Parameter choice in Banach space regularization under variational inequalities

Bernd Hofmann

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

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

Peter Mathé

Weierstraß Institute for Applied Analysis and Stochastics, Mohrenstraße 39, 10117 Berlin, Germany

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

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 Lepskiĭ 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.

AMS classification scheme numbers: 65J20, 47J06, 47A52, 49J40

Submitted to: *Inverse Problems*

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

$$F(x) = y \tag{1.1}$$

with an (in general nonlinear) forward operator $F : \mathcal{D}(F) \subseteq X \to Y$, 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. [28] and [29, 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

$$\|y^\delta - y\|_Y \le \delta\,. \tag{1.2}$$

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

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

with a *convex* penalty functional $\mathcal{R} : X \to [0,\infty]$ and some positive exponent $1 < p < \infty$. We suppose in the sequel that the standard assumptions on $F, \mathcal{D}(F)$, and $\mathcal{R}$, made for the Tikhonov regularization in [15] and in the recent monographs [28, 29] are fulfilled. In particular, we assume that $\mathcal{R}$ is *stabilizing*, which means that for all $c \ge 0$ the sublevel sets

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

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 $\mathcal{R}$-minimizing solution, i.e., from a solution $x^\dagger$ to (1.1) with noise-free data y, for which we have

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

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

$$E(x, x^\dagger) = D_{\xi^\dagger}^{\mathcal{R}}(x, x^\dagger) := \mathcal{R}(x) - \mathcal{R}(x^\dagger) - \langle \xi^\dagger, x - x^\dagger\rangle_{X^*\times X} \tag{1.4}$$

is often used, where we denote by $\xi^\dagger \in \partial\mathcal{R}(x^\dagger) \subseteq X^*$ the subdifferential of the convex functional $\mathcal{R}$ at the point $x^\dagger$. The Bregman distance was introduced into the regularization theory by the study [7] in 2004, and henceforth this concept was adopted, refined and developed by many authors (cf., e.g., [14, 15, 19, 26, 27]). When considering E from (1.4) we always assume that $x \in \mathcal{D}(\mathcal{R}) := \{\tilde{x} \in X : \mathcal{R}(\tilde{x}) < \infty\}$ and that $x^\dagger$ belongs to the Bregman domain

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

The goal of the present paper is to study convergence rates of $E(x_\alpha^\delta, x^\dagger)$ as $\delta \to 0$ 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 $\mathcal{R}$,

(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 [15] as *variational inequalities*. Here we refer to the following variant (cf. [8, 9, 10, 11]), which uses the concept of index functions. We call a function $\varphi : (0,\infty) \to (0,\infty)$ an index function if it is continuous, strictly increasing, and satisfies the limit condition $\lim\limits_{t\to+0} \varphi(t) = 0$, see e.g. [16, 21].

Assumption VI (variational inequality) *We assume to have a constant* $0 < \beta \le 1$, *a concave index function* φ, *and a domain of validity* $\mathcal{M}$ *such that*

$$\beta\, E(x, x^\dagger) \le \mathcal{R}(x) - \mathcal{R}(x^\dagger) + \varphi(\|F(x) - F(x^\dagger)\|_Y) \quad \textit{for all}\ \ x \in \mathcal{M}. \tag{1.5}$$

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

Moreover, there are good reasons to restrict in (1.5) to concave *index functions. Namely, for index functions* φ *with* $\lim\limits_{t\to+0} \frac{\varphi(t)}{t} = 0$, *including the family of strictly convex index functions, the variational inequality degenerates in the sense that* $\mathcal{R}(x^\dagger) \le \mathcal{R}(x)$ *for all* $x \in \mathcal{M}$ *(see [9, Proposition 12.10], and for a special case [17, Proposition 4.3]). If* $0 < \lim\limits_{t\to+0} \frac{\varphi(t)}{t} < \infty$ *then the situation is equivalent to the case* $\varphi(t) = c\,t,\ c > 0$, *in (1.5) (see [9, Proposition 12.11]), and for an index function* φ *with* $\lim\limits_{t\to+0} \frac{\varphi(t)}{t} \nearrow +\infty$ *we can find a concave majorant index function that can be used in (1.5).*

The outline of the paper 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. A concluding discussion is given in Section 5. In an appendix we shall indicate on the basis of some examples from the literature how (1.5) may be derived for linear and nonlinear problems. Mostly, the examples employ varieties of solution smoothness and nonlinearity structure for obtaining Assumption VI. Our *focus* in this paper *is on choices of the regularization parameter* and their properties and consequences, but it was *not our intention* to present new results and examples concerning the verification of Assumption VI. On the other hand, we believe that it will be relevant in the future for the credibility of the approach to provide variational inequalities for important classes of nonlinear inverse problems, as far as possible also without using explicit source conditions or approximate source conditions and explicit nonlinearity conditions.

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., [15, 28, 29]). Under natural assumptions, stated there, $\mathcal{R}$-minimizing solutions

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

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

$$\alpha_* \to 0 \quad \text{and} \quad \frac{\delta^p}{\alpha_*} \to 0 \quad \text{as} \quad \delta \to 0 \tag{2.1}$$

we have convergence for both

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

Hence, all regularized solutions $x_{\alpha_*}^\delta$ for sufficiently small $\delta > 0$ belong to $\mathcal{M}^{\mathcal{R}}(\mathcal{R}(x^\dagger)+c)$, for some $c > 0$, and moreover if $\delta_n \to 0$ 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 $\mathcal{R}$-minimizing solution $x^\dagger$ is not unique. For more details see, for example, [29, 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 stability region is of interest.

