An extension of the variational inequality approach for obtaining convergence rates in regularization of nonlinear ill-posed problems

Radu Ioan Boţ and Bernd Hofmann *

February 4, 2010

Dedicated to Charles W. Groetsch

Abstract

Convergence rates results for Tikhonov regularization of nonlinear ill-posed operator equations in abstract function spaces require the handling of both smoothness conditions imposed on the solution and structural conditions expressing the character of nonlinearity. Recently, the distinguished role of variational inequalities holding on some level sets was outlined for obtaining convergence rates results. When lower rates are expected such inequalities combine the smoothness properties of solution and forward operator in a sophisticated manner. In this paper, using a Banach space setting we are going to extend the variational inequality approach from Hölder rates to more general rates including the case of logarithmic convergence rates.

MSC2000 subject classification: 47J06, 65J20, 47A52, 49N45

Keywords: Ill-posed problems, Tikhonov regularization, convergence rates, variational inequalities, source conditions, distance function, structure of nonlinearity, generalized Young inequality.

1 Introduction

With the monograph [8] Charles Groetsch presented an extremely well-readable introduction to the theory of Tikhonov regularization of ill-posed operator equations in Hilbert spaces. For linear ill-posed problems in that book the ingredients and conditions for obtaining convergence rates, the role of source conditions and the phenomenon of saturation are outlined. The ill-posedness of a linear operator equation describing an inverse problem with 'smoothing' forward operator in Hilbert spaces corresponds with the fact

*Department of Mathematics, Chemnitz University of Technology, 09107 Chemnitz, Germany. Email: bot@mathematik.tu-chemnitz.de, hofmannb@mathematik.tu-chemnitz.de.

that the Moore-Penrose inverse of the forward operator is unbounded and only densely defined on the image space. In that sense, solving linear ill-posed problems based on noisy data can be considered as the application of that unbounded operator to such data elements. For further theoretic extensions we refer to the recent monograph [9]. In 1989 Engl, Kunisch, and Neubauer published a seminal paper [5] on convergence rates results for the Tikhonov regularization of nonlinear ill-posed problems in the Hilbert space setting (see also [4, Chapter 10]). After the turn of the millennium, motivated by specific applications, for example in imaging, there occurred numerous publications on the Banach space treatment of linear and nonlinear operator equations including convergence rates results (see, e.g., [1, 7, 20, 23, 24, 27]). Initiated by the paper [2] of Burger and Osher Bregman distances were systematically exploited for evaluating the regularization error. Because of a completely different methodology for obtaining convergence rates in (generalized) Tikhonov regularization one can distinguish between *low rate results* up to Bregman errors of order $\mathcal{O}(\delta)$ for the noise level $\delta > 0$ and *enhanced rate results* up to the Bregman saturation order $\mathcal{O}(\delta^{4/3})$. Recently, in the papers [13, 16], moreover in [6, 12], in the thesis [22] and in the monograph [25] by Scherzer et al. the distinguished role of *variational inequalities* for proving low rate convergence rates of Hölder type was worked out. This paper tries to extend the variational inequality approach to obtain more general Bregman rates of form $\mathcal{O}(\varphi(\delta))$ with concave index functions φ. This includes the case of logarithmic convergence rates (see the papers [17, 18] by Hohage and Kaltenbacher).

The paper is organized as follows: In Section 2 we present for a nonlinear ill-posed problem in Banach spaces a general setting of Tikhonov type variational regularization with convex stabilizing penalty functional and a misfit functional built by a strictly convex index function of the residual norm. A linear combination of both functionals with some positive regularization parameter as multiplier forms the Tikhonov functional. This functional is to be minimized for obtaining stable approximate solutions of the nonlinear ill-posed problem under consideration. The standing assumptions of the setting and assertions on weak convergence and level sets are also outlined in Section 2. The subsequent Section 3 discusses structural conditions on the nonlinearity of the problem and source conditions as well as approximate source conditions imposed on the solution. The first main result, yielding an extension of the variational inequality approach from convergence rates results of Hölder type to results for general convex index functions, is formulated and proven as Theorem 4.3 in Section 4. As an essential ingredient the proof applies the generalization (4.12) of Young's inequality. The second main result of the paper is given in the concluding Section 5 by the couple of Theorems 5.1 and 5.2, that provide sufficient conditions for obtaining the general variational inequalities required in Theorem 4.3. The canonical source condition for low rates in Banach spaces and distance functions for measuring its violation form the basis for that conditions.

2 Problem setting and assumptions

In this paper, ill-posed operator equations

$$F(x) = y \tag{2.1}$$

are under consideration, where the operators $F : \mathcal{D}(F) \subseteq X \to Y$ with domain $\mathcal{D}(F)$ are mapping between real Banach spaces X and Y, respectively. For some noise level $\delta \ge 0$ let y^δ denote noisy data of the exact right-hand side $y = y^0 \in F(\mathcal{D}(F))$ with

$$\|y^\delta - y\|_Y \le \delta\,. \tag{2.2}$$

Based on that data we consider for regularization parameters $\alpha > 0$ stable approximate solutions x_α^δ as minimizers of the (generalized) Tikhonov type functional

$$T_\alpha^\delta(x) := \psi(\|F(x) - y^\delta\|_Y) + \alpha\,\Omega(x) \tag{2.3}$$

with a *misfit function* $\psi : [0,\infty) \to [0,\infty)$ and a *penalty functional* $\Omega : \mathcal{D}(\Omega) \subseteq X \to [0,\infty)$. The set of admissible elements for the minimization of (2.3) is the intersection $\mathcal{D} := \mathcal{D}(F) \cap \mathcal{D}(\Omega)$ of the occurring domains.

Index functions play a central role in our considerations. Originally coming from the theory of variable Hilbert scales and expressing the function-valued index of such a scale element (see [10, 14]), we use this concept as follows:

Definition 2.1 *We call a real function $\eta : [0,\infty) \to [0,\infty)$ (and also its restriction to any segment $[0,\bar{c}]$ $(0 < \bar{c} < \infty)$) index function if it is continuous and strictly increasing with $\eta(0) = 0$.*

Note that for index functions η, η_1, η_2 also the inverse function η^{-1} and the antiderivative $\Theta(s) := \int\limits_0^s \eta(t)dt$ are index functions, furthermore also all positive linear combinations $\lambda_1\eta_1 + \lambda_2\eta_2$ $(\lambda_1, \lambda_2 \ge 0,\ \lambda_1^2 + \lambda_2^2 > 0)$ and compositions $\eta_1 \circ \eta_2$.

