta) - y^\delta\|_Y \leq \tau \delta.$$

Suppose that x^\dagger obeys Assumption VI for some concave index function φ . Then the following holds true.

- (i) *If $x_\alpha^\delta \in \mathcal{M}$, $\alpha \geq \alpha_*$, then $x_{\alpha_*}^\delta \in \mathcal{F}_{K,C}^{\text{conv}}(\delta)$ with*

$$K = \max \left\{ \frac{1}{2q} \left(\frac{2}{\tau-1} \right)^p \frac{\tau^p + 1}{\tau^p - 1}, 1 \right\} \quad \text{and} \quad C = \tau + 1.$$

- (ii) *If $x_{\alpha_*}^\delta \in \mathcal{M}$ for all $0 < \delta \leq \delta_{\max}$ and some $\delta_{\max} > 0$, then this a posteriori parameter choice yields the convergence rates*

$$E(x_{\alpha_*}^\delta, x^\dagger) = \mathcal{O}(\varphi(\delta)), \quad \|F(x_{\alpha_*}^\delta) - F(x^\dagger)\|_Y = \mathcal{O}(\delta), \quad \text{and}$$

$$|\Omega(x_{\alpha_*}^\delta) - \Omega(x^\dagger)| = \mathcal{O}(\varphi(\delta)) \quad \text{as } \delta \rightarrow 0.$$

Proof. We first bound

$$\|F(x_{\alpha_*}^\delta) - F(x^\dagger)\|_Y \leq \|F(x_{\alpha_*}^\delta) - y^\delta\|_Y + \|F(x^\dagger) - y^\delta\|_Y \leq \tau\delta + \delta = (\tau + 1)\delta.$$

Under Assumption VI this also gives $\Omega(x^\dagger) - \Omega(x_{\alpha_*}^\delta) \leq \varphi((\tau + 1)\delta)$, cf. Lemma 1. For bounding the negative Ω -difference we use Corollary 2 as follows. The (previous) parameter α_*/q fulfills the assumption from Corollary 2, and we bound from below as

$$\alpha_*/q \geq \frac{1}{p2^{p-1}} \frac{\tau^p - 1}{\tau^p + 1} \Phi_p((\tau - 1)\delta).$$

This, together with the second assertion of Lemma 1, yields

$$\begin{aligned} \Omega(x_{\alpha_*}^\delta) - \Omega(x^\dagger) &\leq \frac{\delta^p}{p\alpha_*} \leq \frac{2^{p-1} \left(\frac{\tau^p+1}{\tau^p-1}\right) \delta^p}{q \Phi_p((\tau-1)\delta)} \\ &= \frac{1}{2q} \left(\frac{2}{\tau-1}\right)^p \frac{\tau^p+1}{\tau^p-1} \varphi((\tau-1)\delta) \\ &\leq \frac{1}{2q} \left(\frac{2}{\tau-1}\right)^p \frac{\tau^p+1}{\tau^p-1} \varphi((\tau+1)\delta). \end{aligned}$$

The convergence rates in Item (ii) are again consequences of Propositions 1 and 2. \square

Remark 3. *We emphasize that here we bounded*

$$\Omega(x_{\alpha_*}^\delta) - \Omega(x^\dagger) \leq \frac{1}{2q} \left(\frac{2}{\tau-1}\right)^p \frac{\tau^p+1}{\tau^p-1} \varphi((\tau+1)\delta),$$

whereas the strong discrepancy principle yields $\Omega(x_{\alpha_}^\delta) - \Omega(x^\dagger) \leq 0$, which seems to be chicken-hearted, and this points at the limitations of this strong principle.*