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

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

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

Notice that $\mathcal{F}_{K,C}(\delta) \subset \mathcal{M}^{\mathcal{R}}(\mathcal{R}(x^\dagger) + K\varphi(C\delta))$, such that minimizers which are directed 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*

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

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* $\mathcal{R}(x) - \mathcal{R}(x^\dagger)$, but not its modulus $|\mathcal{R}(x) - \mathcal{R}(x^\dagger)|$. In the proofs given below, we shall obtain the following strengthening. The parameter choices will direct 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*

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

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

Plainly, the inclusion $\mathcal{F}^{\text{conv}}_{K,C}(\delta) \subset \mathcal{F}_{K,C}(\delta)$ holds. But the *additional* requirement provides us with a rate of convergence for $\mathcal{R}(x^\delta_\alpha) \to \mathcal{R}(x^\dagger)$, 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 $\mathcal{F}^{\text{conv}}_{K,C}(\delta)$ for some $K > 0,\ C \geq 1$ and $\delta > 0$ then, in addition to the assertion from Proposition 1, we have*

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

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* in Section 4 yielding convergence rates for the error measure E and deviations of Ω of the type $\mathcal{O}(\varphi(\delta))$ as $\delta \to 0$ whenever φ in (1.5) is a concave index function. The methodology based on Assumption VI is not helpful for providing *enhanced convergence rates* as proven for the Bregman distance (1.4) as error measure in [23, 25] up to the order $\mathcal{O}(\delta^{4/3})$ by using source conditions of higher order in combination with duality mappings. Note that a first step to extend the variational inequality approach to higher rates was made in [12].

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*

$$\mathcal{R}(x^\dagger) - \mathcal{R}(x_\alpha^\delta) \le \varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y), \text{ and}$$
$$\mathcal{R}(x_\alpha^\delta) - \mathcal{R}(x^\dagger) \le \frac{\delta^p}{p\alpha}.$$

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

$$\frac{\|F(x_\alpha^\delta) - y^\delta\|_Y^p}{p} + \alpha\mathcal{R}(x_\alpha^\delta) \le \frac{\|F(x^\dagger) - y^\delta\|_Y^p}{p} + \alpha\mathcal{R}(x^\dagger) \le \frac{\delta^p}{p} + \alpha\mathcal{R}(x^\dagger), \quad (3.1)$$

from which the assertion follows. □

Another conclusion is less obvious.

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

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

Proof: By using the first inequality of Lemma 1 and formula (3.1) we can estimate as

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

As a consequence of the inequality

$$(a+b)^p \le 2^{p-1}(a^p + b^p), \qquad a, b \ge 0, \ \ p \ge 1,$$

we also have the following lower bound

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

Inserting this we see that

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

which completes the proof. □

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

$$\Phi_p(t) := \frac{t^p}{\varphi(t)}, \quad t > 0, \quad (3.2)$$

where φ is an arbitrary concave index function. Since $p > 1$, also Φ_p is an index function.

Corollary 1 *Let α_* be given from*

$$\alpha_* := \Phi_p(\delta). \tag{3.3}$$

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

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

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

$$\begin{aligned}
\frac{\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y^p}{p} &\le 2^p\frac{\delta^p}{p} + \alpha 2^{p-1}\varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y)\\
&\le 2^p\frac{\delta^p}{p} + \alpha_* 2^{p-1}\varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y)\\
&= 2^p\frac{\delta^p}{p} + \delta^p 2^{p-1}\frac{\varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y)}{\varphi(\delta)}\\
&= \frac{\delta^p}{p}\left(2^p + p2^{p-1}\frac{\varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y)}{\varphi(\delta)}\right)\\
&\le 2^{p-1}(2+p)\frac{\delta^p}{p}\frac{\varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y)}{\varphi(\delta)}\\
&\le 2^{p-1}(2+p)\frac{\delta^p}{p}\frac{\delta^{-1}\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y\varphi(\delta)}{\varphi(\delta)}\\
&= 2^{p-1}(2+p)\frac{\delta^{p-1}}{p}\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y.
\end{aligned}$$

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

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*

$$\alpha \ge \frac{1}{p2^{p-1}}\frac{\tau^p - 1}{\tau^p + 1}\Phi_p((\tau-1)\delta). \tag{3.4}$$

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

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

Thus

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

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

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

Since $\tau\delta \le t_\alpha + \delta$, we arrive, using the function Φ_p from (3.2), at

$$\Phi_p((\tau-1)\delta) \le \Phi_p(t_\alpha) \le p2^{p-1}\frac{\tau^p+1}{\tau^p-1}\alpha,$$

and the proof is complete. □

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. [5, 8]). Here we present an intuitive parameter choice, which was obtained in [11] 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 $\mathcal{F}_{K,C}^{\mathrm{conv}}(\delta)$. We recall the index function Φ_p from (3.2).