Throughout this paper we make the following assumptions:

Assumption 2.2

1. *X and Y are Banach spaces with topological duals X^* and Y^*, respectively, where $\|\cdot\|_X$, $\|\cdot\|_Y$ and $\langle\cdot,\cdot\rangle_{X^*,X}$ and $\langle\cdot,\cdot\rangle_{Y^*,Y}$ denote the associated norms and dual pairings. In X and Y we consider in addition to the strong convergence $\to$ based on norms the weak convergence $\rightharpoonup$ based on the weak topology.*

2. *$F : \mathcal{D}(F) \subseteq X \to Y$ is weakly-weakly sequentially continuous and $\mathcal{D}(F)$ is weakly sequentially closed, i.e.,*

   $$x_k \rightharpoonup x \quad \text{in} \quad X \ \text{with} \ x_k \in \mathcal{D}(F) \quad \Longrightarrow \quad x \in \mathcal{D}(F) \quad \text{and} \quad F(x_k) \rightharpoonup F(x) \quad \text{in} \quad Y.$$

3. *The set $\mathcal{D}(\Omega)$ is convex and the functional Ω is convex and weakly sequentially lower semi-continuous.*

4. *The domain $\mathcal{D} := \mathcal{D}(F) \cap \mathcal{D}(\Omega)$ is non-empty.*

5. *For every $\alpha > 0$, $c \ge 0$, and for the exact right-hand side $y = y^0$ of (2.1), the sets*

   $$\mathcal{M}_\alpha(c) := \left\{x \in \mathcal{D} : T_\alpha^0(x) \le c\right\} \tag{2.4}$$

   are weakly sequentially pre-compact in the following sense: every sequence $\{x_k\}_{k=1}^\infty$ in $\mathcal{M}_\alpha(c)$ has a subsequence, which is weakly convergent in X to some element in X.

6. $\psi : [0,\infty) \to [0,\infty)$ *is an index function with the property that there exist numbers* $\bar{a} = \bar{a}(\psi) > 0,\ \bar{b} = \bar{b}(\psi) > 0$ *fulfilling*

$$\psi(u+v) \le \bar{a}\,\psi(u) + \bar{b}\,\psi(v) \qquad \text{for all} \quad u,v \in [0,\infty). \tag{2.5}$$

One should notice that item 6 in Assumption 2.2 is fulfilled in case $\psi : [0,\infty) \to [0,\infty)$ is a *p-homogeneous* (with $p > 0$) and *convex* index function. We recall that ψ is said to be p-homogeneous (with $p > 0$) whenever we have $\psi(tx) = t^p\psi(x)$ for all $x \in [0,\infty)$ and all $t \ge 0$.

Under the stated assumptions existence and stability of regularized solutions x_α^δ can be shown in the lines of the proof of [25, Theorems 3.22 and 3.23] (see also [13, Section 3]).

For the convex functional Ω with subdifferential $\partial\Omega$ regularization errors in a Banach space setting are frequently measured by means of Bregman distances

$$D_\xi(\tilde{x},x) := \Omega(\tilde{x}) - \Omega(x) - \langle \xi, \tilde{x} - x\rangle_{X^*,X}\ , \quad \tilde{x} \in \mathcal{D}(\Omega) \subseteq X\,,$$

at $x \in \mathcal{D}(\Omega) \subseteq X$ and $\xi \in \partial\Omega(x) \subseteq X^*$. The set

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

is called Bregman domain. An element $x^\dagger \in \mathcal{D}$ is called an Ω-*minimizing solution* to (2.1) if

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

Such Ω-minimizing solutions exist under Assumption 2.2 if (2.1) has a solution $x^\dagger$ in $\mathcal{D}$. This can be shown in analogy to the proof of [25, Lemma 3.2].

We close this section by proving that the regularized solutions associated with data possessing a sufficiently small noise level δ belong to a level set like the one in (2.4), provided that the regularization parameters $\alpha = \alpha(\delta)$ are chosen such that weak convergence to Ω-minimizing solutions $x^\dagger$ is enforced.

Proposition 2.3 *Consider an a priori choice* $\alpha = \alpha(\delta) > 0,\ 0 < \delta < \infty$, *for the regularization parameter in* (2.3) *depending on the noise level* δ *such that*

$$\alpha(\delta) \to 0 \ \text{and}\ \frac{\psi(\delta)}{\alpha(\delta)} \to 0. \tag{2.6}$$

Provided that (2.1) *has a solution* $x^\dagger$ *in* $\mathcal{D}$ *then under Assumption 2.2 every sequence* $\{x_n\}_{n=1}^\infty := \{x_{\alpha(\delta_n)}\}^{\delta_n}\}_{n=1}^\infty$ *of regularized solutions corresponding to a sequence* $\{y^{\delta_n}\}_{n=1}^\infty$ *of data with* $\lim_{n\to\infty}\delta_n = 0$ *has a subsequence* $\{x_{n_k}\}_{k=1}^\infty$, *which is weakly convergent in* X, *i.e.* $x_{n_k} \rightharpoonup x^\dagger$ *and its limit* $x^\dagger$ *is an* Ω-*minimizing solution of* (2.1) *with* $\Omega(x^\dagger) = \lim_{k\to\infty}\Omega(x_{n_k})$.

For given $\alpha_{\max} > 0$, *let* $x^\dagger$ *denote an* Ω-*minimizing solution of* (2.1). *If we set*

$$\rho = \alpha_{\max}(1 + \Omega(x^\dagger)), \tag{2.7}$$

then we have $x^\dagger \in \mathcal{M}_{\alpha_{\max}}(\rho)$ *and there exists some* $\delta_{\max} > 0$ *such that*

$$x_{\alpha(\delta)}^\delta \in \mathcal{M}_{\alpha_{\max}}(\rho) \ \text{for all}\ 0 < \delta \le \delta_{\max}. \tag{2.8}$$

Proof: The first part of the result can be proved in the same manner as [25, Theorem 3.26]. Here the properties of the index function ψ play a determinant role.

We come now to the second part of the above statement and consider an $\alpha_{\max} > 0$. Because of (2.6) there exists some $\delta_{\max} > 0$ such that $\alpha(\delta) \le \alpha_{\max}$ and $\frac{\psi(\delta)}{\alpha(\delta)} \le \min\{\frac{1}{2}, \frac{1}{2\bar{b}}\}$ for all $0 < \delta \le \delta_{\max}$. In the following we write for simplicity α instead of $\alpha(\delta)$.

For all $0 < \delta \le \delta_{\max}$, by (2.5), we have

$$
\begin{aligned}
T^0_{\alpha_{\max}}(x^\delta_\alpha) &= \psi(\|F(x^\delta_\alpha) - y\|_Y) + \alpha_{\max}\Omega(x^\delta_\alpha) \le \bar{a}\psi(\|F(x^\delta_\alpha) - y^\delta\|_Y) + \bar{b}\psi(\delta) + \alpha_{\max}\Omega(x^\delta_\alpha)\\
&= \bar{a}[\psi(\|F(x^\delta_\alpha) - y^\delta\|_Y) + \alpha\Omega(x^\delta_\alpha)] + \bar{b}\psi(\delta) + (\alpha_{\max} - \bar{a}\alpha)\Omega(x^\delta_\alpha)\\
&\le \bar{a}T^\delta_\alpha(x^\dagger) + \bar{b}\psi(\delta) + (\alpha_{\max} - \bar{a}\alpha)\Omega(x^\delta_\alpha) \le (\bar{a} + \bar{b})\psi(\delta) + \bar{a}\alpha\Omega(x^\dagger) + (\alpha_{\max} - \bar{a}\alpha)\Omega(x^\delta_\alpha).
\end{aligned}
$$