Remark 4. *We shall conclude this section with a discussion on immediate stop of the discrepancy principle. Notice that in this principle we start from the largest value α_1 . In fact, it may happen that for $\alpha := \alpha_1$ the assumption $\|F(x_{\alpha_1}^\delta) - y^\delta\|_Y \leq \tau\delta$ is already fulfilled. In this case, the above proof will not work, and the question is, whether there is still the rate $\mathcal{O}(\varphi(\delta))$ to be observed. This can indeed be proved for linear problems in Hilbert space (cf. [3] for a recent treatment): For Tikhonov regularization in Hilbert space this corresponds to the case of having small data, because (for linear problems in Hilbert space) $x_\alpha^\delta \rightarrow 0$ as $\alpha \rightarrow \infty$, and hence immediate stop refers to $\|y^\delta\|_Y \leq \tau\delta$. Within the present context we make the following observation. If the discrepancy bound $\tau\delta$ holds, then the error bound $E(x_{\alpha_1}^\delta, x^\dagger)$ holds if only $\Omega(x_{\alpha_1}^\delta) - \Omega(x^\dagger)$ is small. A look at the second bound given in Lemma 1 reveals that a bound $\delta^p/(p\alpha_1)$ is valid. So, the desired overall error bound holds provided that $\delta^p/(p\alpha_1) \leq \bar{K}\varphi(\delta)$, or equivalently, by using the index function Φ_p from (3.1), that*

$$(4.2) \quad \Phi_p(\delta) \leq \bar{K} p \alpha_1.$$

Plainly, for each solution x^\dagger there is δ_0 such that $\Phi_p(\delta) \leq \Phi_p(\delta_0) \leq p\alpha_1$. Thus, for $0 < \delta \leq \delta_0$ we can bound

$$\Omega(x_{\alpha_1}^\delta) - \Omega(x^\dagger) \leq \frac{\delta^p}{p\alpha_1} = \frac{\delta^p}{\Phi_p(\delta)} = \varphi(\delta).$$

This shows that if the initial value α_0 is chosen large enough then immediate stop yields an error bound of the form $\mathcal{O}(\varphi(\delta))$. However, for any particular instance x^\dagger at hand we cannot verify whether (4.2) holds, since φ is not known to us.

4.2.2. The Lepskiĭ principle. The Lepskiĭ (balancing) principle is studied here for the first time within the context of nonlinear equations in Banach space regularization. However, it was used for nonlinear equations in Hilbert space, and we refer to [2]. Actually, by its very construction this parameter choice is not sensitive to the problem at hand, a generic formulation for this principle was given in [15]. This principle requires that the error functional is a metric, and we assume this within the present section without further mentioning. We will need the following fact.

Lemma 3. *Suppose that x^\dagger obeys Assumption VI, and that the parameter α_{AP} is given as in Theorem 1. Then for all $\alpha_m \leq \alpha \leq \alpha_{AP}$ and for $x_\alpha^\delta \in \mathcal{M}$ we have that*

$$\beta E(x_\alpha^\delta, x^\dagger) \leq \left(\frac{1}{p} + 2(2+p)^{1/(p-1)} \right) \frac{\delta^p}{\alpha}.$$

Proof. Under Assumption VI, and using Lemma 1 and Corollary 1 we see that

$$\begin{aligned} \beta E(x_\alpha^\delta, x^\dagger) &\leq \Omega(x_\alpha^\delta) - \Omega(x^\dagger) + \varphi(\|F(x_\alpha^\delta) - x^\dagger\|_Y) \\ &\leq \frac{\delta^p}{p\alpha} + \varphi(C_p \delta) \leq \frac{\delta^p}{p\alpha} + C_p \varphi(\delta) \\ &= \frac{\delta^p}{p\alpha} + C_p \frac{\delta^p}{\alpha} = \left(\frac{1}{p} + C_p \right) \frac{\delta^p}{\alpha}, \end{aligned}$$

where we abbreviated $C_p := 2(2+p)^{1/(p-1)}$, the constant from Corollary 1. \square

We want to use the Lepskiĭ principle, as this is outlined in [15], by using a multiple of the decreasing function $\alpha \rightarrow \delta^p/\alpha$, and we let

$$(4.3) \quad \Psi(\alpha) := \frac{1 + pC_p}{p\beta} \frac{\delta^p}{\alpha}, \quad \alpha > 0.$$

From [15, Prop. 1] we draw the following conclusion.