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 $x_{\alpha_*}^\delta \in \mathcal{F}_{K,C}^{\mathrm{conv}}(\delta)$ with $K = 1$ and $C = 2(2+p)^{1/(p-1)}$.

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

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

as $\delta \to 0$.

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 $\mathcal{R}(x^\dagger) - \mathcal{R}(x_{\alpha_*}^\delta)$, appropriately. Furthermore, from the second assertion of Lemma 1 we have that $\mathcal{R}(x_{\alpha_*}^\delta) - \mathcal{R}(x^\dagger) \le \delta^p/(p\alpha_*) \le \varphi(\delta)$, by the choice of α_*. The convergence rates in Item (ii) are a consequence of Propositions 1 & 2. □

Remark 2 *An inspection of the proofs in Section 3 shows that the first two convergence rates in Theorem 1 use only the implication $x_{\alpha_*}^\delta \in \mathcal{M} \Rightarrow x_{\alpha_*}^\delta \in \mathcal{F}_{K,C}(\delta)$, for sufficiently small $\delta > 0$. Only for the third $\mathcal{R}$-rate the membership $x_{\alpha_*}^\delta \in \mathcal{F}_{K,C}^{\mathrm{conv}}(\delta)$ is required. The a priori parameter choice from (3.3) satisfies the condition (2.1).*

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

$$\Delta_q := \left\{\alpha_j : \quad \alpha_j := q^j \alpha_0, \quad j = 1, 2, \dots\right\}. \tag{4.1}$$

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

$$\tau_1\delta \le \|F(x^\delta_{\alpha_*}) - y^\delta\|_Y \le \tau_2\delta \tag{4.2}$$

(cf. [1, 2, 9]). We shall call this the *strong discrepancy principle*. It is not clear that this is always possible, and it was mentioned in [29, Chapt. 4] that for nonlinear operators F there may be a duality gap due to the non-convexity of the functional $\mathcal{T}^\delta_\alpha$ 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. For another alternative version of the discrepancy principle in the context of Tikhonov regularization we also refer to [24].

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*

$$\|F(x^\delta_\alpha) - y^\delta\|_Y \le \tau\delta.$$

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

(i) If $x^\delta_\alpha \in \mathcal{M}$, $\alpha \ge \alpha_*$, then $x^\delta_{\alpha_*} \in \mathcal{F}^{\mathrm{conv}}_{K,C}(\delta)$ with $K = \frac{1}{2q}\max\left\{\left(\frac{2}{\tau-1}\right)^p \frac{\tau^p+1}{\tau^p-1}, 2q\right\}$, and $C = \tau + 1$.

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

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

as $\delta \to 0$.

Proof: We first bound

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

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

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

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

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

The convergence rates in Item (ii) are again consequences of Propositions 1 & 2. □

Remark 3 *We emphasize that here we bounded*

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

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

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 [3]. 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 [20]. 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 $\alpha_m \le \alpha \le \alpha_{AP}$ and for $x_\alpha^\delta \in \mathcal{M}$ we have that*

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

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

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

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

We want to use the Lepskiĭ principle, as this is outlined in [20], by using a multiple of the decreasing function $\alpha \to \delta^p/\alpha$, and we let

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

From [20, 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*

$$E(x^\delta_{\alpha'}, x^\delta_\alpha) \le 2\Psi(\alpha'), \quad \textit{for all } \alpha' \in \Delta_q,\ \alpha_m \le \alpha' < \alpha.$$

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

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

Proof: As in [20, Prop. 1] we introduce the parameter

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

(Caution: the notation in [20] 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 $\alpha_+ \ge \alpha_{AP}$. Otherwise, we consider the index $k \le m$ for which $\alpha_k < \alpha_{AP} \le \alpha_k/q$, which results in $\alpha_+ \ge \alpha_k$. Proposition 1 in [20] states that $E(x^\delta_{\alpha_*}, x^\dagger) \le 3\Psi(\alpha_+)$. Thus in either case this yields

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

which completes the proof. □

Remark 4 *The above application of the Lepskiĭ principle does not include the Bregman distance as error measure $E(x, x^\dagger) := D^{\mathcal{R}}_{\xi^\dagger}(x, x^\dagger)$, because this is not a metric, in general. However, if the Bregman distance is q-coercive, $D^{\mathcal{R}}_{\xi^\dagger}(x, x^\dagger) \ge c\frac{\|x-x^\dagger\|^q_X}{q}$, then the validity of a variational inequality for $D^{\mathcal{R}}_{\xi^\dagger}(x, x^\dagger)$ implies the one for $\|x - x^\dagger\|^q_X/q$, and we can apply the Lepskiĭ parameter choice to the differences $x^\delta_\alpha - x^\delta_{\alpha'}$, i.e., test whether $\|x^\delta_\alpha - x^\delta_{\alpha'}\|_X \le (2\frac{1+pC_p}{p\beta})^{1/q}\delta^{p/q}/(\alpha')^{1/q}$ for $\alpha' \le \alpha$.*

5. Concluding discussion

We summarize the above findings on the parameter choice strategies. First, we emphasize that the a priori parameter choice from § 4.1 is related 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 $E(x, x^\dagger) = \|x - x^\dagger\|^2_X$ measures the squared error, we obtain a rate $\|x^\delta_{\alpha_*} - x^\dagger\|_X = \mathcal{O}(\sqrt{\varphi(\delta)})$. 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 $x^\dagger = \psi(A^*A)v,\ \|v\|_X \le 1$. Taking this for granted we see that

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

and hence $\Theta_\psi(\alpha_*) = \delta$, which is the 'ordinary' a priori parameter choice in linear problems in Hilbert space under given source condition $x^\dagger = \psi(A^*A)v,\ \|v\|_X \le 1$ (cf. [21]).

