REGULARIZATION PROPERTIES OF THE SEQUENTIAL DISCREPANCY PRINCIPLE FOR TIKHONOV REGULARIZATION IN BANACH SPACES

STEPHAN W. ANZENGRUBER, BERND HOFMANN*, AND PETER MATHÉ

Abstract. The stable solution of ill-posed non-linear operator equations in Banach space requires regularization. One important approach is based on Tikhonov regularization, in which case a one-parameter family of regularized solutions is obtained. It is crucial to choose the parameter appropriately. Here, a sequential variant of the discrepancy principle is analyzed. In many cases such parameter choice exhibits the feature, called regularization property below, that the chosen parameter tends to zero as the noise tends to zero, but slower than the noise level. Here we shall show such regularization property under two natural assumptions. First, exact penalization must be excluded, and secondly, the discrepancy principle must stop after a finite number of iterations. We conclude this study with a discussion of some consequences for convergence rates obtained by the discrepancy principle under the validity of some kind of variational inequality, a recent tool for the analysis of inverse problems.

1. Introduction

In this study, we are concerned with *asymptotic properties of regularization parameters* for Tikhonov-regularized solutions obtained by a variant of Morozov's discrepancy principle which we will call *sequential discrepancy principle (SDP)*. Precisely, we focus on some, in general non-linear, ill-posed operator equation

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

which acts as a mathematical model for an inverse problem. The forward operator $F : \mathrm{dom}(F) \subseteq X \to Y$, with domain $\mathrm{dom}(F)$, acts between the Banach spaces X and Y. We shall assume that problem (1.1) is solvable for the right-hand side $y^\dagger \in Y$. However, data y^δ are given only up to some known noise level $\delta > 0$ as

$$\left\| y^\delta - y^\dagger \right\|_Y \le \delta. \tag{1.2}$$

Date: Updated and revised version: May 13, 2013.
2010 *Mathematics Subject Classification.* 65J20, 47J06, 47A52, 49J40.
Key words and phrases. Inverse problems, Tikhonov-type regularization, discrepancy principle, parameter choice properties, convergence, exact penalization veto.

To compensate for the ill-posedness of the problem we use, for given $\alpha > 0$, as approximate solution the minimizers x_α^δ of the Tikhonov-type functional

$$J_{\alpha,y^\delta}(x) = \|F(x) - y^\delta\|_Y^p + \alpha\mathcal{R}(x), \tag{1.3}$$

for a non-negative penalty functional $\mathcal{R}$. Throughout this paper, let the exponent p be fixed with $1 \le p < \infty$.

In recent years there has been a strong interest in *a posteriori* rules $\alpha = \alpha(\delta, y^\delta)$ for choosing the regularization parameter when minimizing J_{α,y^δ} and, in particular, in variants of the discrepancy principle. It is a natural question to ask, whether for a given parameter choice rule $\alpha(\delta, y^\delta)$, the corresponding regularized solutions x_α^δ converge to a solution of (1.1). A first positive answer can be given in case the limit conditions

$$\alpha(\delta, y^\delta) \to 0 \qquad \text{and} \qquad \frac{\delta^p}{\alpha(\delta, y^\delta)} \to 0 \qquad \text{as } \delta \to 0, \tag{1.4}$$

hold, which highlights the importance of such asymptotic relations. Indeed, it is well-known that then a subsequence of the regularized solutions x_α^δ converges to exact solutions of (1.1) as $\delta \to 0$. This *convergence* does not necessarily occur in norm, but possibly with respect to a coarser topology in X such as the weak or weak* topology, for example. However, stronger results may consequently be obtained for suitable penalty terms or using additional knowledge about the solution.

Now, *a priori* parameter choices $\alpha = \alpha(\delta)$ may be easily constructed such that they fulfill the conditions (1.4) in order to take advantage of the convergence properties that go along with these asymptotics. The situation becomes more subtle for a posteriori rules. Using a continuous formulation of the discrepancy principle, where $\alpha = \alpha(\delta, y^\delta)$ is chosen such that

$$\tau_1\delta \le \|F(x_\alpha^\delta) - y^\delta\| \le \tau_2\delta \tag{1.5}$$

holds for prescribed $1 \le \tau_1 \le \tau_2$, convergence results were obtained in [4, 2] for linear and non-linear forward operators, respectively, and we refer to [1] for further details. Interestingly, the rather strong requirement (1.5) allows for proving convergence without knowing about the limit conditions (1.4). Even so, it has first been shown in [2] that (1.4) holds under mild assumptions which we generalize and extend for our purposes here. A main drawback of the formulation (1.5) is that such a discrepancy principle with upper and lower bounds may not always be feasible, especially for certain non-linear operators where duality gaps occur, see e.g. [1, 2, 13] and [26, p. 87].

The main result of the current study asserts that the parameter $\alpha = \alpha(\delta, y^\delta)$ chosen according to a sequential discrepancy principle (SDP), considered below, obeys (1.4) under two natural assumptions.

Namely, we assume that *exact penalization* is excluded and that the data are *compatible*. This means that the SDP stops after a finite number of steps. If the parameter choice SDP obeys (1.4) then the regularized solutions converge as $\delta \to 0$. If, in addition, a kind of *variational inequality* is valid then SDP also yields *convergence rates*. Such rates have been obtained for the continuous discrepancy principle (1.5) in [1, 3, 13], for example, and for the sequential discrepancy principle in [20].

The paper is organized as follows: In Section 2 we recall, for the convenience of the reader, the common assumptions for Tikhonov-type regularization in Banach space (cf., e.g. [19, 25, 26]), and summarize some mathematical consequences of those standard assumptions. We also introduce and discuss the two additional assumptions, the exact penalization veto, and the data compatibility. The major part is Section 3, in which the sequential discrepancy principle (SDP) is introduced and the main result, Theorem 1, is stated and proven. Variational inequalities ensuring convergence rates of regularized solutions are the subject of Section 4. Such inequalities combining solution smoothness and structural conditions concerning the nonlinearity of F allow us to bound the maximum decay rate of the regularization parameters $\alpha(\delta, y^\delta) \to 0$ obtained from SDP as $\delta \to 0$. Several technical proofs are collected in the final Section 5.

2. Assumptions and auxiliary results

In the following we formulate and discuss our standing assumptions. The proofs of Propositions 1–6 are postponed to Section 5.

Assumption 1 (basic assumptions)**.** *Let* $(X, \|.\|_X)$ *and* $(Y, \|.\|_Y)$ *be arbitrary Banach spaces with duals* X^*, Y^**, and let* τ_X, τ_Y *be topologies on* X, Y*, respectively, that are weaker than the norm topologies. Moreover we assume that*

A1 *The operator* $F : \mathrm{dom}(F) \subseteq X \to Y$ *is* τ_X-τ_Y *continuous, its domain* $\mathrm{dom}(F)$ *is convex and* τ_X*-sequentially closed.*

A2 *The penalty functional* $\mathcal{R} : X \to [0, \infty]$ *with domain*

$$\mathrm{dom}(\mathcal{R}) = \{x \in X \mid \mathcal{R}(x) < \infty\}$$

is proper, i.e. $\mathrm{dom}(\mathcal{R}) \neq \emptyset$, τ_X*-sequentially lower semicontinuous, and stabilizing in the sense that the sublevel sets*

$$\mathcal{M}^{\mathcal{R}}(c) = \left\{x \in X \mid \mathcal{R}(x) \le c\right\}$$

are τ_X*-sequentially precompact.*

A3 *The intersection* $\mathcal{D} := \mathrm{dom}(\mathcal{R}) \cap \mathrm{dom}(F)$ *is non-empty and we have*

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

A4 *The norm* $\|.\|_Y$ *is* τ_Y*-sequentially lower semicontinuous.*

Remark 1. Originally, Tikhonov regularization was studied for Hilbert spaces X, Y with τ_X, τ_Y the respective weak (or even strong) topologies and $\mathcal{R}(x) = \|x\|_X^2$. This classical setting can be extended under Assumption 1 to the case of Banach spaces X, Y, where frequently

$$\mathcal{R}(x) = \|x\|_X^q \tag{2.1}$$

is used (cf., e.g., [26, Chap. 5]) as penalty functional, which is convex for $1 \le q < \infty$. If the space X is reflexive then the unit ball is weakly sequentially precompact. Hence we can use the weak topology in X as τ_X, in which case $\mathcal{R}$ from (2.1) is stabilizing, thus **A2** holds. If, alternatively, X is a non-reflexive Banach space, then this is not the case and the weak*-topology connected with a separable predual Banach space Z ($X = Z^*$) has to be exploited as τ_X in order to ensure the stabilizing property of $\mathcal{R}$ from (2.1). The latter is for example the case for ℓ^1-regularization (cf., e.g., [7, 5]) and total variation regularization in BV (cf., e.g., [8, 10]). Note that then **A1** requires F to be weak-to-weak continuous, or weak*-to-weak continuous, respectively.