Theorem 3. *Fix $m > 1$ (large) and let $\alpha_* \in \Delta_q$ be the largest parameter α for which*

$$E(x_{\alpha'}^\delta, x_\alpha^\delta) \leq 2\Psi(\alpha'), \quad \text{for all } \alpha' \in \Delta_q, \alpha_m \leq \alpha' < \alpha.$$

Moreover, suppose that x^\dagger obeys Assumption VI for some concave index function φ . If $x_\alpha^\delta \in \mathcal{M}$ for all $\alpha_m \leq \alpha \leq \alpha_*$ and if $E(x_{\alpha_m}^\delta, x^\dagger) \leq \Psi(\alpha_m)$, then

$$E(x_{\alpha_*}^\delta, x^\dagger) \leq 3 \frac{1 + pC_p}{pq\beta} \varphi(\delta).$$

Proof. As in [15, Prop. 1] we introduce the parameter

$$\alpha_+ := \max \left\{ \alpha : E(x_{\alpha'}^\delta, x^\dagger) \leq \Psi(\alpha'), \quad \alpha_m \leq \alpha' \leq \alpha \right\}.$$

(Caution: the notation in [15] differs from here, and some care is needed to transfer the results.) Let $\alpha_{AP} = \Phi_p(\delta)$ be the a priori choice from Theorem 1. If $\alpha_{AP} \in \Delta_q$ then Lemma 3 yields that $\alpha_+ \geq \alpha_{AP}$. Otherwise, we consider the index $k \leq m$ for which $\alpha_k < \alpha_{AP} \leq \alpha_k/q$, which results in $\alpha_+ \geq \alpha_k$. Proposition 1 in [15] states that $E(x_{\alpha_*}^\delta, x^\dagger) \leq 3\Psi(\alpha_+)$. Thus in either case this yields

$$E(x_{\alpha_*}^\delta, x^\dagger) \leq 3\Psi(\alpha_+) \leq 3\Psi(\alpha_k) = \frac{3}{q}\Psi(\alpha_k/q) \leq \frac{3}{q}\Psi(\alpha_{AP}) \leq \frac{1 + pC_p}{pq\beta} \varphi(\delta),$$

which completes the proof. \square

Remark 5. The parameter choice à la Lepskiĭ provides us with an error bound, which is obtained regardless whether the approximating $x_{\alpha_*}^\delta$ belongs to some set $\mathcal{F}_{K,C}(\delta)$. In fact, the only information which can be deduced from Theorem 3 is the following lower bound for α_* : Actually, in [15, Prop. 2.1] the information is given that $\alpha_* \geq \alpha_+$, such that $\alpha_* \geq \alpha_+ \geq q\alpha_{AP} = q\Phi_p(\delta)$. This bounds the excess penalty $\Omega(x_{\alpha_*}^\delta) - \Omega(x^\dagger) \leq \varphi(\delta)/q$. However, it is not clear whether the discrepancy $\|F(x_{\alpha_*}^\delta) - y^\delta\|_Y$ is of the order δ .

Remark 6. For the Lepskiĭ principle we have to start with the smallest chosen value α_m , and in the formulation of Theorem 3 we assumed that $E(x_{\alpha_m}^\delta, x^\dagger) \leq \Psi(\alpha_m)$. Since $\Psi(\alpha_m)$ is known to the user, some exogenous knowledge about the expected error size may allow to adjust for the choice of m , and hence of α_m . However, we cannot verify this condition, based on information of δ and the given data y^δ . So, if the Lepskiĭ principle stops immediately, one should decrease the initial value α_m until this will not be the case.