We turn to the a posteriori parameter choices, and we notice that the discrepancy principle requires to start with some 'large' parameter α_1, whereas the Lepskiĭ principle one has to start with some smallest $\alpha_m \in \Delta_q$. Thus the latter parameter choice is more involved as the discrepancy principle. However, for linear problems in Hilbert space, the discrepancy principle is known for its early saturation, a drawback which is not present in the Lepskiĭ parameter choice. A similar effect is not known for nonlinear problems in Banach space, and this discussion still has to be done.

The proofs of both principles require that the search for the parameter α does not stop immediately, i.e., at α_1 for the discrepancy principle, or at α_m for the Lepskiĭ principle, and these cases require additional attention. In fact, for the discrepancy principle it may happen that for $\alpha := \alpha_1$ the assumption $\|F(x_{\alpha_1}^\delta) - y^\delta\|_Y \le \tau\delta$ is already fulfilled. The question is, whether there is still the rate $\mathcal{O}\left(\varphi(\delta)\right)$ to be observed. This can indeed be proved for linear problems in Hilbert space (cf. [4] 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) $x_\alpha^\delta \to 0$ as $\alpha \to \infty$, and hence immediate stop refers to $\|y^\delta\|_Y \le \tau\delta$. 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 $\mathcal{R}(x_{\alpha_1}^\delta) - \mathcal{R}(x^\dagger)$ 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 $\delta^p/(p\alpha_1) \le \bar{K}\varphi(\delta)$, or equivalently, by using the index function Φ_p from (3.2), that

$$\Phi_p(\delta) \le \bar{K}\, p\, \alpha_1. \tag{5.1}$$

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

$$\mathcal{R}(x_{\alpha_1}^\delta) - \mathcal{R}(x^\dagger) \le \frac{\delta^p}{p\alpha_1} = \frac{\delta^p}{\Phi_p(\delta)} = \varphi(\delta).$$

This shows that if the initial value α_0 is chosen large enough then immediate stop yields an error bound of the form $\mathcal{O}\left(\varphi(\delta)\right)$. However, for any particular instance $x^\dagger$ at hand we cannot verify whether (5.1) holds, since φ is not known to us.

The situation is similar for the Lepskiĭ principle. In the formulation of Theorem 3 we assumed that $E(x_{\alpha_m}^\delta, x^\dagger) \le \Psi(\alpha_m)$. 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.

Both a posteriori parameter choices do not require to know the function φ involved in Assumption VI. However, the functional Ψ, as it is used in the Lepskiĭ principle, requires the parameter $0 < \beta \le 1$. In some cases in which Assumption VI holds the factor β is known to be one, see Appendix A. In general, the functional Ψ could be increased by some multiplicative safeguard factor, say $\nu > 1$. In this case the conclusion of Theorem 3 remains true whenever a variational inequality with factor $\beta > 1/\nu$ holds on some domain of validity.

As Theorem 2 shows, the discrepancy principle directs the chosen parameter towards the region of convergence. The parameter choice à la Lepskiĭ from Theorem 3 provides us with an error bound, which is obtained regardless whether the approximating $x^\delta_{\alpha_*}$ 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_*: In [20, Prop. 2.1] the information is given that $\alpha_* \ge \alpha_+$, such that $\alpha_* \ge \alpha_+ \ge q\alpha_{AP} = q\Phi_p(\delta)$. This bounds the excess penalty $\mathcal{R}(x^\delta_{\alpha_*}) - \mathcal{R}(x^\dagger) \le \varphi(\delta)/q$. However, it is not clear whether the discrepancy $\|F(x^\delta_{\alpha_*}) - y^\delta\|_Y$ is of the order δ.

Finally, we mention that the roughness parameter $q < 1$, which describes the construction for Δ_q in (4.1) influences the accuracy by a multiplicative factor $1/q$. This can be seen from the constant K in Theorem 2, and also from Theorem 3.

The obtained rates are valid in all cases where a suitable variational inequality as in Assumption VI holds. In the Appendix A we highlight several important cases, where such variational inequalities can be obtained. Thus, in all such cases, the parameter choices as discussed in this study will yield the rates described through the index function φ in Assumption VI. In particular, we sketch the case of sparse recovery by using a weighted l^q-norm as penalty $\mathcal{R}$ in the Tikhonvov functional. For $q = 1$, and if $\varphi(t) = t$ in (1.5), the maximal rate $\|x^\delta_{\alpha_*} - x^\dagger\|_X = \mathcal{O}(\delta)$ can be obtained for Tikhonov regularization under sparsity constraints.