On the other hand, from $T^\delta_\alpha(x^\delta_\alpha) \le T^\delta_\alpha(x^\dagger)$ it yields $\Omega(x^\delta_\alpha) \le \frac{\psi(\delta)}{\alpha} + \Omega(x^\dagger)$. Consequently,

$$
\begin{aligned}
T^0_{\alpha_{\max}}(x^\delta_\alpha) &\le (\bar{a} + \bar{b})\psi(\delta) + \bar{a}\alpha\Omega(x^\dagger) + \left(\frac{\alpha_{\max}}{\alpha} - \bar{a}\right)\psi(\delta) + (\alpha_{\max} - \bar{a}\alpha)\,\Omega(x^\dagger)\\
&= \bar{b}\psi(\delta) + \frac{\alpha_{\max}}{\alpha}\psi(\delta) + \alpha_{\max}\Omega(x^\dagger) \le \alpha_{\max}(1 + \Omega(x^\dagger)) = \rho.
\end{aligned}
$$

□

3 Source conditions and structural conditions of nonlinearity for the Banach space setting

There are two ingredients influencing the convergence rates for Tikhonov regularized solutions in the case of nonlinear ill-posed problems. On the one hand, the solution smoothness, if possible expressed by source conditions for $x^\dagger$, plays an important role. On the other hand, the structure of nonlinearity of F in a neighborhood of $x^\dagger$ must be in line with the solution smoothness in order to obtain a certain rate. In this context, we are going to restrict the situation a little bit more as follows:

Assumption 3.1

1. *$F, \Omega, \mathcal{D}, X$ and Y satisfy the Assumption 2.2.*
2. *Let $x^\dagger \in \mathcal{D}$ be an Ω-minimizing solution of (2.1).*
3. *The operator F is Gâteaux differentiable in $x^\dagger$ with the Gâteaux derivative $F'(x^\dagger) \in \mathcal{L}(X, Y)$ ($\mathcal{L}(X, Y)$ denotes the space of bounded linear operators from X to Y).*
4. *The functional Ω is Gâteaux differentiable in $x^\dagger$ with the Gâteaux derivative $\xi = \Omega'(x^\dagger) \in X^*$, i.e., $x^\dagger \in \mathcal{D}_B(\Omega)$ and the subdifferential $\partial\Omega(x^\dagger) = \{\xi\}$ is a singleton.*

In the case of Hilbert spaces X and Y by spectral theory one can consider bounded linear operators $\eta(F'(x^\dagger)^* F'(x^\dagger)) \in \mathcal{L}(X,X)$ for any index function η based on the fact that with the Hilbert space adjoint $F'(x^\dagger)^* \in \mathcal{L}(Y,X)$ of $F'(x^\dagger) \in \mathcal{L}(X,Y)$ the operators $F'(x^\dagger)^* F'(x^\dagger) \in \mathcal{L}(X,X)$ are non-negative and self-adjoint and this property carries over to the operators $\eta(F'(x^\dagger)^* F'(x^\dagger))$. For Banach spaces X and Y, however, only the Banach space adjoint $F'(x^\dagger)^* \in \mathcal{L}(Y^*,X^*)$ of $F'(x^\dagger)$ is available, but $F'(x^\dagger)^* F'(x^\dagger)$ and hence $\eta(F'(x^\dagger)^* F'(x^\dagger))$ are not well-defined. In contrast to the Hilbert space setting, where generalized source conditions

$$\xi = \eta(F'(x^\dagger)^* F'(x^\dagger))v, \quad v \in X\,, \tag{3.1}$$

can be exploited for arbitrary index functions η, in our setting only the source condition

$$\xi = F'(x^\dagger)^*\, w, \quad w \in Y^*\,, \tag{3.2}$$

expressing a medium smoothness of ξ, has a canonical character. We will consider this as an upper benchmark source condition, here accepting that only low and medium convergence rates for the regularized solutions are under consideration. For expressing higher solution smoothness with respect to the stabilizing functional Ω duality mappings can be helpful admitting enhanced convergence rates. For such kind of results we refer for example to the papers [11, 20, 21], meanwhile noticing that the higher source conditions used there seem to be a little bit artificial. Searching for low rate results in Banach spaces, with solution smoothness limited by (3.2), our main drawback is the non-existence of generalized source conditions (3.1) with concave index functions η such that $\sqrt{t} = \mathcal{O}(\eta(t))$ as $t \to 0$. This class of index functions includes for $0 < \nu \le 1/2$ the monomials

$$\eta(t) \;=\; t^\nu \quad (t \ge 0) \tag{3.3}$$

and for all $\mu > 0$ the family of logarithmic functions

$$\eta(t) \;=\; \begin{cases} 0 & (t=0) \\ [\log(1/t)]^{-\mu} & (0 < t \le e^{-\mu-1}) \end{cases}. \tag{3.4}$$

Since SCHOCK's paper [26] we know that convergence rates for regularized solutions can be arbitrarily low. This corresponds with arbitrarily weak solution smoothness. For example the very low multiple logarithmic rates for associated generalized source conditions with index function $\eta(t) = \log\log ... \log(1/t)$ really occur in applications of the Hilbert space theory.

One way of compensating the Banach space drawback of missing generalized source conditions consists in applying the *method of approximate source conditions* (see [3, 12]) whenever ξ fails to satisfy the benchmark source condition (3.2) for every $w \in Y$. This method is based on the utilization of the obviously non-increasing distance function $d : [0,\infty) \to [0,\infty)$ defined as

$$d(R) := \inf\{\|\xi - F'(x^\dagger)^*\, w\|_{X^*} :\; w \in Y^*,\; \|w\|_{Y^*} \le R\} \qquad (R \ge 0)\,. \tag{3.5}$$

We notice that this function is continuous, because it is convex. The latter is a consequence of the convexity of the function

$$\Xi(R,w) = \begin{cases} \|\xi - F'(x^\dagger)^* w\|_{X^*}, & \text{if } \|w\|_{Y^*} \le R, \\ \infty, & \text{otherwise,} \end{cases}$$

which implies the convexity of its corresponding infimal value function $d(R) = \inf_{w\in Y^*} \Xi(R,w)$.

The desired limit condition

$$\lim_{R\to\infty} d(R) \,=\, 0 \tag{3.6}$$

is fulfilled if and only if $\xi \in \overline{\mathcal{R}(F'(x^\dagger)^*)}^{\|\cdot\|_{X^*}}$. By a separation theorem one can prove that the latter is guaranteed provided $F'(x^\dagger)^{**}$ is injective. For $A \in \mathcal{L}(X,Y)$ we denote by $A^{**} \in \mathcal{L}(X^{**},Y^{**})$, defined by $\langle A^{**}x^{**}, y^*\rangle_{Y^{**},Y^*} = \langle x^{**}, A^*y^*\rangle_{X^{**},X^*}$ for $x^{**} \in X^{**}$ and $y^* \in Y^*$, its *bi-adjoint operator*. In reflexive Banach spaces X and Y this means assuming that $F'(x^\dagger)$ is injective.