Remark 7. The above application of the Lepskiĭ principle does not include the Bregman distance as error measure $E(x, x^\dagger) := D_{\xi^\dagger}^\Omega(x, x^\dagger)$, because this is not a metric, in general. However, if the Bregman distance is q -coercive, $D_{\xi^\dagger}^\Omega(x, x^\dagger) \geq c \frac{\|x - x^\dagger\|_X^q}{q}$, then the validity of a variational inequality for $D_{\xi^\dagger}^\Omega(x, x^\dagger)$ implies the one for $\|x - x^\dagger\|_X^q/q$, and we can apply the Lepskiĭ parameter choice to the differences $x_\alpha^\delta - x_{\alpha'}^\delta$, i.e., test whether $\|x_\alpha^\delta - x_{\alpha'}^\delta\|_X \leq (2 \frac{1+pC_p}{p\beta})^{1/q} \delta^{p/q} / (\alpha')^{1/q}$ for $\alpha' \leq \alpha$.

APPENDIX A. VALIDITY OF ASSUMPTION VI

For the convenience of the reader we briefly sketch some approaches to show how variational inequalities occur. We suppose that the mapping $F : \mathcal{D}(F) \subseteq X \rightarrow Y$ with some convex domain $\mathcal{D}(F)$ has a one-sided directional derivative at x^\dagger given as a bounded linear operator $F'(x^\dagger) : X \rightarrow Y$ such that

$$(A.1) \quad \lim_{t \rightarrow +0} \frac{1}{t} \left(F(x^\dagger + t(x - x^\dagger)) - F(x^\dagger) \right) = F'(x^\dagger)(x - x^\dagger), \quad x \in \mathcal{D}(F).$$

For the application of several variational inequalities of type (1.5) that we will derive below, in the context of Tikhonov regularization, one still has to show that the minimizers x_α^δ of (1.3) belong to the domain \mathcal{M} of validity of such inequality. This requires special attention, and we do not tackle this question. We leave details to the indicated original references (see also [20, Sections 3.2 and 4.2]).

A.1. Benchmark case. Here we assume that $x^\dagger \in \mathcal{D}_B(\Omega)$ and the subdifferential ξ^\dagger fulfills the benchmark source condition

$$(A.2) \quad \xi^\dagger = F'(x^\dagger)^* v \in \partial\Omega(x^\dagger), \quad \text{for some } v \in Y^*.$$

Such information allows us to bound

$$\begin{aligned} \langle \xi^\dagger, x^\dagger - x \rangle_{X^* \times X} &= \langle (F'(x^\dagger))^* v, x^\dagger - x \rangle_{X^* \times X} = \langle v, F'(x^\dagger)(x^\dagger - x) \rangle_{Y^* \times Y} \\ &\leq \|v\|_{Y^*} \|F'(x^\dagger)(x - x^\dagger)\|_Y. \end{aligned}$$

After adding the term $\Omega(x) - \Omega(x^\dagger)$ on both sides this yields that

$$(A.3) \quad D_{\xi^\dagger}^\Omega(x, x^\dagger) \leq \Omega(x) - \Omega(x^\dagger) + \|v\|_{Y^*} \|F'(x^\dagger)(x - x^\dagger)\|_Y, \quad x \in \mathcal{D}(F).$$

Remark 8. We highlight the special case when X is a Hilbert space and $\Omega(x) = \|x\|_X^2$. Then $D_{\xi^\dagger}^\Omega(x, x^\dagger) = \|x - x^\dagger\|_X^2$ (cf. [19, Example 3.18]), and (A.3) implies

$$\|x - x^\dagger\|_X^2 \leq \|x\|_X^2 - \|x^\dagger\|_X^2 + \|v\|_{Y^*} \|F'(x^\dagger)(x - x^\dagger)\|_Y, \quad x \in \mathcal{D}(F).$$

It was emphasized in [6, Chapter 13] that within the Hilbert space setting solution smoothness can always be expressed by variational inequalities (1.5) with general index functions φ .

Inequality (A.3) results in a variational inequality for bounded linear operators

$F = A : X \rightarrow Y$, because then $A = F'(x^\dagger)$, with $\beta = 1$, $E(x, x^\dagger) = D_{\xi^\dagger}^\Omega(x, x^\dagger)$ and $\varphi(t) = \|v\|_{Y^*} t$, $t > 0$ on $\mathcal{M} = X$.