We mention that converse results concluding from the validity of Assumption VI to solution smoothness are not known at present in the general Banach space setting if the index function φ in (1.5) strictly concave, i.e. $t = o(\varphi(t))$ as $t \to +0$. However, for linear problems in Hilbert spaces such an assertion concerning Hölder rates and $\varphi(t) = t^\kappa,\ 0 < \kappa < 1$ was formulated as Proposition 6.8 in [17]. The proof of this proposition is simply based on the converse results from [22].

Appendix A. Examples for verifying Assumption VI

For the convenience of the reader we briefly sketch some approaches to show how variational inequalities occur. We suppose that the mapping $F : \mathcal{D}(F) \subseteq X \to Y$

with some convex domain $\mathcal{D}(F)$ has a one-sided directional derivative at $x^\dagger$ given as a bounded linear operator $F'(x^\dagger): X \to Y$ such that

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

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 [29, Sections 3.2 and 4.2]).

Appendix A.1. Bregman distance as error measure: benchmark case

Here we assume that $x^\dagger \in \mathcal{D}_B(\mathcal{R})$ and some corresponding subgradient element $\xi^\dagger \in \partial\mathcal{R}(x^\dagger)$ fulfills the benchmark source condition

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

Such information allows us to bound for all $x \in X$

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

After adding the term $\mathcal{R}(x) - \mathcal{R}(x^\dagger)$ on both sides this yields that

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

Remark 5 *We highlight the special case when X is a Hilbert space and $\mathcal{R}(x) = \|x\|_X^2$ with $\mathcal{D}(\mathcal{R}) = X$. Then $D_{\xi^\dagger}^{\mathcal{R}}(x, x^\dagger) = \|x - x^\dagger\|_X^2$ (cf. [28, Example 3.18]), and (A.3) implies*

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

It was emphasized in [8, Chapter 13] that for bounded linear operators $F = A$ mapping between Hilbert spaces X and Y solution smoothness can always be expressed by variational inequalities (1.5) with general index functions φ and a domain of validity $\mathcal{M} = X$.

For the general Banach space setting inequality (A.3) also results in a variational inequality for bounded linear operators $F = A : X \to Y$. Then Assumption VI is satisfied with $A = F'(x^\dagger)$, $\beta = 1$, $E(x, x^\dagger) = D_{\xi^\dagger}^{\mathcal{R}}(x, x^\dagger)$ and $\varphi(t) = \|v\|_{Y^*}t,\ t > 0$ on the whole space $\mathcal{M} = X$ as domain of validity.

If the mapping F is nonlinear then we may use certain *structure of nonlinearity* 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

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

for some concave index function σ and constant $\eta > 0$ on some set $\mathcal{M} \subset \mathcal{D}(F)$ (cf. [5]). In this case (A.3) provides us with a variational inequality on $\mathcal{M}$ with $\beta = 1,\ E(x,x^\dagger) = D^{\mathcal{R}}_{\xi^\dagger}(x,x^\dagger)$ and $\varphi(t) = \eta\|v\|_{Y^*}\sigma(t),\ t > 0$.

An alternative structural condition is given in the form

$$\|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\|_Y \le \eta D^{\mathcal{R}}_{\xi^\dagger}(x,x^\dagger), \quad x \in \mathcal{M}, \tag{A.5}$$

again for some set $\mathcal{M} \subset \mathcal{D}(F)$, (cf., e.g., [15, 27]). This allows us to bound

$$\|F'(x^\dagger)(x - x^\dagger)\|_Y \le \eta D^{\mathcal{R}}_{\xi^\dagger}(x,x^\dagger) + \|F(x) - F(x^\dagger)\|_Y,\ x \in \mathcal{M}.$$

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

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

with $0 < \beta = 1 - \eta\|v\|_{Y^*} \le 1,\ E(x,x^\dagger) = D^{\mathcal{R}}_{\xi^\dagger}(x,x^\dagger)$ 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).

We conclude this subsection by mentioning that there is some *converse result* in the sense that a variational inequality (1.5) just for $\varphi(t) = c\,t,\ c > 0$, implies the validity of the benchmark source condition (A.2) (cf. [28, Proposition 3.38]).

Appendix A.2. Bregman distance as error measure: violation of the benchmark

If the assumption (A.2) is violated then we may use the *method of approximate source conditions* (cf. [5, 14]) to derive variational inequalities. To this end we need additionally that the distance function

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

is nonincreasing and obeys the limit condition $d_{\xi^\dagger}(R) \to 0$ as $R \to \infty$. As mentioned in [5] this is the case when $F'(x^\dagger)^{**}\colon X^{**} \to Y^{**}$ is injective. Additionally this approach presumes that the Bregmann distance is q-coercive, i.e., that

$$D^{\mathcal{R}}_{\xi^\dagger}(x,x^\dagger) \ge c_q\,\|x - x^\dagger\|^q_X \qquad \text{for all} \quad x \in \mathcal{M}, \tag{A.7}$$

is satisfied for some exponent $2 \le q < \infty$ and a corresponding constant $c_q > 0$. Such assumption is for example fulfilled if $\mathcal{R}(x) := \|x\|^q_X$ 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