Assumption 1 also covers sparsity promoting regularization, where the focus is on convex penalties

$$\mathcal{R}(x) = \sum_{n\in\mathbb{N}} |\langle x, \varphi_n\rangle|^q \tag{2.2}$$

with exponents $1 \le q < \infty$ and for some basis $\{\varphi_n\}_{n\in\mathbb{N}}$ from the Hilbert space X, see e.g., [1, 8, 18, 25] for details. Even the extension to non-convex sublinear penalties (2.2) with $0 < q < 1$ is admissible, where **A2** is satisfied as well (cf., e.g., [14, 22, 24, 28]).

For **A4** to hold the topology τ_Y should be no coarser than the weak topology on Y.

We collect without proofs some properties of solutions and regularized solutions as well as some auxiliary results needed below which all are valid under Assumption 1. In the sequel we will write $\|\cdot\|$ instead of $\|\cdot\|_X$ and $\|\cdot\|_Y$ if the space is clear from the context.

We introduce, for any $y \in Y$ and parameter $\alpha > 0$ the set

$$\mathcal{M}_{\alpha,y} := \left\{x \in \mathcal{D} \;\middle|\; J_{\alpha,y}(x) \le J_{\alpha,y}(z) \text{ for all } z \in \mathcal{D}\right\}. \tag{2.3}$$

Evidently, all regularized solutions x_α^δ belong to $\mathcal{M}_{\alpha,y^\delta}$. For the derivation of the following assertions we refer the interested reader, e.g., to [19, Section 3], [26, Section 4.1.1], and [12, 13, 25].

Facts 1.

(a) *The sets* $\mathrm{dom}(\mathcal{R})$ *and* $\mathcal{D}$ *are* τ_X*-sequentially closed subsets of* X.

(b) *For all* $\alpha > 0$ *and* $y^\delta \in Y$ *we have that* $\mathcal{M}_{\alpha,y^\delta} \neq \emptyset$. *Thus, the regularized solutions* x_α^δ *minimizing* $J_{\alpha,y^\delta}(x)$ *over* $x \in \mathcal{D}$ *exist, and they are stable with respect to perturbations of the data* y^δ.

(c) *The value*

$$\mathcal{R}_{\min} := \min_{x \in \mathcal{D}} \mathcal{R}(x) \geq 0 \tag{2.4}$$

exists, and the set

$$X_{\min} := \{x \in \mathcal{D} \mid \mathcal{R}(x) = \mathcal{R}_{\min}\} \neq \emptyset \tag{2.5}$$

is τ_X*-sequentially closed and precompact. If moreover* $\mathcal{R}$ *is a convex functional, then* $X_{\min}$ *is a convex subset of* X*.*

(d) *Solutions* $x^\dagger \in \mathcal{D}$ *of equation* (1.1) *satisfying*

$$\mathcal{R}(x^\dagger) = \min \left\{\mathcal{R}(x) \mid F(x) = y^\dagger, \;\; x \in \mathcal{D}\right\},$$

which are called $\mathcal{R}$*-minimizing solutions, always exist and we denote by* $\mathcal{L}$ *the set of all* $\mathcal{R}$*-minimizing solutions.*

(e) *For all* $y \in Y$ *the functional* $\zeta(x) := \|F(x) - y\|$ *is* τ_X*-sequentially lower semicontinuous on* $\mathrm{dom}(F)$*.*

(f) *Let* $Y_{\min} := F(X_{\min})$*. For all* $y \in Y$ *there exists* $x_{\min} \in X_{\min}$*, such that*

$$d(y, Y_{\min}) := \inf_{x \in X_{\min}} \|F(x) - y\| = \|F(x_{\min}) - y\| .$$

For the subsequent analysis it will be important that the set $\mathcal{L}$ of $\mathcal{R}$-minimizing solutions, the set $X_{\min}$ of minimizers of the penalty $\mathcal{R}$ over $\mathcal{D}$, and the set $\mathcal{M}_{\alpha,y^\dagger}$ of regularized solutions in the noise-free case are apart (see the definition of $\mathcal{M}_{\alpha,y}$ from (2.3)). In this context, it is helpful to avoid *exact penalization* (cf. [9]), and therefore we make the following assumption.

Assumption 2 (EP veto)**.** *Let at the right-hand side* $y^\dagger$ *in* (1.1) *the exact penalization veto (EP veto) be satisfied, which means that for all* $\alpha > 0$ *the implication*

$$x^\dagger \in \mathcal{L} \;\wedge\; x^\dagger \in \mathcal{M}_{\alpha,y^\dagger} \quad \Longrightarrow \quad x^\dagger \in X_{\min}$$

is true. In other words, we assume that $\mathcal{L} \cap \bigcup_{\alpha>0} \mathcal{M}_{\alpha,y^\dagger} \subseteq X_{\min}$*.*

With the following three propositions we provide, for the case $p > 1$, and under different requirements on F and $\mathcal{R}$, handy sufficient conditions for Assumption 2 to hold. Under these conditions the main result, Theorem 1, to be formulated in Section 3, ensures the asymptotics (1.4). We also refer to [2, Lemma 4.8 and 4.9], where conditions of similar nature are exploited to obtain the asymptotics (1.4) for a stronger formulation of the discrepancy principle. Note that the case $p = 1$ is always suspicious for violating the EP veto, and we refer for illustration to Proposition 11 at the end of Section 4.

Proposition 1. *Let $p > 1$ and suppose that for all $x \in \mathcal{D}$ there exists a bounded linear operator $F'(x) : X \to Y$ such that*

$$\lim_{t \to +0} \frac{F\big(x + th\big) - F\big(x\big)}{t} = F'(x)h \tag{2.6}$$

holds for all $h \in X$ satisfying $x + th \in \mathcal{D}$ for sufficiently small $t > 0$. Then the EP veto is satisfied for arbitrary $y^\dagger \in F\big(\mathcal{D}\big)$ whenever $\mathcal{R}$ is a convex functional.

Remark 2. The condition (2.6) is weaker than Gâteaux differentiability of F at x for all $x \in \mathcal{D}$, because not all directions $h \in X$ are concerned; and hence x is not necessarily an interior point of $\mathcal{D}$. It is enough that $\mathcal{D}$ is a convex set in X. We thus include the practically important case of 'half-spaces' in $X := L^r(\Omega)$, $1 \le r < \infty$, of the form

$$\mathrm{dom}(F) := \{x \in X \mid x(s) \ge 0 \text{ for almost all } s \in \Omega\}$$

as domain of F, which do not possess interior points at all.

Proposition 1 needs the *convexity* of $\mathcal{R}$ in order to ensure the EP veto. In contrast, the subsequent Propositions 2 and 3 assume alternative properties of $\mathcal{R}$ and partly Gâteaux differentiability of F.

Proposition 2. *Let $p > 1$, $0 \in \mathcal{D}$ with $\mathcal{R}(0) = 0$, and suppose that for all $x \in \mathcal{D}$ there exists a bounded linear operator $F'(x) : X \to Y$ such that*

$$\lim_{t \to +0} \frac{F\big(x + th\big) - F\big(x\big)}{t} = F'(x)h \tag{2.7}$$

holds for all $h \in X$ satisfying $x + th \in \mathcal{D}$ for sufficiently small $t > 0$. Moreover let there exist a function $\theta : (0, 1) \to [0, 1)$ such that

$$\liminf_{t \to +0} \frac{t}{1 - \theta(1 - t)} = C_\theta < \infty \tag{2.8}$$

and

$$\mathcal{R}(\mu x) \le \theta(\mu)\, \mathcal{R}(x) \quad \text{for all } x \in \mathcal{D} \text{ and } 0 < \mu < 1. \tag{2.9}$$

Then the EP veto holds for arbitrary $y^\dagger \in F\big(\mathcal{D}\big)$.