If the mapping F is nonlinear then we may use certain *structure of non-linearity* to bound $\|F'(x^\dagger)(x - x^\dagger)\|_Y$ in terms of $\|F(x^\dagger) - F(x)\|_Y$, and the validity of such structural conditions requires additional assumptions, which we will not discuss here. In its simplest form such condition is given as

$$(A.4) \quad \|F'(x^\dagger)(x - x^\dagger)\|_Y \leq \eta \sigma(\|F(x) - F(x^\dagger)\|_Y), \quad x \in \mathcal{M},$$

for some concave index function σ and constant $\eta > 0$ on some set $\mathcal{M} \subset \mathcal{D}(F)$ (cf. [4]). In this case (A.3) provides us with a variational inequality on \mathcal{M} with $\beta = 1$, $E(x, x^\dagger) = D_{\xi^\dagger}^\Omega(x, x^\dagger)$ and $\varphi(t) = \eta \|v\|_{Y^*} \sigma(t)$, $t > 0$.

An alternative structural condition is given in the form

$$(A.5) \quad \|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\|_Y \leq \eta D_{\xi^\dagger}^\Omega(x, x^\dagger), \quad x \in \mathcal{M},$$

again for some set $\mathcal{M} \subset \mathcal{D}(F)$, (cf., e.g., [10, 18]). This allows us to bound

$$\|F'(x^\dagger)(x - x^\dagger)\|_Y \leq \eta D_{\xi^\dagger}^\Omega(x, x^\dagger) + \|F(x) - F(x^\dagger)\|_Y, \quad x \in \mathcal{M}.$$

Then (A.3) implies a variational inequality (1.5) under

$$(A.6) \quad \eta \|v\|_{Y^*} < 1$$

with $0 < \beta = 1 - \eta \|v\|_{Y^*} \leq 1$, $E(x, x^\dagger) = D_{\xi^\dagger}^\Omega(x, x^\dagger)$ and $\varphi(t) = \|v\|_{Y^*} t$, $t > 0$ on \mathcal{M} . This occurring smallness condition (A.6) indicates that (A.5) is a weaker nonlinearity condition compared with (A.4).

A.2. Violation of the benchmark. If the assumption (A.2) is violated then we may use the *method of approximate source conditions* (cf. [4, 9]) to derive variational inequalities. To this end we need additionally that the distance function

$$d_{\xi^\dagger}(R) := \inf\{\|\xi^\dagger - \xi\|_{X^*} : \xi = F'(x^\dagger)^* v, v \in Y^*, \|v\|_{Y^*} \leq R\}, \quad R > 0,$$

is nonincreasing and obeys the limit condition $d_{\xi^\dagger}(R) \rightarrow 0$ as $R \rightarrow \infty$. As mentioned in [4] this is the case when $F'(x^\dagger)^{**} : X^{**} \rightarrow Y^{**}$ is injective. Additionally this approach presumes that the Bregmann distance is q -coercive, i.e., that

$$(A.7) \quad D_{\xi^\dagger}^\Omega(x, x^\dagger) \geq c_q \|x - x^\dagger\|_X^q \quad \text{for all } x \in \mathcal{M},$$

is satisfied for some exponent $2 \leq q < \infty$ and a corresponding constant $c_q > 0$. Such assumption is for example fulfilled if $\Omega(x) := \|x\|_X^q$ and X is a q -convex Banach space.