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

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

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

Adding, as before, the difference $\mathcal{R}(x) - \mathcal{R}(x^\dagger)$ on both sides gives

$$D^{\mathcal{R}}_{\xi^\dagger}(x,x^\dagger) \le \mathcal{R}(x) - \mathcal{R}(x^\dagger) + R\,\|F'(x^\dagger)(x - x^\dagger)\|_Y + d_{\xi^\dagger}(R)\,\|x - x^\dagger\|_X,\ x \in \mathcal{M}. \tag{A.8}$$

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

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

An application of Young's inequality yields

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

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

$$\beta D_{\xi^\dagger}^{\mathcal{R}}(x,x^\dagger) \le \mathcal{R}(x) - \mathcal{R}(x^\dagger) + R\|F'(x^\dagger)(x-x^\dagger)\|_Y + \frac{c_q^{-q^*/q}}{q^*}\,\left(d_{\xi^\dagger}(R)\right)^{q^*}. \tag{A.9}$$

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

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

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

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

Remark 6 *We observe, with $t = \Phi(R)$, $R>0$, that*

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

Thus the function $\frac{t}{\zeta(t)}$ decreases to zero as $t \to 0$. In this case there is a concave majorant index function $\tilde{\varphi}$ to ζ (cf. [18, Chapt. 5]) such that

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

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

Appendix A.3. On a couple of variational inequalities

For nonlinear operators F and general convex penalties $\mathcal{R}$ in [2, Condition 3.3] a coupled system of two variational inequalities

$$\langle \xi^\dagger, x^\dagger - x\rangle_{X^*\times X} \le \beta_1 D_{\xi^\dagger}^{\mathcal{R}}(x,x^\dagger) + \beta_2\|F'(x^\dagger)(x-x^\dagger)\|_Y + \beta_3\|F(x)-F(x^\dagger)\|_Y^{\kappa} \tag{A.11}$$

and

$$\|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\|_Y$$
$$\leq \gamma_1 D^{\mathcal{R}}_{\xi^\dagger}(x, x^\dagger) + \gamma_2 \|F'(x^\dagger)(x - x^\dagger)\|_Y + \gamma_3 \|F(x) - F(x^\dagger)\|^\kappa_Y \tag{A.12}$$

with constants $0 < \kappa \leq 1,\ \beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2, \gamma_3 \geq 0$ satisfying the relations

$$\beta_1 < 1, \quad \gamma_2 < 1 \quad \text{and} \quad \frac{\beta_2 \gamma_1}{(1 - \beta_1)(1 - \gamma_2)} < 1 \tag{A.13}$$

and valid on some vicinity $\mathcal{M}$ of $x^\dagger$, containing $x \in \mathcal{D}$ satisfying the condition

$$\|F(x) - F(x^\dagger)\|_Y \leq K \tag{A.14}$$

with some constant $K > 0$, has been considered. Based on that couple of variational inequalities convergence rates for the Bregman distance of type $\mathcal{O}(\delta^\kappa),\ 0 < \kappa \leq 1$, could be shown by applying the strong discrepancy principle (4.2), where regularized solutions always belong to the sublevel set $\mathcal{M}^{\mathcal{R}}(\mathcal{R}(x^\dagger))$. Hence for the discussion in [2] the system (A.11) – (A.14) needs to hold only for $\mathcal{M} \subseteq \mathcal{M}^{\mathcal{R}}(\mathcal{R}(x^\dagger))$. For such *restricted domain of validity* $\mathcal{M}$ (see [2, Lemma 3.6]) the couple of variational inequalities is equivalent to our variational inequality (1.5), which is in this case of the form

$$\beta\, D^{\mathcal{R}}_{\xi^\dagger}(x, x^\dagger) \leq \mathcal{R}(x) - \mathcal{R}(x^\dagger) + C\, \|F(x) - F(x^\dagger)\|^\kappa_Y \tag{A.15}$$

with some constants $0 < \beta \leq 1$ and $C > 0$, in combination with a nonlinearity condition

$$\|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\|_Y \leq \gamma\, \|F(x) - F(x^\dagger)\|^\kappa_Y \tag{A.16}$$

with some constant $\gamma > 0$.

In the following we will briefly verify that the fulfillment of the coupled system (A.11) – (A.14) acts as a *sufficient condition* for the variational inequality (A.15) even if $\mathcal{M}$ is *not* a subset of $\mathcal{M}^{\mathcal{R}}(\mathcal{R}(x^\dagger))$. By adding $\mathcal{R}(x) - \mathcal{R}(x^\dagger)$, subtracting $\beta_1 D^{\mathcal{R}}_{\xi^\dagger}(x, x^\dagger)$ and dividing by $1 - \beta_1$ we obtain from (A.11)

$$D^{\mathcal{R}}_{\xi^\dagger}(x, x^\dagger) \leq \frac{\mathcal{R}(x) - \mathcal{R}(x^\dagger)}{1 - \beta_1} + \frac{\beta_2}{1 - \beta_1} \|F'(x^\dagger)(x - x^\dagger)\|_Y + \frac{\beta_3}{1 - \beta_1} \|F(x) - F(x^\dagger)\|^\kappa_Y. \tag{A.17}$$