Remark 3. Proposition 2 allows us to show the validity of Assumption 2 also for frequently used *q-homogeneous, non-convex penalties* $\mathcal{R}$ such as (2.1) and (2.2) with $0 < q < 1$. In this case we let $\theta(\mu) = \mu^q$ and hence $C_\theta = \lim\limits_{t \to +0} \frac{t}{1-(1-t)^q} = \frac{1}{q}$. In this context, we mention the family of penalty functionals

$$\mathcal{R}(x) = \sum_{n \in \mathbb{N}} w_n \psi(\,|\langle x, \varphi_n \rangle|), \qquad 0 < \underline{w} \le w_n \text{ for all } n \in \mathbb{N}, \tag{2.10}$$

occurring in sparsity promoting regularization as a generalization of (2.2) and discussed in [6, 15]. If the function $\psi : [0, \infty) \to [0, \infty)$ is lower semicontinuous with $\psi(0) = 0$, $\lim\limits_{t \to \infty} \psi(t) = \infty$, and if there is some $C > 0$ for which $\psi(t) \geq Ct^2/(1+t^2)$, then $\mathcal{R}$ is a τ_X-weakly semi-continuous and stabilizing functional in the sense of **A2** from Assumption 1, for the weak topology τ_X in the Hilbert space X. Proposition 2 applies to members of this family of penalties whenever there exists a function $\theta : (0,1) \to [0,1)$ such that (2.8) and

$$\psi(\mu t) \leq \theta(\mu)\, \psi(t) \quad \text{for all} \quad t > 0 \text{ and } 0 \leq \mu < 1.$$

Then as a consequence of Theorem 1 below, also non-convex penalties (2.10) generated by such ψ yield the limit relations (1.4) when using the sequential discrepancy principle. These relations are required for example in [6, Thm. 2.3] and [15, Prop. 4.3] in order to obtain convergence of regularized solutions.

The stochastically motivated Cauchy functional is another example of a non-convex penalty that is of the form (2.10) with

$$\psi(t) = \log(1 + \omega t^2), \qquad \omega > 0, \tag{2.11}$$

and the validity of **A2** from Assumption 1 is shown in [23, Chap. 7]. However, Proposition 2 does not apply to ψ from (2.11) as there is no function θ satisfying the conditions (2.8) and (2.9). But, in contrast to (2.1) and (2.2) with $0 < q \leq 1$, the penalty $\mathcal{R}$ based on (2.11) is Gâteaux differentiable and we have $\nabla\mathcal{R}(x) \neq 0$ for $x \neq 0$. Thus, the following Proposition 3 ensures that Assumption 2 holds for the Cauchy functional.

Proposition 3. *Let $p > 1$ and let $\mathcal{R}$ be Gâteaux differentiable with Gâteaux derivative $\nabla\mathcal{R}(x) \neq 0$ for all $x \in \mathcal{D} \setminus X_{\min}$. Moreover, suppose that $y^\dagger \in F\big(\mathcal{D}\big)$ is such that F is Gâteaux differentiable with Gâteaux derivative $F'(x)$ for all $x \in \mathcal{L} \setminus X_{\min}$. Then the EP veto is satisfied for such $y^\dagger \in F\big(\mathcal{D}\big)$.*

When using the discrepancy principle, we are, conceptually speaking, interested in finding the largest value $\alpha > 0$ such that for prescribed $\tau > 1$

$$\left\| F\big(x_\alpha^\delta\big) - y^\delta \right\| \leq \tau\delta \tag{2.12}$$

(or the largest such $\alpha \in \Delta_q$ in the discrete formulation, below). In this context we mention the following properties of the discrepancy functional.

Proposition 4. *Let $y^\delta \in Y$ and, for each $\alpha > 0$, $x_\alpha^\delta \in \mathcal{M}_{\alpha,y^\delta}$ be arbitrary but fixed. Then, the functional*

$$g(\alpha) = \left\| F\big(x_\alpha^\delta\big) - y^\delta \right\|, \quad \alpha > 0,$$

is non-decreasing. Moreover, it holds that

$$\lim_{\alpha \to +0} g(\alpha) = \inf_{x \in \mathcal{D}} \left\| F\big(x\big) - y^\delta \right\| \quad \textit{and} \quad \lim_{\alpha \to \infty} g(\alpha) \geq d(y^\delta, Y_{\min}).$$

In order to make sure that a largest finite $0 < \alpha < \infty$ with $\|F(x_\alpha^\delta) - y^\delta\| \le \tau\delta$ exists, we impose the following assumption. Recall that we have $\|y^\dagger - y^\delta\| \le \delta$ from (1.2).

Assumption 3 (data compatibility)**.** *For prescribed $\tau > 1$ there is some $\bar\delta > 0$ such that the data $y^\delta \in Y$ satisfy*

$$\tau\delta < d(y^\delta, Y_{\min}) \qquad \textit{for all} \quad 0 < \delta \le \bar\delta. \tag{2.13}$$

There are intuitive and handy conditions which ensure the validity of both the Assumptions 2 and 3.

Proposition 5. *Suppose that*

$$\mathcal{L} \cap \left(X_{\min} \cup \bigcup_{\alpha>0} \mathcal{M}_{\alpha,y^\dagger} \right) = \emptyset, \tag{2.14}$$

then Assumptions 2 and 3 hold. If $\mathcal{L} \cap X_{\min} \ne \emptyset$ then Assumption 3 cannot hold.

It is seen from Proposition 5 that (2.13) excludes $x^\dagger \in X_{\min}$. Related to this the following observation is interesting.

Proposition 6. *If $\tau > 1$ is prescribed such that $\|F(x_{\min}) - y^\delta\| \le \tau\delta$, for some $x_{\min} \in X_{\min}$, then $\|F(x_\alpha^\delta) - y^\delta\| \le \tau\delta$ for all $\alpha > 0$.*

Thus, in the absence of (2.13) there is no chance to stop the sequential discrepancy principle. Actually, in the case considered in Proposition 6, we would have no choice but to pick $\alpha = +\infty$, and

$$x_\alpha^\delta \in \operatorname*{arg\,min}_{x_{\min} \in X_{\min}} \|F(x_{\min}) - y^\delta\|$$

as the corresponding regularized solution. To avoid this degenerate case, we restrict our attention to data satisfying Assumption 3.

We close this preliminary section with the following important result, proven in [27] in a more general framework, and we also refer to [26, Section 4.1.2].

Lemma 1. *Assume that the sequence of positive noise levels $\{\delta_n\}$ tends to zero as $n \to \infty$. If a corresponding sequence $\{x_n\} \subseteq \mathcal{D}$ satisfies the limit conditions*

$$\lim_{n\to\infty} \|F(x_n) - y^{\delta_n}\| = 0 \qquad \textit{and} \qquad \limsup_{n\to\infty} \mathcal{R}(x_n) \le \mathcal{R}(x^\dagger). \tag{2.15}$$

for $x^\dagger \in \mathcal{L}$, then $\{x_n\}$ is in the sense of subsequences τ_X-convergent to elements of $\mathcal{L}$ and we have $\lim_{n\to\infty} \mathcal{R}(x_n) = \mathcal{R}(x^\dagger)$.

This result, together with the standard pair of inequalities,

$$\|F(x_\alpha^\delta) - y^\delta\|^p \le \delta^p + \alpha\,\mathcal{R}(x^\dagger), \qquad \alpha > 0, \tag{2.16}$$

and

$$\mathcal{R}(x_\alpha^\delta) \le \frac{\delta^p}{\alpha} + \mathcal{R}(x^\dagger), \qquad \alpha > 0, \tag{2.17}$$

valid for Tikhonov regularization, yields τ_X-convergence of the regularized solutions to elements of the set $\mathcal{L}$ of $\mathcal{R}$-minimizing solutions whenever the regularization parameters $\alpha(\delta, y^\delta)$ obey the asymptotics (1.4).

3. The sequential discrepancy principle and the main result

We start by defining the sequential discrepancy principle for choosing the regularization parameter α. For prescribed $0 < q < 1$ and $\alpha_0 > 0$, we let

$$\Delta_q := \{\alpha_j \mid \alpha_j = q^j \alpha_0, \quad j \in \mathbb{Z}\}.$$

Given any $\delta > 0$ and data y^δ, the sublevel sets $\mathcal{M}_{\alpha,y^\delta}$ are non-empty. From now on we fix some selection $x_\alpha^\delta \in \mathcal{M}_{\alpha,y^\delta}$, $\alpha \in \Delta_q$.

Definition 1 (sequential discrepancy principle)**.** We say that an element $\alpha \in \Delta_q$ is chosen according to the *sequential discrepancy principle (SDP)*, if

$$\left\|F\left(x_\alpha^\delta\right) - y^\delta\right\| \le \tau\delta < \left\|F\left(x_{\alpha/q}^\delta\right) - y^\delta\right\|. \tag{3.1}$$

It must be shown that the SDP from Definition 1 can be satisfied under data compatibility introduced by Assumption 3.

Lemma 2. *Under Assumption 3 the SDP is feasible, i.e., for all $0 < \delta \le \overline{\delta}$ there exists a unique $j = j(\delta, y^\delta) \in \mathbb{Z}$ such that* (3.1) *holds for $\alpha = \alpha_j \in \Delta_q$.*

Proof. This is a consequence of Proposition 4. Indeed, from (2.13) and the asymptotic relations in Proposition 4 we know that

$$\begin{aligned}\lim_{\alpha\to+0} \left\|F\left(x_\alpha^\delta\right) - y^\delta\right\| &= \inf_{x\in\mathcal{D}} \left\|F\left(x\right) - y^\delta\right\| \le \delta < \tau\delta < d(y^\delta, Y_{\min}) \\ &\le \lim_{\alpha\to\infty} \left\|F\left(x_\alpha^\delta\right) - y^\delta\right\|.\end{aligned}$$

This ensures that there always exists $j \in \mathbb{Z}$ such that (3.1) holds. □

Now we are ready to formulate the main result concerning the asymptotic behavior of regularization parameters $\alpha = \alpha(\delta, y^\delta)$, chosen by the sequential discrepancy principle, when δ tends to zero.