The following lemma will be used in order to guarantee that the distance function defined in (3.5) *strictly* decreases to zero as $R \to \infty$.

Lemma 3.2 *Let X, Y be reflexive Banach spaces and $A \in \mathcal{L}(X,Y)$ an injective operator. For $\xi \in X^*$ we assume that $\xi \notin \mathcal{R}(A^*)$. Then the distance function $d : [0,+\infty) \to (0,+\infty)$, defined by*

$$d(R) = \inf\{\|\xi - A^*w\|_{X^*} :\; w \in Y^*,\; \|w\|_{Y^*} \le R\},$$

is strictly decreasing and satisfies the limit condition (3.6).

Proof: The limit condition (3.6) follows from $\xi \in \overline{\mathcal{R}(A^*)}^{\|\cdot\|_{X^*}}$, which is a consequence of the injectivity of A when X and Y are reflexive Banach spaces. We still have to show the strict decay of $d(R)$ with respect to R. First let us notice that for all $R \ge 0$ there exists $\bar w \in Y^*$, $\|\bar w\|_{Y^*} \le R$, such that $d(R) = \|\xi - A^*\bar w\|_{X^*}$. This is because of the fact that the dual norm function is weak* lower semicontinuous and the unit ball in Y^* is weak* compact (Theorem of Alaoglu-Bourbaki).

Let $R \ge 0$. Next we prove that if we have for $\bar w \in Y^*$, $\|\bar w\|_{Y^*} \le R$, the equation $d(R) = \|\xi - A^*\bar w\|_{X^*}$, then one necessarily must have $\|\bar w\|_{Y^*} = R$. In case $R = 0$, this fact is obvious. Suppose now that $R > 0$. Indeed, in this case $\bar w$ is an optimal solution of the convex optimization problems

$$\inf_{\|w\|_{Y^*} - R \le 0} \|\xi - A^*w\|_{X^*}.$$

As the Slater constraint qualification is fulfilled (for $w' = 0$ we have $\|w'\|_{Y^*} - R < 0$), there exists a Lagrange multiplier $\bar\lambda \ge 0$ such that (see, for instance, [29, Theorem 2.9.2])

$$\bar\lambda(\|\bar w\|_{Y^*} - R) = 0$$

and

$$0 \in \partial(\|\xi - A^*(\cdot)\|_{X^*} + \bar\lambda(\|\cdot\|_{Y^*} - R))(\bar w).$$

If we prove that $\bar\lambda > 0$, then the assertion follows. We assume the contrary. This means that

$$0 \in \partial(\|\xi - A^*(\cdot)\|_{X^*})(\bar w).$$

Next we evaluate the above subdifferential. Let $L : X^* \to \mathbb{R}$, $L(w) = \|\xi + w\|_{X^*}$. Since L is continuous, by [29, Theorem 2.8.2] we have that

$$\partial(\|\xi - A^*(\cdot)\|_{X^*})(\bar w) = \partial(L \circ (-A^*))(\bar w) = -A(\partial L(-A^*\bar w)).$$

As A is injective,

$$0 \in \partial L(-A^*\bar{w}) = \partial \|\cdot\|_{X^*}(\xi - A^*\bar{w}). \tag{3.7}$$

For the subdifferential of the norm we have the following expressions

$$\partial \|\cdot\|_{X^*}(v) = \{u \in X : \|u\|_X \leq 1\}, \text{ if } v = 0,$$

while

$$\partial \|\cdot\|_{X^*}(v) = \{u \in X : \|u\|_X = 1, \langle v, u \rangle_{X^*,X} = \|v\|_{X^*}\}, \text{ if } v \neq 0.$$

By (3.7) it follows that only the first situation is possible. Consequently, $\xi - A^*\bar{w} = 0$. But this is a contradiction to $\xi \notin \mathcal{R}(A^*)$. Thus $\bar{\lambda} > 0$ and, so, $\|\bar{w}\|_{Y^*} = R$.

Let us prove now that d is strictly decreasing. To this aim take $R_1, R_2 \in [0, +\infty)$ such that $0 \leq R_1 < R_2$. Then $d(R_1) \geq d(R_2)$. Assume that $d(R_1) = d(R_2)$. Then there exists $w_1, w_2 \in Y^*$, $\|w_1\|_{Y^*} = R_1$, $\|w_2\|_{Y^*} = R_2$, such that $d(R_1) = d(R_2) = \|\xi - A^*w_1\|_{X^*} = \|\xi - A^*w_2\|_{X^*}$. As $\|w_1\|_{Y^*} < R_2$, this leads to a contradiction to the above considerations. Consequently, $d(R_1) > d(R_2)$ and this concludes the proof. □

Let us mention that the speed of the decay of $d(R) \to 0$ as $R \to \infty$ under the condition $\xi \in \overline{\mathcal{R}(F'(x^\dagger)^*)}^{\|\cdot\|_{X^*}} \setminus \mathcal{R}(F'(x^\dagger)^*)$ expresses for the element ξ the degree of violation of (3.2) and thus it can be handled as a replacement information for the missing index function η from (3.1) in the Banach space setting.

As an adaption of the *local degree of nonlinearity* introduced for a Hilbert space setting in [15, Definition 1] to the Banach space situation with Bregman distance in [16, Definition 3.2] it has been suggested a definition, which attains here under Assumption 3.1 the form:

Definition 3.3 *Let $0 \leq c_1, c_2 \leq 1$ and $0 < c_1 + c_2 \leq 1$. We define F to be nonlinear of degree (c_1, c_2) at $x^\dagger$ for the Bregman distance $D_\xi(\cdot, x^\dagger)$ of Ω with $\xi = \Omega'(x^\dagger)$ if there is a constant $K > 0$ such that*

$$\left\|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\right\|_Y \leq K \left\|F(x) - F(x^\dagger)\right\|_Y^{c_1} D_\xi(x, x^\dagger)^{c_2} \tag{3.8}$$

for all $x \in \mathcal{M}_{\alpha_{max}}(\rho)$.