Moreover, we have from the triangle inequality

$$\|F'(x^\dagger)(x - x^\dagger)\|_Y \leq \|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\|_Y + \|F(x) - F(x^\dagger)\|_Y$$

together with (A.12) and (A.14)

$$\|F'(x^\dagger)(x - x^\dagger)\|_Y \leq \gamma_1 D^{\mathcal{R}}_{\xi^\dagger}(x, x^\dagger) + \gamma_2 \|F'(x^\dagger)(x - x^\dagger)\|_Y + \gamma_4 \|F(x) - F(x^\dagger)\|^\kappa_Y$$

for some constant $\gamma_4 > 0$ and hence

$$\|F'(x^\dagger)(x - x^\dagger)\|_Y \leq \frac{\gamma_1}{1 - \gamma_2} D^{\mathcal{R}}_{\xi^\dagger}(x, x^\dagger) + \frac{\gamma_4}{1 - \gamma_2} \|F(x) - F(x^\dagger)\|^\kappa_Y.$$

Substituting this into the estimate (A.17) yields with some constant $\hat{C} > 0$

$$D^{\mathcal{R}}_{\xi^\dagger}(x, x^\dagger) \leq \frac{\mathcal{R}(x) - \mathcal{R}(x^\dagger)}{1 - \beta_1} + \frac{\beta_2 \gamma_1}{(1 - \beta_1)(1 - \gamma_2)} D^{\mathcal{R}}_{\xi^\dagger}(x, x^\dagger) + \hat{C} \|F(x) - F(x^\dagger)\|^\kappa_Y.$$

By subtracting $\frac{\beta_2 \gamma_1}{(1-\beta_1)(1-\gamma_2)} D^{\mathcal{R}}_{\xi^\dagger}(x, x^\dagger)$ and multiplying by $1 - \beta_1$ we arrive at (A.15), where in view of (A.13) $0 < \beta = \frac{\beta_2 \gamma_1}{1-\gamma_2} \leq 1$ and $C = (1 - \beta_1)\hat{C} > 0$.

Appendix A.4. Error under sparsity constraints

The authors in [2] also study variational inequalities under *sparsity constraints*, where a frame $\{\phi_\lambda\}_{\lambda\in\Lambda}$ in the Hilbert space X, and penalty functionals

$$\mathcal{R}_{q,w}(x) := \sum_{\lambda\in\Lambda} w_\lambda \,|\langle\phi_\lambda, x\rangle|^q,\ 1 \le q \le 2,$$

with weights $\{w_\lambda\}_{\lambda\in\Lambda}$ are considered. We confine our analysis to a linear forward operators $A : X \to Y$ with values in the Hilbert space Y. As shown in Appendix A.3, the coupled system in [2, Condition 3.3] yields a variational inequality

$$\tilde\beta D^{\mathcal{R}_{q,w}}_{\xi^\dagger}(x, x^\dagger) \le \mathcal{R}_{q,w}(x) - \mathcal{R}_{q,w}(x^\dagger) + C_1\|A(x - x^\dagger)\|_Y^\kappa, \quad x \in X, \tag{A.18}$$

with $0 < \kappa \le 1,\ 0 < \tilde\beta \le 1,\ C_1 > 0$, and for the Bregmann distance $D^{\mathcal{R}_{q,w}}_{\xi^\dagger}(x, x^\dagger)$ as error measure. We shall sketch that the inequality (A.18) turns to a variational inequality for the error measure $E(x, x^\dagger) = \|x - x^\dagger\|_X^q$ under sparsity. Suppose that the solution $x^\dagger$ is sparse, i.e., it has coefficients within a finite subset J of the index set Λ, and that the restriction $A|_U$ of the operator A to the subspace $U = \operatorname{span}\{\varphi_\lambda,\ \lambda \in J\}$ is injective (see [2, Condition 4.1] for details). Then, based on techniques from [13], the authors in [2, Lemma 4.2] assert that there are constants $\mu_1, \mu_2 > 0$ such that we have

$$\mathcal{R}_{q,w}(x - x^\dagger) \le \mu_1 D^\xi_{\mathcal{R}_{q,w}}(x, x^\dagger) + \mu_2\|A(x - x^\dagger)\|_Y^q, \quad x \in X. \tag{A.19}$$

Combining the inequalities (A.18)–(A.19), we deduce with $0 < \beta := \min\left(1, \tilde\beta/\mu_1\right) \le 1$ that

$$\beta\mathcal{R}_{q,w}(x - x^\dagger) \le \mathcal{R}_{q,w}(x) - \mathcal{R}_{q,w}(x^\dagger) + C_2\|A(x - x^\dagger)\|_Y^\kappa + C_3\|A(x - x^\dagger)\|_Y^q$$

for some positive constants C_2, C_3. In a vicinity of $x^\dagger$, i.e., if we have $\|A(x - x^\dagger)\|_Y \le K$ for some $K > 0$, this yields

$$\beta\mathcal{R}_{q,w}(x - x^\dagger) \le \mathcal{R}_{q,w}(x) - \mathcal{R}_{q,w}(x^\dagger) + C\,\|A(x - x^\dagger)\|_Y^\kappa, \tag{A.20}$$

for some $C > 0$ and all x in some neighborhood $\mathcal{M}$ of $x^\dagger$. By virtue of [2, Lemma 2.3], which bounds as $\|x - x^\dagger\|_X^q \le \mu_3\mathcal{R}_{q,w}(x - x^\dagger)$, we finally find the variational inequality

$$\bar\beta\|x - x^\dagger\|_X^q \le \mathcal{R}_{q,w}(x) - \mathcal{R}_{q,w}(x^\dagger) + C\,\|A(x - x^\dagger)\|_Y^\kappa,$$

and x as above. For extensions from monomials of $\|A(x - x^\dagger)\|_Y$ to more general index functions we refer to [6].