Theorem 1. *Under the Assumptions 1, 2, and 3 there is some $\overline{\delta} > 0$ such that regularization parameters $\alpha = \alpha(\delta, y^\delta)$ chosen according to the sequential discrepancy principle (SDP) exist for all $0 < \delta \le \overline{\delta}$. These parameters satisfy the limit conditions*

$$\alpha(\delta, y^\delta) \to 0 \qquad \text{and} \qquad \frac{\delta^p}{\alpha(\delta, y^\delta)} \to 0 \qquad \text{as} \quad \delta \to 0. \tag{3.2}$$

Then the associated regularized solutions $x^\delta_{\alpha(\delta,y^\delta)}$ *are in the sense of subsequences* τ_X*-convergent to elements of* $\mathcal{L}$ *as* $\delta \to 0$*, and we have* $\lim_{\delta\to 0} \mathcal{R}(x^\delta_{\alpha(\delta,y^\delta)}) = \mathcal{R}(x^\dagger)$.

The existence of the regularization parameter according to the SDP, under the assumptions of Theorem 1, was shown in Lemma 2. Now, we still have to establish the limit behavior from (3.2). Both limit conditions are immediate consequences of the two propositions formulated and proven, next. Then, the τ_X-convergence of the regularized solutions immediately follows from (3.2) together with (2.16) and (2.17).

Proposition 7. *Under the assumptions of Theorem 1 the parameters* $\alpha(\delta, y^\delta)$*, chosen according to the SDP, obey*

$$\alpha(\delta, y^\delta) \to 0 \qquad \text{as} \qquad \delta \to 0.$$

Proof. Let $\overline{\delta} \geq \delta_n \to 0$ and $\alpha_n = \alpha(\delta_n, y^{\delta_n})$ be chosen according to the SDP. As a shorthand we write $x_n = x^{\delta_n}_{\alpha_n}$ for the corresponding regularized solutions satisfying (3.1).

Assume to the contrary that there is a subsequence of $\{\alpha_{n_k}\}$ of $\{\alpha_n\}$, and a constant $\underline{\alpha} > 0$ such that $\alpha_{n_k} \geq \underline{\alpha},\ k \in \mathbb{N}$. If we denote by

$$\bar{x}_{n_k} = \operatorname*{arg\,min}_{x\in\mathcal{D}} \left\{ \left\|F(x) - y^{\delta_{n_k}}\right\|^p + \underline{\alpha}\mathcal{R}(x) \right\}$$

the minimizers of $J_{\underline{\alpha},y^{\delta_{n_k}}}$ and use the Proposition 4 and (3.1), then we have that

$$\left\|F(\bar{x}_{n_k}) - y^{\delta_{n_k}}\right\| \leq \left\|F(x_{n_k}) - y^{\delta_{n_k}}\right\| \leq \tau_2\, \delta_{n_k} \to 0,$$

and

$$\limsup_{k\to\infty} \underline{\alpha}\,\mathcal{R}(\bar{x}_{n_k}) \leq \limsup_{k\to\infty} \{ \left\|F(\bar{x}_{n_k}) - y^{\delta_{n_k}}\right\|^p + \underline{\alpha}\,\mathcal{R}(\bar{x}_{n_k}) \} \leq \underline{\alpha}\,\mathcal{R}(x^\dagger).$$

Therefore, $\{\bar{x}_{n_k}\}$ satisfies the assumptions of Lemma 1 and we can extract a subsubsequence $\{\bar{x}_{n_{k_l}}\}$ which is τ_X-convergent for $l \to \infty$ to some element, say $z \in \mathcal{L}$. Because of the τ_X-sequential lower semicontinuity of the functionals $\mathcal{R}$ and ζ (cf. Fact 1(e)) it holds for any $x \in \mathcal{D}$ that

$$\begin{aligned}
\left\|F(z) - y^\dagger\right\|^p + \underline{\alpha}\mathcal{R}(z) &\leq \liminf_{l\to\infty} \left(\left\|F(\bar{x}_{n_{k_l}}) - y^{\delta_{n_{k_l}}}\right\|^p + \underline{\alpha}\mathcal{R}(\bar{x}_{n_{k_l}}) \right) \\
&\leq \liminf_{l\to\infty} \left(\left\|F(x) - y^{\delta_{n_{k_l}}}\right\|^p + \underline{\alpha}\mathcal{R}(x) \right) \\
&= \left\|F(x) - y^\dagger\right\|^p + \underline{\alpha}\mathcal{R}(x),
\end{aligned}$$

which shows that $z \in \mathcal{M}_{\underline{\alpha},y^\dagger}$. Assumption 2 implies $z \in \mathcal{L} \cap X_{\min}$ which, according to Proposition 5, violates Assumption 3 and we have reached a contradiction. □

Proposition 8. *Under the assumptions of Theorem 1 the parameters $\alpha(\delta, y^\delta)$, chosen according to the SDP, satisfy the limit condition*

$$\frac{\delta^p}{\alpha(\delta, y^\delta)} \to 0 \qquad as \quad \delta \to 0.$$

Proof. Let $\overline{\delta} \geq \delta_n \to 0$ and $\alpha_n = \alpha(\delta_n, y^{\delta_n}) \in \Delta_q^{(n)}$ be chosen according to the SDP. Now, let $x_n \in \mathcal{M}_{\alpha_n, y^{\delta_n}}$ and $x_n^{(q)} \in \mathcal{M}_{\alpha_n/q, y^{\delta_n}}$ be such that

$$\left\|F\left(x_n\right) - y^{\delta_n}\right\| \leq \tau\delta_n < \left\|F\left(x_n^{(q)}\right) - y^{\delta_n}\right\|$$

holds. Due to the minimizing property of $x_n^{(q)}$ we thus obtain for any $x^\dagger \in \mathcal{L}$

$$\begin{aligned}(\tau\delta_n)^p + \frac{\alpha_n}{q}\mathcal{R}(x_n^{(q)}) &\leq \left\|F\left(x_n^{(q)}\right) - y^{\delta_n}\right\|^p + \frac{\alpha_n}{q}\mathcal{R}(x_n^{(q)}) \\ &\leq \left\|F\left(x^\dagger\right) - y^{\delta_n}\right\|^p + \frac{\alpha_n}{q}\mathcal{R}(x^\dagger) \\ &\leq \delta_n^p + \frac{\alpha_n}{q}\mathcal{R}(x^\dagger),\end{aligned}$$

where we have used $\left\|F\left(x^\dagger\right) - y^{\delta_n}\right\| = \|y - y^{\delta_n}\| \leq \delta_n$. Hence we get the estimate

$$0 \leq q(\tau^p - 1)\frac{\delta_n^p}{\alpha_n} \leq \mathcal{R}(x^\dagger) - \mathcal{R}(x_n^{(q)}). \tag{3.3}$$

In particular we infer that $\limsup_{n\to\infty} \mathcal{R}(x_n^{(q)}) \leq \mathcal{R}(x^\dagger)$.

Also, from the minimizing properties we see that

$$\left\|F\left(x_n^{(q)}\right) - y^{\delta_n}\right\|^p + \frac{\alpha_n}{q}\mathcal{R}(x_n^{(q)}) \leq \left\|F\left(x^\dagger\right) - y^{\delta_n}\right\|^p + \frac{\alpha_n}{q}\mathcal{R}(x^\dagger),$$

which implies that

$$\left\|F\left(x_n^{(q)}\right) - y^{\delta_n}\right\|^p \leq \delta^p + \frac{\alpha_n}{q}\mathcal{R}(x^\dagger).$$

From Proposition 7 we obtain that $\left\|F\left(x_n^{(q)}\right) - y^{\delta_n}\right\|^p \to 0$ as $n \to \infty$ and thus Lemma 1 yields $\lim_{n\to\infty} \mathcal{R}(x_n^{(q)}) = \mathcal{R}(x^\dagger)$, which in turn by virtue of (3.3) allows to complete the proof. □

Remark 4. Here we have used that, under our assumptions, α obtained from the SDP tends to zero in order to obtain $\delta^p/\alpha \to 0$ as $\delta \to 0$. However, the latter also remains true if α is bounded from below by some constant $\bar{\alpha} > 0$. When combining $\left\|F\left(x_\alpha^\delta\right) - y^\delta\right\| \leq \tau\delta$ from (3.1) and (2.17) we easily see that Lemma 1 is applicable under such lower bound, too. Thus, the regularized solutions x_α^δ converge with respect to the τ_X-topology to elements of $\mathcal{L}$ whenever $\alpha(\delta, y^\delta) \to 0$ or $\alpha(\delta, y^\delta) \geq \bar{\alpha} > 0$ can be ensured for the problem at hand.