Then, for every $R > 0$ one can find elements $v_R \in Y^*$ and $u_R \in X^*$ such that

$$\xi^\dagger = \left(F'(x^\dagger)\right)^* v_R + u_R \quad \text{with} \quad \|v_R\|_{Y^*} = R, \quad \|u_R\|_{X^*} \leq d_{\xi^\dagger}(R),$$

and we can estimate for all $R > 0$ and $x \in \mathcal{M}$ as

$$\begin{aligned} -\langle \xi^\dagger, x - x^\dagger \rangle_{X^* \times X} &= -\langle (F'(x^\dagger))^* v_R + u_R, x - x^\dagger \rangle_{X^* \times X} \\ &= -\langle v_R, F'(x^\dagger)(x - x^\dagger) \rangle_{Y^* \times Y} + \langle u_R, x^\dagger - x \rangle_{X^* \times X} \\ &\leq R \|F'(x^\dagger)(x - x^\dagger)\|_Y + d_{\xi^\dagger}(R) \|x - x^\dagger\|_X. \end{aligned}$$

Adding, as before, the difference $\Omega(x) - \Omega(x^\dagger)$ on both sides gives

$$(A.8) \quad D_{\xi^\dagger}^\Omega(x, x^\dagger) \leq \Omega(x) - \Omega(x^\dagger) + R \|F'(x^\dagger)(x - x^\dagger)\|_Y + d_{\xi^\dagger}(R) \|x - x^\dagger\|_X, \quad x \in \mathcal{M}.$$

Using the q -coercivity (A.7) we see that

$$d_{\xi^\dagger}(R) \|x - x^\dagger\|_X \leq c_q^{-1/q} d_{\xi^\dagger}(R) \left(D_{\xi^\dagger}^\Omega(x, x^\dagger) \right)^{1/q}.$$

An application of Young's inequality yields

$$d_{\xi^\dagger}(R) \|x - x^\dagger\|_X \leq \frac{1}{q} D_{\xi^\dagger}^\Omega(x, x^\dagger) + \frac{c_q^{-q^*/q}}{q^*} (d_{\xi^\dagger}(R))^{q^*}.$$

Plugging this into (A.8) we obtain (with $\beta = 1 - 1/q$) that
(A.9)

$$\beta D_{\xi^\dagger}^\Omega(x, x^\dagger) \leq \Omega(x) - \Omega(x^\dagger) + R \|F'(x^\dagger)(x - x^\dagger)\|_Y + \frac{c_q^{-q^*/q}}{q^*} (d_{\xi^\dagger}(R))^{q^*}.$$

The term $\|F'(x^\dagger)(x - x^\dagger)\|_Y$ may be treated under structural conditions, used before in the benchmark case. To avoid this step we confine ourselves to the linear case $F'(x^\dagger) = A$, below.

We equilibrate the second and the third term, depending of R and $d_{\xi^\dagger}(R)$, respectively, by means of the auxiliary continuous and strictly decreasing function

$$(A.10) \quad \Phi(R) := \frac{(d_{\xi^\dagger}(R))^{q^*}}{R}, \quad R > 0,$$

which fulfills the limit conditions $\lim_{R \rightarrow 0} \Phi(R) = \infty$ and $\lim_{R \rightarrow \infty} \Phi(R) = 0$, thus it has a continuous decreasing inverse $\Phi^{-1} : (0, \infty) \rightarrow (0, \infty)$. By setting $R := \Phi^{-1}(\|A(x - x^\dagger)\|_Y)$ and introducing the index function $\zeta(t) := [d_{\xi^\dagger}(\Phi^{-1}(t))]^{q^*}$, $t > 0$, we get from (A.9), with some constant $\hat{K} > 0$, a variational inequality of the form

$$\beta D_{\xi^\dagger}^\Omega(x, x^\dagger) \leq \Omega(x) - \Omega(x^\dagger) + \hat{K} \zeta(\|A(x - x^\dagger)\|_Y), \quad x \in \mathcal{M}.$$

Remark 9. We observe, with $t = \Phi(R)$, $R > 0$, that

$$\frac{t}{\zeta(t)} = \frac{\Phi(R)}{[d_{\xi^\dagger}(R)]^{q^*}} = \frac{1}{R} \rightarrow 0 \quad \text{as } t \rightarrow 0.$$

Thus the function $\frac{t}{\zeta(t)}$ decreases to zero as $t \rightarrow 0$. In this case there is a concave majorant index function $\tilde{\varphi}$ to ζ (cf. [13, Chapt. 5]) such that

$$\beta D_{\xi^\dagger}^\Omega(x, x^\dagger) \leq \Omega(x) - \Omega(x^\dagger) + \varphi(\|A(x - x^\dagger)\|_Y), \quad x \in \mathcal{M},$$

with the constant $\beta = 1 - 1/q > 0$, and an index function φ which is a multiple of $\tilde{\varphi}$.