In [12] it was shown that the method of approximate source conditions yields convergence rates for Tikhonov regularized solutions x_α^δ minimizing (2.3) with misfit function $\psi(t) = t^p$ ($p > 1$) whenever we have $c_1 > 0$ in the degree of nonlinearity, even if ξ fails to satisfy the benchmark source condition (3.2). The corresponding rates depend on the distance function (3.5). If $c_1 > 0$ and the source condition (3.2) holds, then one even obtains Hölder convergence rates with Hölder exponents $\kappa = \frac{c_1}{1-c_2}$ (see [16]). If the nonlinearity of F at $x^\dagger$ is such that (3.8) can only be satisfied for $c_1 = 0$, then rate results are only known if, additionally, $c_2 = 1$ and (3.2) under the smallness condition $K\|w\|_{Y^*} < 1$ is valid (see [25]). As already mentioned in [18] for the Hilbert space setting, there are no rate assertions for $c_1 = 0$ and $c_2 = 1$, if ξ fails to satisfy the benchmark source condition (3.2). We even *conjecture* that convergence rate results, in principle, cannot be proven whenever the structure of nonlinearity of F at $x^\dagger$ is too *rough* and moreover ξ fails to satisfy the benchmark source condition (3.2). But Definition 3.3, with focus on powers with exponents c_1 and c_2, seems to be inappropriate for characterizing that roughness clear enough. Precisely, we will introduce and analyze in this paper a weaker condition for the structure of nonlinearity as

$$\left\|F'(x^\dagger)(x-x^\dagger)\right\|_Y \le C\,\sigma(\left\|F(x)-F(x^\dagger)\right\|_Y) \tag{3.9}$$

for all $x \in \mathcal{M}_{\alpha_{max}}(\rho)$ with some constant $C>0$ and some index function σ. We will exploit this new condition in Section 5 by using variational inequalities as main tool. In this context, let us note that the validity of (3.9) for $\sigma(t)=t^{c_1}$ and $0<c_1\le 1$ implies with the triangle inequality that we have

$$\left\|F(x)-F(x^\dagger)-F'(x^\dagger)(x-x^\dagger)\right\|_Y \le \left\|F'(x^\dagger)(x-x^\dagger)\right\|_Y + \left\|F(x)-F(x^\dagger)\right\|_Y$$
$$\le C\left\|F(x)-F(x^\dagger)\right\|_Y^{c_1} + \left\|F(x)-F(x^\dagger)\right\|_Y \le K\left\|F(x)-F(x^\dagger)\right\|_Y^{c_1}$$

on the associated level sets, fact that shows a degree $(c_1,0)$ of nonlinearity. As the condition (3.9) is typical for low order convergence rates, *concave* index functions σ are of main interest in this context. The concept of roughness mentioned in the conjecture above can be rendered more precisely now. We conjecture that the structure of nonlinearity of F at $x^\dagger$ is too rough if there is no index function σ such that (3.9) is valid. The considerations in Section 5 will assume that (3.9) is satisfied for some index function σ. This situation characterizes a boundary layer between the case $c_1>0$ well-discussed in the literature and the case of too much roughness for the structure of nonlinearity. We will prove in the sequel that for the boundary layer situations convergence rates results still do exist. It is future work to exhibit examples for which the structural condition (3.9) is just typical.

4 Variational inequalities and convergence rates

In recent publications (see [12, 13, 16, 25]) variational inequalities of the form

$$\left\langle \xi, x^\dagger - x\right\rangle_{X^*,X} \le \beta_1 D_\xi(x,x^\dagger) + \beta_2\left\|F(x)-F(x^\dagger)\right\|_Y^\kappa \qquad \text{for all} \quad x\in\mathcal{M}_{\alpha_{max}}(\rho) \tag{4.1}$$

with two multipliers $0\le\beta_1<1$, $\beta_2>0$ and an exponent $\kappa>0$ have been exploited for obtaining convergence rates in Tikhonov regularization in Banach spaces, where in the functional (2.3) to be minimized the strictly convex misfit function $\psi(t)=t^p$ $(p>1)$ was used. We repeat in our context the Proposition 3.3 from [16]:

Proposition 4.1 *Set $\psi(t):=t^p$ $(p>1)$ in (2.3) and assume that $F,\Omega,\mathcal{D},X,Y,x^\dagger$ and ξ satisfy the Assumption 3.1. If there exist constants $0\le\beta_1<1$, $\beta_2>0$, and $0<\kappa\le1$ such that the variational inequality (4.1) holds with ρ from (2.7), then we have the convergence rate*

$$D_\xi(x^\delta_{\alpha(\delta)},x^\dagger) = \mathcal{O}(\delta^\kappa) \quad \text{as} \quad \delta\to0 \tag{4.2}$$

for an a priori parameter choice $\alpha(\delta)\asymp\delta^{p-\kappa}$.

The proof of this proposition is based on the inequality $T_\alpha^\delta(x_\alpha^\delta) \le T_\alpha^\delta(x^\dagger)$ that holds for all regularized solutions x_α^δ and on the variant

$$a\,b \le a^{p_1} + \frac{b^{p_2}}{{p_1}^{p_2/p_1}p_2} \qquad (a,b \ge 0,\;\; p_1,p_2>1 \;\text{ with }\; \frac{1}{p_1}+\frac{1}{p_1}=1) \tag{4.3}$$

of Young's inequality. Note that due to Proposition 4.3 in [16] the case $\kappa > 1$ is not of interest, since (4.1) with $\kappa > 1$ implies the singular case $\xi = 0$.

For obtaining more general low order convergence rates we change (4.1) as follows: We assume that there holds a variational inequality

$$\langle \xi, x^\dagger - x\rangle_{X^*,X} \le \beta_1 D_\xi(x,x^\dagger) + \beta_2\,\varphi(\|F(x)-F(x^\dagger)\|_Y) \qquad \text{for all} \quad x \in \mathcal{M}_{\alpha_{max}}(\rho) \tag{4.4}$$

with two multipliers $0 \le \beta_1 < 1,\ \beta_2 > 0$ and an index function φ.

As outlined comprehensively in [16] for the monomial case $\varphi(t) = t^\kappa$ $(0 < \kappa \le 1)$ we have a variational inequality (4.4) if and only if the interplay between the solution smoothness expressed by conditions imposed on ξ and the structural condition for the nonlinearity of F at $x^\dagger$ is appropriate. The most clearly represented assertion in this context was formulated for $\varphi(t) = t$ (see [7], [22, Section 1.6] and [25, Section 3.2]), where the variational inequality (4.4) and the canonical source condition (3.2) are equivalent. On the other hand, in order to obtain (3.2) from (4.4), it is sufficient to guarantee the existence of a constant $\hat{C} > 0$ such that $\varphi(t) \le \hat{C}\,t$ for all t in a right neighborhood of zero. In this and in the following section, however, we focus on concave index functions φ in (4.4), which in general do not imply (3.2), and their correspondence with the functions ψ in the misfit functional of (2.3) and σ in the structural condition (3.9).

Assumption 4.2 *Regarding the functions ψ from (2.3) and φ from (4.4) we make the following assumptions:*

1. *ψ and φ are index functions which are twice differentiable on the interior of their domains.*
2. *ψ is strictly convex with $\lim_{s\to 0+} \psi'(s) = 0$ and φ is concave.*

Under Assumption 4.2 we can define another index function f as follows:

$$f(0) = 0 \text{ and } f(s) = \left[\frac{\psi'}{\varphi'} \circ \varphi^{-1}\right](s) \text{ when } s > 0. \tag{4.5}$$

Let us first show that f is well-defined, by proving that $\varphi'(s) > 0$ when $s > 0$. Indeed, suppose that there exists $\bar{s} > 0$ in the interior of the domain of φ such that $\varphi'(\bar{s}) = 0$. Take $t > \bar{s}$. By the concavity assumption one has

$$0 = \varphi'(\bar{s})(t-\bar{s}) \ge \varphi(t) - \varphi(\bar{s}),$$

which contradicts the fact that φ is strictly increasing.