At the end of this section we provide two simple examples illustrating that the regularization parameters $\alpha > 0$ chosen according to the SDP need not tend to zero as $\delta > 0$ tends to zero if the assumptions of Theorem 1 are not completely fulfilled. Precisely, the EP veto of Assumption 2 is not satisfied in both examples. In particular, for differentiable F, Example 1 refers to the exponent $p = 1$ in the misfit term $\|F(x) - y^\delta\|^p$ of the Tikhonov functional $J_{\alpha,y^\delta}(x)$, which violates one main assumption of Proposition 1. Another variety of violating the EP veto is presented in Example 2 (cf. [1, 2]), working with the exponent $p = 2$, but F does not meet the differentiability requirement (2.6) of Proposition 1.

Example 1. For $X = Y = \mathbb{R}$ with norms $\|\cdot\| := |\cdot|$ we consider here $\mathcal{R}(x) = |x|$, $F(x) := x,\ x \in \mathcal{D} = \mathbb{R}$, $x^\dagger = y^\dagger = 1$, and $y^\delta = 1 \pm \delta$ for $0 < \delta \le 1/3$. When setting the exponent $p := 1$ in the Tikhonov functional

$$J_{\alpha,y^\delta}(x) = |x - (1 \pm \delta)| + \alpha|x|,$$

for $0 < \alpha < 1$ the uniquely determined regularized solution is $x_\alpha^\delta = 1 \pm \delta$, and moreover for any $\tau > 1$ we have $\|F(x_\alpha^\delta) - y^\delta\| = 0 < \tau\delta$. For $\alpha = 1$ the closed interval $\mathcal{M}_{\alpha,y^\delta} = [0, 1 \pm \delta]$ characterizes the regularized solutions and the corresponding values $\|F(x_\alpha^\delta) - y^\delta\|$ run through the same interval. On the other hand, for $\alpha > 1$ we have $x_\alpha^\delta = 0$ with $\|F(x_\alpha^\delta) - y^\delta\| = 1 \pm \delta$ which dominates the value $\tau\delta$ if $\delta < 1/(\tau + 1)$. For $\alpha_0 > 1$ and $0 < q < 1$ we always find some $j \in \mathbb{N}$ such that $\alpha_j = \alpha_0 q^j < 1$ satisfies the SDP and hence that the regularization parameter remains constant and positive for all sufficiently small $\delta > 0$. Note that we have here $\mathcal{L} = \{1\}$, $X_{\min} = \{0\}$, $\mathcal{L} \cap X_{\min} = \emptyset$,

$$\mathcal{M}_{\alpha,y^\dagger} = \begin{cases} \mathcal{L} & \text{if } \ 0 < \alpha < 1, \\ [0,1] & \text{if } \ \alpha = 1, \\ X_{\min} & \text{if } \ \alpha > 1, \end{cases}$$

and hence $\mathcal{L} \cap \bigcup_{\alpha > 0} \mathcal{M}_{\alpha,y^\dagger} = \{1\} \not\subseteq X_{\min}$ which violates the EP veto.

Example 2. For obtaining this example we only amend Example 1 in a few details, namely we set $p := 2$ and use the function

$$F(x) := 1 + \sqrt{|1 - x|}, \qquad x \in \mathbb{R},$$

which is non-differentiable at the solution point $x = 1$. Moreover we consider data $y^\delta = 1 + \delta,\ 0 < \delta < 1/3$, such that the Tikhonov functional attains the form

$$J_{\alpha,y^\delta}(x) = (\sqrt{|1 - x|} - \delta)^2 + \alpha|x|,$$

and we have for $0 < \alpha < 1 - \delta$ the uniquely determined regularized solution $x_\alpha^\delta = 1 - \frac{\delta^2}{(1-\alpha)^2}$ with $\left\|F\big(x_\alpha^\delta\big) - y^\delta\right\| = \frac{\alpha\delta}{1-\alpha}$ (see also [1]).

For $\tau \in (1,2)$ and any $q<1$, the SDP will always select $\alpha \in \Delta_q$ such that

$$0 < q\frac{\tau}{\tau+1} < \alpha \le \frac{\tau}{\tau+1},$$

and α does not tend to zero as $\delta \to 0$. Since the sets $\mathcal{L}$, $X_{\min}$ and $\mathcal{M}_{\alpha,y^\dagger}$ are the same as in Example 1, the same conclusions concerning the violation of the EP veto can be drawn, although the reason in Example 2 is now the non-differentiability of F.

4. Impact on rates for the parameter choice under variational inequalities

In Proposition 7 we have shown that under the Assumptions 1–3 there is some $\bar{\delta} > 0$ such that regularization parameters $\alpha = \alpha(\delta, y^\delta)$ satisfying the sequential discrepancy principle (SDP) exist for all $0 < \delta \le \bar{\delta}$ and tend to zero as $\delta \to 0$. Moreover, $\delta^p/\alpha(\delta, y^\delta) \to 0$. The latter has an important consequence. Let $\delta > 0$ and $\alpha > 0$ be fixed. Then, for the true solution $x^\dagger$, we have that $J_{\alpha,y^\delta}\left(x_\alpha^\delta\right) \le J_{\alpha,y^\delta}\left(x^\dagger\right)$, which in particular implies that $\mathcal{R}(x_\alpha^\delta) - \mathcal{R}(x^\dagger) \le \delta^p/\alpha$. This yields that the minimizers x_α^δ belong to certain sublevel sets, specifically

$$x_\alpha^\delta \in \mathcal{M}^{\mathcal{R}}\left(\mathcal{R}(x^\dagger) + \frac{\delta^p}{\alpha}\right).$$

From Theorem 1 we deduce that for every $c > \mathcal{R}(x^\dagger)$ there is $\bar{\delta} > 0$ such that for $0 < \delta \le \bar{\delta}$ the minimizer x_α^δ with parameter $\alpha(\delta, y^\delta)$ chosen according to the sequential discrepancy principle, obeys $x_\alpha^\delta \in \mathcal{M}^{\mathcal{R}}(c)$. That assertion holds without using any additional condition on the $\mathcal{R}$-minimizing solutions $x^\dagger \in \mathcal{L}$ to which the corresponding regularized solutions x_α^δ converge (in the sense of subsequences) with respect to the τ_X-toplogy in the Banach space X.

This has implications for the results which were established in [20]. In that study the authors discuss, among others, the sequential discrepancy principle under the validity of some kind of variational inequalities, and we briefly recall this concept. The goal is to establish results beyond convergence, and to turn to convergence rates results. If one requires convergence rates measured by a non-negative error measure $E(x, x^\dagger)$ then smoothness conditions have to be imposed on the solutions $x^\dagger$ which fit to the (non-linearity) structure of the forward operator F. An appropriate way of combining such conditions on smoothness and non-linearity is provided by the variational inequality approach, where the solution $x^\dagger$ fulfills the inequality

$$\beta E(x, x^\dagger) \le \mathcal{R}(x) - \mathcal{R}(x^\dagger) + \varphi(\|F(x) - F(x^\dagger)\|) \qquad \text{for all} \quad x \in \mathcal{M}, \tag{4.1}$$

with a constant $\beta > 0$, some concave index functions φ (strictly increasing continuous function $\varphi: (0,\infty) \to (0,\infty)$ satisfying the limit condition $\lim_{t\to+0} \varphi(t) = 0$), and for some set $\mathcal{M}$ containing $x^\dagger$.

Remark 5. The considerations in [20] refer to convex functionals $\mathcal{R}$, but inequalities of the form (4.1) also occur for non-convex penalties (cf., e.g., [6, 16]). The most prominent choice for the error measure E is, however, the *Bregman distance* for *convex* $\mathcal{R}$, which was introduced to regularization theory in [9]. For $x^\dagger \in \mathcal{L}$ and $\xi \in \partial\mathcal{R}(x^\dagger)$ it is defined as

$$D_\xi^{\mathcal{R}}(x, x^\dagger) := \mathcal{R}(x) - \mathcal{R}(x^\dagger) - \langle \xi, x - x^\dagger \rangle_{X^* \times X}. \tag{4.2}$$

It is well-known that, for $E(x, x^\dagger) = D_\xi^{\mathcal{R}}(x, x^\dagger)$, variational inequalities (4.1) only make sense with *concave* index functions φ and $0 < \beta \le 1$ provided that the operator F is Gâteaux differentiable at $x^\dagger$ (cf. [13, 21, 25]). Therefore we restrict our considerations to such φ and β. However, recent results from [17] show that non-concave φ, for example $\varphi(t) = C\,t^2$, in (4.1) are possible if the operator F is non-differentiable.