REFERENCES

- [1] S. W. Anzengruber and R. Ramlau. Morozov's discrepancy principle for Tikhonov-type functionals with nonlinear operators. *Inverse Problems*, 26:025001, 17 pp., 2010.
- [2] F. Bauer and T. Hohage. A Lepskij-type stopping rule for regularized Newton methods. *Inverse Problems*, 21(6):1975–1991, 2005.

- [3] G. Blanchard and P. Mathé. Conjugate gradient regularization under general smoothness and noise assumptions. *J. Inverse Ill-Posed Probl.*, 18(6):701–726, 2010.
- [4] R. I. Boţ and B. Hofmann. An extension of the variational inequality approach for obtaining convergence rates in regularization of nonlinear ill-posed problems. *Journal of Integral Equations and Applications*, 22(3):369–392, 2010.
- [5] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Problems*, 20(5):1411–1421, 2004.
- [6] J. Flemming. *Generalized Tikhonov Regularization: Basic Theory and Comprehensive Results on Convergence Rates*. Technische Universität Chemnitz, Dept. Mathematics, Chemnitz, Germany, 2011. PhD thesis (Dissertation), <http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-78152>.
- [7] J. Flemming. *Generalized Tikhonov Regularization and Modern Convergence Rate Theory in Banach Spaces*. Shaker Verlag, Aachen, 2012.
- [8] M. Grasmair. Generalized Bregman distances and convergence rates for non-convex regularization methods. *Inverse Problems*, 26(11):115014, 16 pp., 2010.
- [9] T. Hein and B. Hofmann. Approximate source conditions for nonlinear ill-posed problems—chances and limitations. *Inverse Problems*, 25(3):035003, 16 pp., 2009.
- [10] 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.
- [11] B. Hofmann and P. Mathé. Analysis of profile functions for general linear regularization methods. *SIAM J. Numer. Anal.*, 45(3):1122–1141, 2007.
- [12] B. Hofmann and M. Yamamoto. On the interplay of source conditions and variational inequalities for nonlinear ill-posed problems. *Appl. Anal.*, 89(11):1705–1727, 2010.
- [13] N. Korneïchuk. *Exact constants in approximation theory*, volume 38 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1991.
- [14] D. A. Lorenz. Convergence rates and source conditions for Tikhonov regularization with sparsity constraints. *J. Inverse Ill-Posed Probl.*, 16(5):463–478, 2008.
- [15] P. Mathé. The Lepskiï principle revisited. *Inverse Problems*, 22(3):L11–L15, 2006.
- [16] P. Mathé and S. V. Pereverzev. Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19(3):789–803, 2003.
- [17] E. Resmerita. Regularization of ill-posed problems in Banach spaces: convergence rates. *Inverse Problems*, 21(4):1303–1314, 2005.
- [18] E. Resmerita and O. Scherzer. Error estimates for non-quadratic regularization and the relation to enhancement. *Inverse Problems*, 22(3):801–814, 2006.
- [19] 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.
- [20] T. Schuster, B. Kaltenbacher, B. Hofmann, and K.S. Kazimierski. *Regularization Methods in Banach Spaces*. Walter de Gruyter, Berlin, 2012.

DEPARTMENT OF MATHEMATICS, CHEMNITZ UNIVERSITY OF TECHNOLOGY, 09107 CHEMNITZ, GERMANY

E-mail address: `hofmannb@mathematik.tu-chemnitz.de`

WEIERSTRASS INSTITUTE FOR APPLIED ANALYSIS AND STOCHASTICS, MOHRENSTRASSE 39, 10117 BERLIN, GERMANY

E-mail address: `peter.mathe@wias-berlin.de`