By employing similar arguments, since ψ is convex, whenever $s > 0$ one has that $\psi'(s) > 0$ and so $f(s) > 0$.

In the following we prove that f is an index function. For showing the strict monotonicity of f take $0 < s_1 < s_2$. Then $\varphi^{-1}(s_1) < \varphi^{-1}(s_2)$. As ψ is strictly convex, ψ' is strictly increasing and so $0 < \psi'(\varphi^{-1}(s_1)) < \psi'(\varphi^{-1}(s_2))$. On the other hand, since φ is concave, φ' is non-increasing, consequently, $\varphi'(\varphi^{-1}(s_1)) \geq \varphi'(\varphi^{-1}(s_2)) > 0$. From here one has $f(s_1) < f(s_2)$.

As the continuity of f on $(0, +\infty)$ is automatically satisfied, one only needs to show that $\lim_{s \to 0+} f(s) = 0$. But, this limit condition is a consequence of $\lim_{s \to 0+} \psi'(s) = 0$ and of the monotonicity of $[\varphi^{-1}]'$, taking into account that we can write f in the form

$$f(s) = \psi'(\varphi^{-1}(s))[\varphi^{-1}]'(s) \quad (s > 0)\,.$$

Hence f is an index function and so is the antiderivative

$$H(s) := \int_0^s f(\tau) d\tau\,. \tag{4.6}$$

For $s > 0$ it follows from (4.5) that $\psi(s) = \int_0^{\varphi(s)} f(t)dt + C$. As $\psi(0) = 0$, this yields $C = 0$ and consequently

$$\psi(s) = H(\varphi(s)) = \int_0^{\varphi(s)} f(t)dt\,.$$

Now aspects of the interplay between ψ, φ, f and H can be written in different manner by the equations

$$\psi = H \circ \varphi, \quad H = \psi \circ \varphi^{-1}$$

and

$$f(s) = [\psi \circ \varphi^{-1}]'(s) \ \ (s > 0),$$

where the last equation yields (4.5) by differentiation and use of the chain rule. Further, let

$$G(s) := \int_0^s f^{-1}(\tau) d\tau \tag{4.7}$$

be the antiderivative of the inverse function to f.

Now we are ready to present the main convergence rate result of this paper:

Theorem 4.3 *Assume that $F, \Omega, \mathcal{D}, X, Y, x^\dagger$, ξ and ψ satisfy Assumption 3.1 and assume that ψ and φ satisfy Assumption 4.2 which ensures the existence of an index function f defined by (4.5). Let there exist constants $0 \leq \beta_1 < 1$, $\beta_2 > 0$, such that the variational inequality (4.4) holds with ρ from (2.7). Then we have the convergence rate of Tikhonov regularized solutions*

$$D_\xi(x^\delta_{\alpha(\delta)}, x^\dagger) \;=\; \mathcal{O}(\varphi(\delta)) \quad as \quad \delta \to 0 \tag{4.8}$$

for an a priori parameter choice

$$\alpha(\delta) = \frac{1}{\bar{a}\beta_2} f(\varphi(\delta))\,, \tag{4.9}$$

where the constant $\bar{a}$ is from (2.5).

Proof: Throughout this proof $\bar{a}$ and $\bar{b}$ are the constants introduced in formula (2.5). For all $\alpha > 0$ regularized solutions x_α^δ minimizing (2.3) have to satisfy the inequality $T_\alpha^\delta(x_\alpha^\delta) \le T_\alpha^\delta(x^\dagger)$. Using the definition of the Bregman distance this implies for the noise model (2.2) the estimate

$$\psi(\left\|F(x_\alpha^\delta) - y^\delta\right\|_Y) + \alpha D_\xi(x_\alpha^\delta, x^\dagger) \le \psi(\delta) + \alpha\left(\Omega(x^\dagger) - \Omega(x_\alpha^\delta) + D_\xi(x_\alpha^\delta, x^\dagger)\right) . \quad (4.10)$$

Moreover, from the variational inequality (4.4) we obtain that

$$\begin{aligned} &\Omega(x^\dagger) - \Omega(x_\alpha^\delta) + D_\xi(x_\alpha^\delta, x^\dagger) = -\left\langle \xi, x_\alpha^\delta - x^\dagger \right\rangle_{X^*,X} \\ &\le \beta_1\, D_\xi(x_\alpha^\delta, x^\dagger) \;+\; \beta_2\, \varphi(\left\|F(x_\alpha^\delta) - F(x^\dagger)\right\|_Y). \end{aligned}$$

Therefore from (4.10) it follows that

$$\begin{aligned} \psi(\|F(x_\alpha^\delta) - y^\delta\|_Y) + \alpha D_\xi(x_\alpha^\delta, x^\dagger) &\le \psi(\delta) + \alpha\beta_1 D_\xi(x_\alpha^\delta, x^\dagger) + \alpha\beta_2\varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y) \quad (4.11) \\ &= \psi(\delta) + \alpha\beta_1 D_\xi(x_\alpha^\delta, x^\dagger) + \frac{1}{\bar{a}}(\alpha\bar{a}\beta_2)\varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y) . \end{aligned}$$

Using the generalization of Young's inequality

$$a\, b \le \int_0^a f(t)dt + \int_0^b f^{-1}(\tau)d\tau \qquad (a, b \ge 0) \quad (4.12)$$

(see, for instance, [19]) with the index function f we obtain for sufficiently small $\alpha > 0$

$$\begin{aligned} (\alpha\bar{a}\beta_2)\varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y) &\le \int_0^{\varphi(\|F(x_\alpha^\delta)-F(x^\dagger)\|_Y)} f(t)dt + \int_0^{\alpha\bar{a}\beta_2} f^{-1}(\tau)d\tau \quad (4.13) \\ = H(\varphi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y) + G(\alpha\bar{a}\beta_2) &= \psi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y) + G(\alpha\bar{a}\beta_2). \end{aligned}$$

From (4.11) and (4.13) it follows that

$$\begin{aligned} &\psi(\|F(x_\alpha^\delta) - y^\delta\|_Y) + \alpha D_\xi(x_\alpha^\delta, x^\dagger) \\ &\le \psi(\delta) + \alpha\beta_1 D_\xi(x_\alpha^\delta, x^\dagger) + \frac{1}{\bar{a}}\,\psi(\|F(x_\alpha^\delta) - F(x^\dagger)\|_Y) + \frac{1}{\bar{a}}\,G(\alpha\bar{a}\beta_2) \\ &\le \psi(\delta) + \alpha\beta_1 D_\xi(x_\alpha^\delta, x^\dagger) + \psi(\|F(x_\alpha^\delta) - y^\delta\|_Y) + \frac{\bar{b}}{\bar{a}}\psi(\delta) + \frac{1}{\bar{a}}\,G(\alpha\bar{a}\beta_2). \end{aligned}$$

Consequently,

$$D_\xi(x_\alpha^\delta, x^\dagger) \le \frac{1}{(1-\beta_1)\bar{a}}\frac{(\bar{a}+\bar{b})\psi(\delta) + G(\alpha\bar{a}\beta_2)}{\alpha}, \quad (4.14)$$

for sufficiently small $\alpha > 0$.