Under the variational inequality (4.1) and for $p > 1$ it was proven in [20, Thm. 2] that for every non-negative error measure E we have a convergence rate

$$E(x_\alpha^\delta, x^\dagger) = \mathcal{O}(\varphi(\delta)) \qquad \text{as} \quad \delta \to 0 \tag{4.3}$$

whenever the sequential discrepancy principle is used for choosing the regularization parameter $\alpha = \alpha(\delta, y^\delta)$. However, those results could be formulated only under the restrictive requirement that $x_\alpha^\delta \in \mathcal{M}$ *for all* $0 < \delta \le \bar{\delta}$, see Theorem 2, ibid.

Now in the light of Theorem 1 and under Assumptions 1–3 we can extend this as follows.

Proposition 9. *Suppose that a variational inequality (4.1) holds true on a set* $\mathcal{M}$. *If there is some* $c > \mathcal{R}(x^\dagger)$ *such that* $\mathcal{M} \supseteq \mathcal{M}^{\mathcal{R}}(c)$, *then there is a* $\bar{\delta} > 0$ *such that for every* $0 < \delta \le \bar{\delta}$ *we have that* $x_\alpha^\delta \in \mathcal{M}$, *where* $\alpha = \alpha(\delta, y^\delta)$ *is chosen according to the SDP. Consequently, for* $p > 1$ *the rate result (4.3) is valid.*

From [20, Cor. 2] we obtain for $p > 1$ the following δ-dependent lower bound for the associated regularization parameters $\alpha > 0$. To this end, we assign to the function φ from (4.1) the related function

$$\Phi(t) := \frac{t^p}{\varphi(t)}, \quad t > 0. \tag{4.4}$$

We note that for $p > 1$, and because φ is concave, the function Φ is an index function. This function controls the decay rate of the regularization parameter α chosen according to the SDP, when a variational inequality holds true.

Proposition 10 (cf. [20, Cor. 2]). *If for $p > 1$ and under the Assumptions 1–3 the $\mathcal{R}$-minimizing solution $x^\dagger$ satisfies a variational inequality (4.1) and $\alpha = \alpha(\delta, y^\delta)$ is chosen by the sequential discrepancy principle such that $x_\alpha^\delta \in \mathcal{M}$ for all $0 < \delta \le \overline{\delta}$, then we have the lower bound*

$$\alpha(\delta, y^\delta) \ge \frac{q}{2^{p-1}} \frac{\tau^p - 1}{\tau^p + 1} \Phi((\tau - 1)\delta), \quad 0 < \delta \le \bar{\delta}. \tag{4.5}$$

An inspection of the proof of Corollary 2 in [20] shows that the convergence rate result (4.3) remains true for $p = 1$, whenever the function Φ from (4.4) is an index function.

In the alternative situation when $p = 1$ and when the function φ is of the form $\varphi(t) = Ct,\ t > 0$, with some $C > 0$, then the related function Φ attains the constant value $1/C > 0$, for all $t > 0$. Thus, Φ does not constitute an index function. Also, the sequential discrepancy principle yields regularization parameters $\alpha(\delta, y^\delta)$, which are bounded below by a positive constant, i.e.,

$$\alpha(\delta, y^\delta) \ge \frac{q(\tau - 1)}{C(\tau + 1)}, \quad 0 < \delta \le \bar{\delta}.$$

Nevertheless, the results from [20] extend to the case $p = 1$, even if the regularization properties of the parameter choice do not hold. Indeed, these results were only based on bounding the excess penalty $\mathcal{R}(x) - \mathcal{R}(x^\dagger)$, and the data misfit $\|F(x) - F(x^\dagger)\|$, separately. In the present context, the SDP bounds the data misfit by $\tau\delta$, and in case that the chosen parameter does not tend to zero the excess penalty is bounded as $\mathcal{R}(x) - \mathcal{R}(x^\dagger) \le C\delta$, such that overall a rate of the order $\mathcal{O}(\delta)$ can be established in this case.

The lack of the regularization properties for the SDP parameter choice, specifically the violation of the exact penalization veto, can be established in the following situation. Recall, that $\mathcal{L}$ denotes the set of all $\mathcal{R}$-minimizing solutions of $F\big(x\big) = y^\dagger$ and the definition of $\mathcal{M}_{\alpha,y}$ in (2.3).

Proposition 11. *Let $p = 1$ and let Assumption 3 hold. If $x^\dagger \in \mathcal{L}$ satisfies a variational inequality*

$$\beta E(x, x^\dagger) \le \mathcal{R}(x) - \mathcal{R}(x^\dagger) + C\,\|F(x) - F(x^\dagger)\| \qquad \text{for all} \quad x \in \mathcal{M}, \tag{4.6}$$

where $0 < \beta \le 1$, $C > 0$, E is a non-negative error measure, and $\mathcal{M}_{\alpha,y^\dagger} \subseteq \mathcal{M}$ for $0 < \alpha \le \overline{\alpha}$, then the EP veto (Assumption 2) is violated at $y^\dagger = F(x^\dagger)$.

Proof. In this proof we extend some ideas presented in [9] regarding exact penalization to the situation of variational inequalities. For all $\alpha > 0$ and $x_\alpha \in \mathcal{M}_{\alpha,y^\dagger}$ we have $\frac{1}{\alpha}\|F(x_\alpha) - y^\dagger\| + \mathcal{R}(x_\alpha) - \mathcal{R}(x^\dagger) \le 0$. Adding $C\,\|F(x_\alpha) - y^\dagger\|$ on both sides of the inequality and using (4.6) we arrive at the estimate

$$\frac{1}{\alpha}\|F(x_\alpha)-y^\dagger\| \le \frac{1}{\alpha}\|F(x_\alpha)-y^\dagger\| + (\mathcal{R}(x_\alpha)-\mathcal{R}(x^\dagger)+C\,\|F(x_\alpha)-y^\dagger\|)$$
$$\le C\,\|F(x_\alpha)-y^\dagger\|.$$

For $0<\alpha<\min(1/C,\overline{\alpha})$ this gives $0\le\left(\frac{1}{\alpha}-C\right)\|F(x_\alpha)-y^\dagger\|\le 0$, and consequently, that $F(x_\alpha)=y^\dagger$ for such α. Therefore we see that $x_\alpha\in\mathcal{L}\cap\bigcup_{\alpha>0}\mathcal{M}_{\alpha,y^\delta}\neq\emptyset$. However, according to Proposition 5 we have that $x_\alpha\notin X_{\min}$ whenever Assumption 3 holds. This contradicts Assumption 2 and the proof is complete. □

Note that the variational inequality (4.6) for the Bregman distance (4.2) with respect to a convex penalty $\mathcal{R}$ as error measure and under Gâteaux differentiability of F is equivalent to a *benchmark source condition* written in Banach spaces as

$$\xi=F'(x^\dagger)^*w, \qquad \xi\in\partial\mathcal{R}(x^\dagger),\ w\in Y^*$$

(cf. [25, 26]). The flavor of exact penalization expressed by Proposition 11 was presented in [19]. Another example of such situation is discussed in [11].

5. Proofs of the auxiliary propositions 1–6

Proof of Proposition 1. Let $x^\dagger\in\mathcal{D}$ and $y^\dagger=F(x^\dagger)$ be such that $x^\dagger\in\mathcal{M}_{\alpha,y^\dagger}$ for some $\alpha>0$. Due to the convexity of $\mathcal{D}$ we then have for every $x_{\min}\in X_{\min}$ and $0<t<1$ that $(1-t)x^\dagger+tx_{\min}\in\mathcal{D}$ and

$$\alpha\mathcal{R}(x^\dagger)=J_{\alpha,y^\dagger}(x^\dagger)\le J_{\alpha,y^\dagger}\big((1-t)x^\dagger+tx_{\min}\big)$$
$$\le \big\|F\big((1-t)x^\dagger+tx_{\min}\big)-y^\dagger\big\|^p+\alpha\mathcal{R}((1-t)x^\dagger+tx_{\min}).$$

The convexity of $\mathcal{R}$ yields that

$$\mathcal{R}((1-t)x^\dagger+tx_{\min})\le(1-t)\mathcal{R}(x^\dagger)+t\,\mathcal{R}(x_{\min})=(1-t)\mathcal{R}(x^\dagger)+t\mathcal{R}_{\min}.$$

Therefore, since $x^\dagger\in\mathcal{M}_{\alpha,y^\dagger}$, we get

$$\alpha\,t\mathcal{R}(x^\dagger)\le\big\|F\big(x^\dagger+t(x_{\min}-x^\dagger)\big)-F\big(x^\dagger\big)\big\|^p+\alpha t\mathcal{R}_{\min}$$

and, after dividing by αt and letting $t\to+0$,

$$\mathcal{R}(x^\dagger)\le\frac{1}{\alpha}\liminf_{t\to+0}\left\{\frac{1}{t}\,\big\|F\big(x^\dagger+t(x_{\min}-x^\dagger)\big)-F\big(x^\dagger\big)\big\|^p\right\}+\mathcal{R}_{\min}.$$