Acknowledgments

The authors thank two anonymous referees for hints that allowed us to improve the presentation of the results. This work started while the first author was visiting the Weierstraß Institute for Applied Analysis and Stochastics, Berlin, in February 2012. B. Hofmann thanks P. Mathé, and the Weierstrass Institute for the kind invitation and hospitality. The research of B. Hofmann was also supported by Deutsche Forschungsgemeinschaft (DFG) under Grant HO 1454/8-1.

References

[1] S. W. Anzengruber and R. Ramlau. Morozov's discrepancy principle for Tikhonov-type functionals with nonlinear operators. *Inverse Problems*, 26:025001, 17pp., 2010.

[2] S. W. Anzengruber and R. Ramlau. Convergence rates for Morozov's discrepancy principle using variational inequalities. *Inverse Problems*, 27:105007, 18pp., 2011.

[3] F. Bauer and T. Hohage. A Lepskij-type stopping rule for regularized Newton methods. *Inverse Problems*, 21(6):1975–1991, 2005.

[4] G. Blanchard and P. Mathé. Conjugate gradient regularization under general smoothness and noise assumptions. *J. Inverse Ill-Posed Probl.*, 18(6):701–726, 2010.

[5] 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.

[6] K. Bredies and D. A. Lorenz. Regularization with non-convex separable constraints. *Inverse Problems*, 25(8):085011, 14pp., 2009.

[7] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Problems*, 20(5):1411–1421, 2004.

[8] J. Flemming. *Generalized Tikhonov Regularization: Basic Theory and Comprehensive Results on Convergence Rates*. Technische Universität Chemnitz, Germany, 2011. PhD thesis (Dissertation), `http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-78152`.

[9] J. Flemming. *Generalized Tikhonov Regularization and Modern Convergence Rate Theory in Banach Spaces*. Shaker Verlag, Aachen, 2012.

[10] K. Frick and M. Grasmair. Regularization of linear ill-posed problems by the augmented Lagragnian method and variational inequalities. Technical report, ArXiv:1204.0771, 2012. `http://arxiv.org/pdf/1204.0771v1.pdf`.

[11] M. Grasmair. Generalized Bregman distances and convergence rates for non-convex regularization methods. *Inverse Problems*, 26(11):115014, 16pp., 2010.

[12] M. Grasmair. Variational inequalities and improved convergence rates for Tikhonov regularization on Banach spaces. Technical report, University of Vienna, Preprint, 2011. `http://arxiv.org/pdf/1107.2471v1.pdf`.

[13] M. Grasmair, M. Haltmeier, and O. Scherzer. Sparse regularization with l^q penalty term. *Inverse Problems*, 24(5):055020, 13pp., 2008.

[14] T. Hein and B. Hofmann. Approximate source conditions for nonlinear ill-posed problems—chances and limitations. *Inverse Problems*, 25(3):035003, 16pp., 2009.

[15] 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.

[16] B. Hofmann and P. Mathé. Analysis of profile functions for general linear regularization methods. *SIAM J. Numer. Anal.*, 45(3):1122–1141, 2007.

[17] 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.

[18] N. Korneĭchuk. *Exact Constants in Approximation Theory*, volume 38 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1991.

[19] D. A. Lorenz. Convergence rates and source conditions for Tikhonov regularization with sparsity constraints. *J. Inverse Ill-Posed Probl.*, 16(5):463–478, 2008.

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

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

[22] A. Neubauer. On converse and saturation results for Tikhonov regularization of linear ill-posed problems. *SIAM J. Numer. Anal.*, 34(2):517–527, 1997.

[23] A. Neubauer. On enhanced convergence rates for Tikhonov regularization of nonlinear ill-posed

problems in Banach spaces. *Inverse Problems*, 25(6):065009, 10pp, 2009.

[24] A. Neubauer. Modified Tikhonov regularization for nonlinear ill-posed problems in Banach spaces. *J. Integral Equations Appl.*, 22(2):341–351, 2010.

[25] A. Neubauer, T. Hein, B. Hofmann, S. Kindermann, and U. Tautenhahn. Improved and extended results for enhanced convergence rates of Tikhonov regularization in Banach spaces. *Appl. Anal.*, 89(11):1729–1743, 2010.

[26] E. Resmerita. Regularization of ill-posed problems in Banach spaces: convergence rates. *Inverse Problems*, 21(4):1303–1314, 2005.

[27] E. Resmerita and O. Scherzer. Error estimates for non-quadratic regularization and the relation to enhancement. *Inverse Problems*, 22(3):801–814, 2006.

[28] 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.

[29] T. Schuster, B. Kaltenbacher, B. Hofmann, and K.S. Kazimierski. *Regularization Methods in Banach Spaces*. Walter de Gruyter, Berlin/Boston, 2012. ISBN: 978-3-11-025524-9.