Next we prove that we obtain for $\alpha(\delta) := \frac{1}{\bar{a}\beta_2} f(\varphi(\delta))$ the estimate

$$\frac{(\bar{a}+\bar{b})\psi(\delta) + G(\alpha(\delta)\bar{a}\beta_2)}{\alpha(\delta)} \le (\bar{a}+\bar{b}+1)\bar{a}\beta_2\varphi(\delta) \quad (4.15)$$

for sufficiently small $\delta > 0$. Indeed, (4.15) is equivalent to

$$(\bar{a} + \bar{b} + 1)\bar{a}\beta_2\varphi(\delta)\alpha(\delta) - (\bar{a} + \bar{b})\psi(\delta) - G(\alpha(\delta)\bar{a}\beta_2) \geq 0 \tag{4.16}$$

for sufficiently small $\delta > 0$. Let

$$K(\delta) := (\bar{a}+\bar{b}+1)\bar{a}\beta_2\varphi(\delta)\alpha(\delta) - (\bar{a}+\bar{b})\psi(\delta) - G(\alpha(\delta)\bar{a}\beta_2) \ (\delta > 0), \quad K(0) := \lim_{\delta\to 0+} K(\delta)\,.$$

Regarding $\lim_{\delta\to 0+} \alpha(\delta) = 0$ for (4.9) one has $K(0) = 0$. We prove that $K'(\delta) > 0$, for sufficiently small $\delta > 0$, and this will have as consequence the fact that $K(\delta) > K(0) = 0$, for sufficiently small $\delta > 0$. Indeed, one has for sufficiently small $\delta > 0$

$$\begin{aligned} K'(\delta) &= (\bar{a} + \bar{b} + 1)\varphi'(\delta)f(\varphi(\delta)) + (\bar{a} + \bar{b} + 1)\varphi(\delta)f'(\varphi(\delta))\varphi'(\delta) \\ &\quad -(\bar{a} + \bar{b})\varphi'(\delta)f(\varphi(\delta)) - f^{-1}(f(\varphi(\delta)))f'(\varphi(\delta))\varphi'(\delta) \\ &= \varphi'(\delta)f(\varphi(\delta)) + (\bar{a} + \bar{b})\varphi(\delta)f'(\varphi(\delta))\varphi'(\delta) > 0, \end{aligned}$$

since by construction f has a positive derivative $f'(s)$ for all $s > 0$. Thus (4.15) holds and this yields the estimate

$$D_\xi(x^\delta_{\alpha(\delta)}, x^\dagger) \leq c_0\varphi(\delta)$$

for sufficiently small $\delta > 0$ and some constant $c_0 > 0$. □

Example 4.4 We conclude this section with the example situation of monomials (power functions) $\varphi(t) = t^\kappa$ $(0 < \kappa \leq 1)$ and $\psi(t) = t^p$ $(p > 1)$ discussed in [16] for which Proposition 4.1 was repeated above. Then our assumptions are satisfied and we have

$$H(t) = t^{p/\kappa}, \ \ f(t) = \frac{p}{\kappa}\, t^{(p-\kappa)/\kappa}, \ \ G(t) \sim t^{p/(p-\kappa)}, \ \ \alpha(\delta) \sim \delta^{p-\kappa}, \ \ D_\xi(x^\delta_{\alpha(\delta)}, x^\dagger) \ = \ \mathcal{O}(\delta^\kappa)\,.$$

We should mention here that $\alpha(\delta) \sim G^{-1}(\psi(\delta))$ for the example with monomials. Hence the regularization parameter is chosen such that both terms in the numerator of the second fraction in (4.14) are equilibrated up to constant. Such equilibration yields frequently order optimal convergence rates in regularization.

Furthermore, we would like to notice that one comes to the same conclusion also in the case $0 < \kappa < p \leq 1$ discussed in [6]. The reason therefore lays in the fact that f remains an index function and, consequently, Theorem 4.3 is still applicable, even if in this situation ψ fails to be strictly convex. In fact, in order to obtain the convergence rate (4.8) in Theorem 4.3 one needs only to guarantee that the function f defined as in (4.5) is an index function which is differentiable on the interior of its domain. This happens when Assumption 4.2 is satisfied, but can be the case also in other settings.

5 Variational inequalities based on canonical source conditions and approximate source conditions

In this section we are going to formulate sufficient conditions for variational inequalities (4.4) when only some weak structural assumption of the form (3.9) on the nonlinearity of F with concave index function σ is imposed.

Theorem 5.1 *Assume that $F, \Omega, \mathcal{D}, X, Y, x^\dagger, \xi$ and ψ satisfy the Assumption 3.1. Let ξ satisfy the canonical source condition (3.2) and the structural condition (3.9) with some index function σ and some constant $C > 0$ for all $x \in \mathcal{M}_{\alpha_{max}}(\rho)$. Then a variational inequality (4.4) holds with two multipliers $0 \le \beta_1 < 1,\ \beta_2 > 0$ and with the index function $\varphi = \sigma$.*

Proof: Owing to (3.2) and (3.9) we can estimate for all $x \in \mathcal{M}_{\alpha_{max}}(\rho)$ as

$$\langle \xi, x^\dagger - x\rangle_{X^*,X} = \langle F'(x^\dagger)^* w, x^\dagger - x\rangle_{X^*,X} = \langle w, F'(x^\dagger)(x^\dagger - x)\rangle_{Y^*,Y}$$
$$\le \|w\|_{Y^*}\|F'(x^\dagger)(x - x^\dagger)\|_Y \le C\,\|w\|_{Y^*}\sigma(\|F(x) - F(x^\dagger)\|_Y)\,.$$

This, however, yields the variational inequality (4.4) with $\beta_1 = 0 < 1,\ \beta_2 = C\,\|w\|_{Y^*}$ and with $\varphi = \sigma$, where σ is the index function from (3.9). This proves the theorem. $\square$

Theorem 5.2 *Assume that X, Y are reflexive Banach spaces, $F, \Omega, \mathcal{D}, X, Y, x^\dagger, \xi$ and ψ satisfy the Assumption 3.1, and $F'(x^\dagger)$ is an injective operator. Let $\xi \notin \mathcal{R}(F'(x^\dagger)^*)$. Moreover, assume that the structural condition (3.9) is fulfilled with some index function σ and some constant $C > 0$ for all $x \in \mathcal{M}_{\alpha_{max}}(\rho)$ and that the Bregman distance is locally q-coercive with $2 \le q < \infty$, i.e. there is some constant $c_q > 0$ such that*

$$D_\xi(x, x^\dagger) \ge c_q\,\|x - x^\dagger\|_X^q \tag{5.1}$$

holds for all $x \in \mathcal{M}_{\alpha_{max}}(\rho)$. Then a variational inequality (4.4) holds for all $x \in \mathcal{M}_{\alpha_{max}}(\rho)$ with two multipliers $0 \le \beta_1 < 1,\ \beta_2 > 0$ and with the index function $\varphi(0) = 0,\ \varphi(t) = [d(\Psi^{-1}(\sigma(t)))]^{q^*}\ (t > 0)$, where $\frac{1}{q} + \frac{1}{q^*} = 1$ and $\Psi : (0,\infty) \to (0,\infty)$, $\Psi(R) := \frac{d(R)^{q^*}}{R}$.