Thus $x^\dagger\in X_{\min}$ follows since

$$\text{(5.1)}\quad \lim_{t\to+0}\left\{\frac{1}{t}\,\big\|F\big(x^\dagger+t(x_{\min}-x^\dagger)\big)-F\big(x^\dagger\big)\big\|^p\right\}$$
$$=\left(\lim_{t\to+0}t^{p-1}\right)\,\big\|F'(x^\dagger)(x_{\min}-x^\dagger)\big\|^p=0$$

and the proof is complete. □

Proof of Proposition 2. Again let $x^\dagger \in \mathcal{D}$ and $y^\dagger = F(x^\dagger)$ be such that $x^\dagger \in \mathcal{M}_{\alpha,y^\dagger}$ for some $\alpha > 0$. By $0 \in \mathcal{D}$ and $\mathcal{R}(0) = 0$ we obtain here that $0 \in X_{\min}$ and owing to (2.9) we have that $tx^\dagger \in \mathcal{D}$ for all $0 < t < 1$. Then we derive from condition (2.9) that

$$\begin{aligned} \alpha\mathcal{R}(x^\dagger) = J_{\alpha,y^\dagger}\big(x^\dagger\big) &\le J_{\alpha,y^\dagger}\big((1-t)x^\dagger\big) \\ &= \big\|F\big((1-t)x^\dagger\big) - y^\dagger\big\|^p + \alpha\mathcal{R}((1-t)x^\dagger) \\ &\le \big\|F\big((1-t)x^\dagger\big) - y^\dagger\big\|^p + \alpha\,\theta(1-t)\,\mathcal{R}(x^\dagger). \end{aligned}$$

Using similar arguments as in the proof of Proposition 1 and also (2.8), we obtain

$$\begin{aligned} \alpha\mathcal{R}(x^\dagger) &\le \liminf_{t\to+0}\left\{\frac{1}{1-\theta(1-t)}\ \big\|F\big((1-t)x^\dagger\big) - F\big(x^\dagger\big)\big\|^p\right\} \\ &\le C_\theta \cdot \lim_{t\to+0}\left\{\frac{1}{t}\ \big\|F\big((1-t)x^\dagger)\big) - F\big(x^\dagger\big)\big\|^p\right\} = 0, \end{aligned}$$

where the last equality follows from (5.1) with the choice $x_{\min} = 0$. Thus $x^\dagger \in X_{\min}$ which completes the proof. □

Proof of Proposition 3. Let, for the element $y^\dagger \in F(\mathcal{D})$ under consideration, $x^\dagger \in \mathcal{L} \setminus X_{\min}$ be a minimizer of $J_{\alpha,y^\dagger}$ for some $\alpha > 0$. Because of the Gâteaux differentiability of F and $\mathcal{R}$ on $\mathcal{L} \setminus X_{\min}$ we have $x^\dagger \in \operatorname{int}(\mathcal{D})$ and the Tikhonov functional is also differentiable with

$$\nabla J_{\alpha,y^\dagger}\big(x^\dagger\big) = \nabla\{\|F(\cdot) - y^\dagger\|^p\}(x^\dagger) + \alpha\nabla\mathcal{R}(x^\dagger) = 0.$$

Using for $t \to +0$ limit considerations as in the proof of Proposition 1 we obtain here

$$\nabla\{\|F(\cdot) - y^\dagger\|^p\}(x) = 0 \qquad \text{for all} \qquad x \in \mathcal{L} \setminus X_{\min}.$$

Indeed, the directional derivative of $\|F(\cdot) - y^\dagger\|^p$ at $x \in \mathcal{L} \setminus X_{\min}$ in any given direction $h \in X$ vanishes as

$$\begin{aligned} &\lim_{t\to 0}\frac{\|F(x+th) - y^\dagger\|^p - \|F(x) - y^\dagger\|^p}{t} \\ &\quad = \lim_{t\to 0}\|F(x+th) - y^\dagger\|^{p-1}\left\|\frac{F(x+th) - F(x)}{t}\right\| \\ &\quad = \lim_{t\to 0}\|F(x+th) - y^\dagger\|^{p-1}\|F'(x)(h)\| = 0. \end{aligned}$$

Consequently $\nabla\mathcal{R}(x^\dagger) = 0$ which, however, violates the premise that $\nabla\mathcal{R}(x) \ne 0$ is valid for all $x \in \mathcal{D} \setminus X_{\min}$ and proves the proposition. □

Proof of Proposition 4. In order to prove this proposition, we provide a more detailed picture of monotonicity and asymptotics beyond Proposition 4. The proofs of the asymptotics also rely on the monotonicity of these functionals.

Lemma 3 (see e.g., [1, Lemma 4.7])**.** *If $y^\delta \in Y$ is fixed and $0 < \alpha < \beta$, then*

$$\begin{aligned} \|F(x_\alpha^\delta) - y^\delta\| &\le \|F(x_\beta^\delta) - y^\delta\|, \\ \mathcal{R}(x_\alpha^\delta) &\ge \mathcal{R}(x_\beta^\delta), \\ J_{\alpha,y^\delta}(x_\alpha^\delta) &\le J_{\beta,y^\delta}(x_\beta^\delta), \end{aligned}$$

holds for all $x_\alpha^\delta \in \mathcal{M}_{\alpha,y^\delta}$ and $x_\beta^\delta \in \mathcal{M}_{\beta,y^\delta}$.

The following lemma extends the assertions of Proposition 4

Lemma 4. *Let $y^\delta \in Y$ be fixed and for $\alpha > 0$ let $x_\alpha^\delta \in \mathcal{M}_{\alpha,y^\delta}$. Then,*

$$\begin{aligned} \lim_{\alpha\to+0} \alpha\mathcal{R}(x_\alpha^\delta) &= 0, \\ \lim_{\alpha\to\infty} \mathcal{R}(x_\alpha^\delta) &= \mathcal{R}_{\min}, \\ \lim_{\alpha\to+0} J_{\alpha,y^\delta}(x_\alpha^\delta) &= \inf_{x\in\mathcal{D}} \|F(x) - y^\delta\|^q \\ \lim_{\alpha\to\infty} J_{\alpha,y^\delta}(x_\alpha^\delta) &= \begin{cases} d(y^\delta, Y_{\min})^q & \text{if } \mathcal{R}_{\min} = 0 \\ +\infty & \text{otherwise,} \end{cases} \\ \lim_{\alpha\to+0} \|F(x_\alpha^\delta) - y^\delta\| &= \inf_{x\in X_{\min}} \|F(x) - y^\delta\|, \\ \lim_{\alpha\to\infty} \|F(x_\alpha^\delta) - y^\delta\| &\ge d(y^\delta, Y_{\min}) \end{aligned}$$

and, if $\mathcal{R}_{\min} = 0$, then equality also holds in the last line.

Proof. For $\alpha \to +0$ the asymptotic relations as are well known, see e.g. [1, Lemma 4.15]. For $\alpha \to \infty$ we argue as follows. Since

$$J_{\alpha,y^\delta}(x_\alpha^\delta) \le \|F(x_{\min}) - y^\delta\|^q + \alpha\mathcal{R}_{\min} = d(y^\delta, Y_{\min})^q + \alpha\mathcal{R}_{\min},$$

we obtain that

$$\begin{aligned} \mathcal{R}_{\min} &\le \liminf_{\alpha\to\infty} \mathcal{R}(x_\alpha^\delta) \le \limsup_{\alpha\to\infty} \mathcal{R}(x_\alpha^\delta) \\ &\le \lim_{\alpha\to\infty} \left\{ \frac{1}{\alpha} J_{\alpha,y^\delta}(x_\alpha^\delta) \right\} \le \lim_{\alpha\to\infty} \left\{ \frac{1}{\alpha} d(y^\delta, Y_{\min})^q + \mathcal{R}_{\min} \right\} = \mathcal{R}_{\min}, \end{aligned}$$

and we have that $\mathcal{R}(x_\alpha^\delta) \to \mathcal{R}_{\min}$ as $\alpha \to \infty$. This also shows that $x_\alpha^\delta \in \mathcal{M}^{\mathcal{R}}(c)$ for some c large enough, and we may thus find a sequence $\alpha_k \to \infty$ and corresponding minimizers $x_k \in \mathcal{M}_{\alpha_k,y^\delta}$ such that $x_k \to \bar{x} \in \mathcal{D}$ with respect to τ_X. Then, the lower semicontinuity of $\mathcal{R}$ yields $\mathcal{R}(\bar{x}) \le \liminf_{k\to\infty} \mathcal{R}(x_k) = \mathcal{R}_{\min}$, such that $\bar{x} \in X_{\min}$. Therefore,

$$\begin{aligned} d(y^\delta, Y_{\min}) \le \|F(\bar{x}) - y^\delta\| &\le \liminf_{k\to\infty} \|F(x_k) - y^\delta\| \\ &= \lim_{\alpha\to\infty} \|F(x_\alpha^\delta) - y^\delta\|, \end{aligned} \tag{5.2}$$

where the last identity holds due to the monotonicity asserted in the preceding Lemma.

Finally, if $\mathcal{R}_{\min} = 0$, then for all $x_{\min} \in X_{\min}$

$$0 \le J_{\alpha,y^\delta}\big(x_\alpha^\delta\big) \le \big\|F\big(x_{\min}\big) - y^\delta\big\|^q,$$

which together with (5.2) yields

$$d(y^\delta, Y_{\min})^q \le \lim_{\alpha\to\infty} \big\|F\big(x_\alpha^\delta\big) - y^\delta\big\|^q \le \lim_{\alpha\to\infty} J_{\alpha,y^\delta}\big(x_\alpha^\delta\big) \le d(y^\delta, Y_{\min})^q.$$

If, on the other hand, $\mathcal{R}_{\min} > 0$, then $J_{\alpha,y^\delta}\big(x_\alpha^\delta\big) \to \infty$ follows from

$$J_{\alpha,y^\delta}\big(x_\alpha^\delta\big) \ge \alpha\mathcal{R}(x_\alpha^\delta) \ge \alpha\mathcal{R}_{\min} \to \infty \qquad \text{as } \alpha \to \infty,$$

which completes the proof. □

Proof of Proposition 5. The statement regarding Assumption 2 is trivially true, and we shall show the validity of Assumption 3 under (2.14). By virtue of Fact 1 (c), we have that

$$\kappa := d\left(y^\dagger, Y_{\min}\right) = \min_{x\in X_{\min}} \big\|F(x) - y^\dagger\big\| > 0.$$

Thus, for any $\tau > 1$ we can choose

$$0 < \bar{\delta} < \frac{\kappa}{\tau + 1}.$$

Then, for all $0 < \delta \le \bar{\delta}$, data y^δ satisfying $\big\|y^\dagger - y^\delta\big\| \le \delta$, and $x_{\min} \in X_{\min}$, we obtain

$$\big\|F\big(x_{\min}\big) - y^\delta\big\| \ge \big\|F\big(x_{\min}\big) - y^\dagger\big\| - \big\|y^\dagger - y^\delta\big\| \ge \kappa - \delta > \tau\delta,$$

which is (2.13).

For the last assertion we notice that, if $y^\dagger \in Y_{\min}$, then

$$d(y^\delta, Y_{\min}) \le \big\|y^\dagger - y^\delta\big\| \le \delta,$$

and (2.13) cannot hold no matter how $\tau > 1$ is chosen. □

Proof of Proposition 6. This assertion is a consequence of

$$\begin{aligned}\big\|F\big(x_\alpha^\delta\big) - y^\delta\big\|^p + \alpha\mathcal{R}(x_\alpha^\delta) &\le \big\|F\big(x_{\min}\big) - y^\delta\big\|^p + \alpha\mathcal{R}_{\min} \\ &\le (\tau\delta)^p + \alpha\mathcal{R}_{\min},\end{aligned}$$

which implies

$$\big\|F\big(x_\alpha^\delta\big) - y^\delta\big\|^p \le \big\|F\big(x_{\min}\big) - y^\delta\big\|^p + \alpha(\mathcal{R}_{\min} - \mathcal{R}(x_\alpha^\delta)) \le (\tau\delta)^p.$$

□

Conclusion

We have investigated the regularization property (1.4) for a widely applicable sequential version of the discrepancy principle for Tikhonov's regularization method applied to nonlinear ill-posed problems in Banach spaces, including partly the case of certain non-convex penalties. Ensuring the regularization property is of great importance due to its inherent consequences for the convergence of regularized solutions. We have even shown its impact on the variational inequality approach for obtaining convergence rates, where it allows us to restrict all considerations to sublevel sets of the penalty term.

In the course of showing (1.4), we were able to demonstrate the prominent role of the exact penalization veto formulated as Assumption 2. This veto is, in combination with the required data compatibility, the crucial link for obtaining (1.4) when using the sequential discrepancy principle. In Propositions 1–3 we could verify for three different situations that the exact penalization veto is satisfied, but we have also formulated counterexamples, e.g. in Proposition 11, where the veto is violated.

Acknowledgement

The first two authors greatly appreciate financial support by the German Science Foundation (DFG) under grant HO 1454/8-1.

References

[1] Anzengruber S W 2012 *The discrepancy principle for Tikhonov regularization in Banach spaces: Regularization properties and rates of convergence* (Saarbrücken: Südwestdeutscher Verlag für Hochschulschriften)

[2] Anzengruber S W and Ramlau R 2010 Morozov's discrepancy principle for Tikhonov-type functionals with nonlinear operators *Inverse Problems* 26(2) 025001

[3] Anzengruber S W and Ramlau R 2011 Convergence rates for Morozov's discrepancy principle using variational inequalities *Inverse Problems* 27(10) 105007

[4] Bonesky T 2009 Morozov's discrepancy principle and Tikhonov-type functionals *Inverse Problems* 25(1) 015015

[5] Boţ R I and Hofmann B 2013 The impact of a curious type of smoothness conditions on convergence rates in ℓ^1-regularization *Eurasian Journal of Mathematical and Computer Applications* 1(1) 29–40

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

[7] Burger M, Flemming J and Hofmann B 2013 Convergence rates in ℓ^1-regularization if the sparsity assumption fails *Inverse Problems* 29(2) 025013

[8] Burger M and Osher S 2012 *A Guide to the TV Zoo* (Münster: University of Münster)

[9] Burger M and Osher S 2004 Convergence rates of convex variational regularization *Inverse Problems* 20(5) 1411–21

[10] Burger M, Resmerita E and He L 2007 Error estimation for Bregman iterations and inverse scale space methods in image restoration *Computing* 81(2-3) 109–135

[11] Clason C 2012 L^∞ fitting for inverse problems with uniform noise *Inverse Problems* 28(10) 104007

[12] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* vol. 375 of *Mathematics and its Application* (Dordrecht: Kluwer Academic Publishers)

[13] Flemming J 2012 *Generalized Tikhonov Regularization and Modern Convergence Rate Theory in Banach Spaces* (Aachen: Shaker Verlag)

[14] Grasmair M 2009 Well-posedness and convergence rates for sparse regularization with sublinear l^q penalty term *Inverse Probl. Imaging*, 3(3) 383–387

[15] Grasmair M 2010 Non-convex sparse regularisation *J. Math. Anal. Appl.* 365(1) 19–28

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

[17] Grasmair M 2012 An application of source inequalities for convergence rates of Tikhonov regularization with a non-differentiable operator *Submitted - preliminary version under arXiv:1209.2246v1*

[18] Grasmair M, Haltmeier M and Scherzer O 2008 Sparse regularization with ℓ^q penalty term *Inverse Problems* 24(5) 1–13

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

[20] Hofmann B and Mathé P 2012 Parameter choice under variational inequalities *Inverse Problems* 28(10) 104006

[21] Hofmann B and Yamamoto M 2010 On the interplay of source conditions and variational inequalities for nonlinear ill-posed problems *Applicable Analysis* 89(11) 1705–1727

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

[23] Offtermatt J 2012 *A Projection and Variational Regularization Method for Sparse Inverse Problems* (PhD thesis) (Stuttgart: University of Stuttgart, Dept. Math.)

[24] Ramlau R and and Zarzer C A 2012 On the minimization of a Tikhonov functional with a non-convex sparsity constraint *Electron. Trans. Numer. Anal.* 39 476–507

[25] Scherzer O, Grasmair M, Grossauer H, Haltmeier M and Lenzen F 2009 *Variational Methods in Imaging* (New York: Springer-Verlag)

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

[27] Tikhonov A N, Leonov A S and Yagola A G 1998 *Nonlinear Ill-posed Problems* (London: Chapman & Hall)

[28] Zarzer C A 2009 On Tikhonov regularization with non-convex sparsity constraints *Inverse Problems* 25(2) 025006

Department of Mathematics, Chemnitz University of Technology, 09107 Chemnitz, Germany
E-mail address: stephan.anzengruber@mathematik.tu-chemnitz.de

Department of Mathematics, Chemnitz University of Technology, 09107 Chemnitz, Germany
E-mail address: bernd.hofmann@mathematik.tu-chemnitz.de

Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstrasse 39, 10117 Berlin, Germany
E-mail address: peter.mathe@wias-berlin.de