Proof: Instead of (3.2) we have here for all $R > 0$ the equations $\xi = F'(x^\dagger)^* w_R + r_R$ with $\|w_R\|_{Y^*} \le R$ and $\|r_R\|_{X^*} = d(R)$. By using (3.9) we get for all $x \in \mathcal{M}_{\alpha_{max}}(\rho)$ the following estimate

$$\langle \xi, x^\dagger - x\rangle_{X^*,X} = \langle F'(x^\dagger)^* w_R + r_R, x^\dagger - x\rangle_{X^*,X} = \langle w_R, F'(x^\dagger)(x^\dagger - x)\rangle_{Y^*,Y} + \langle r_R, x^\dagger - x\rangle_{X^*,X}$$
$$\le R\,\|F'(x^\dagger)(x - x^\dagger)\|_Y + d(R)\,\|x - x^\dagger\| \le R\,C\,\sigma(\|F(x) - F(x^\dagger)\|_Y) + d(R)\,\|x - x^\dagger\|\,.$$

Now for q and q^* adjoint exponents with $1/q + 1/q^* = 1$ the inequality

$$\langle \xi, x^\dagger - x\rangle_{X^*,X} \le R\,C\,\sigma(\|F(x) - F(x^\dagger)\|_Y) + c_q^{-1/q}\,d(R)\,D_\xi(x, x^\dagger)^{1/q}$$

obtained from (5.1) can be further handled by using Young's inequality in the standard form

$$a\,b \le \frac{a^{p_1}}{p_1} + \frac{b^{p_2}}{p_2} \qquad (a, b \ge 0,\ \ p_1, p_2 > 1 \ \text{ with } \ \frac{1}{p_1} + \frac{1}{p_1} = 1)$$

when setting $a := D_\xi(x, x^\dagger),\ b := c_q^{-1/q}\,d(R),\ p_1 := q,\ p_2 := q^*$. In that way we derive for all $R > 0$

$$\langle \xi, x^\dagger - x\rangle_{X^*,X} \le R\,C\,\sigma(\|F(x) - F(x^\dagger)\|_Y) + \frac{1}{q}\,D_\xi(x, x^\dagger) + \frac{c_q^{-q^*/q}}{q^*}\,d(R)^{q^*}\,.$$

The continuity of d carries over to the auxiliary function $\Psi : (0,\infty) \to (0,\infty)$, $\Psi(R) = \frac{d(R)^{q^*}}{R}$, which is continuous and strictly decreasing, and which fulfills $\lim_{R\to 0} \Psi(R) = \infty$ and $\lim_{R\to\infty} \Psi(R) = 0$. Its inverse $\Psi^{-1} : (0,\infty) \to (0,\infty)$ is also continuous and strictly decreasing and for all $t > 0$ the equation $\Psi(R) = \sigma(t)$ has a uniquely determined solution $R > 0$. Note that for rates results only sufficiently small $t > 0$ are of interest. Setting $R := \Psi^{-1}\left(\sigma(\|F(x) - F(x^\dagger)\|_Y)\right)$ we get some constant $\hat{C} > 0$ such that the variational inequality

$$\langle \xi, x^\dagger - x \rangle_{X^*,X} \le \frac{1}{q} D_\xi(x, x^\dagger) + \hat{C}\, \left[d\left(\Psi^{-1}(\sigma(\|F(x) - F(x^\dagger)\|_Y))\right)\right]^{q^*}$$

holds for all $x \in \mathcal{M}_{\alpha_{max}}(\rho)$. Now the function defined by $\zeta(s) := d \circ \Psi^{-1} \circ \sigma(s)$ when $s > 0$ with extension $\zeta(0) := 0$ is an index function. Namely, ζ is continuous on $(0,\infty)$, since d is continuous. Moreover, the limit $\lim_{R\to\infty} d(R) = 0$ implies $\lim_{t\to 0} \zeta(t) = 0$ and this ensures the continuity of ζ in 0. On the other hand, by Lemma 3.2 one has that d is strictly decreasing. Thus ζ is strictly increasing, and hence an index function.

Because of $0 < \frac{1}{q} < 1$ this proves the theorem, since $\varphi := \zeta^{q^*}$, namely $\varphi(0) = 0$ and $\varphi(t) = \left[d\left(\Psi^{-1}(\sigma(t))\right)\right]^{q^*}$ when $t > 0$, is an index function, too. □

Remark 5.3 One can easily see that the rate function $[d \circ \Psi^{-1} \circ \sigma]^{q^*}(t)$ in the variational inequality of Theorem 5.2 tends to zero as $t \to 0$ *slower* than the associated rate function $\sigma(t)$ in the variational inequality of Theorem 5.1. Namely, taking into account the one-to-one correspondence between large $R > 0$ and small t via $\Psi(R) = \sigma(t)$ and $\Psi(R) = \frac{d(R)^{q^*}}{R}$ we have for the quotient function

$$\frac{\sigma(t)}{\left[d\left(\Psi^{-1}(\sigma(t))\right)\right]^{q^*}} = \frac{\Psi(R)}{d(R)^{q^*}} = \frac{1}{R} \to 0 \quad \text{as} \quad R \to \infty, \;\text{ resp.} \quad t \to 0\,.$$

As a consequence the situation of approximate source conditions occurring in Theorem 5.2 leads to lower convergence rates of Tikhonov regularization obtained from Theorem 4.3 than the situation of canonical source conditions that appears in Theorem 5.1.

Example 5.4 Concerning logarithmic rates, as an example, we are going to conclude the paper with a brief study that outlines the specific potential of variational inequalities (4.4) for extracting both solution smoothness of ξ and nonlinearity conditions on F at $x^\dagger$ in one index function φ which determines the convergence rate. Let in this example with some $C > 0$

$$\varphi(t) \;=\; \begin{cases} 0 & (t = 0) \\ C\,[\log(1/t)]^{-\mu} & (0 < t \le e^{-\mu-1}) \end{cases} \tag{5.2}$$

hold. By Theorem 4.3 we immediately derive for all $\mu > 0$ the logarithmic convergence rate

$$D_\xi(x^\delta_{\alpha(\delta)}, x^\dagger) \;=\; \mathcal{O}\left([\log(1/\delta)]^{-\mu}\right) \quad \text{as} \quad \delta \to 0\,, \tag{5.3}$$

which is slower than every power rate (4.2) for any $\kappa > 0$. Now the function (5.2) with slow decay to zero as $t \to 0$ can be a consequence of two completely different causes characterized by the following two situations (I) and (II), respectively:

(I) Let $\sigma = \varphi$, i.e., a very weak logarithmic structural condition (3.9) is valid, and assume that the canonical source condition (3.2) holds, which expresses in our context the strong smoothness assumption on the solution. Then by Theorem 5.1 in connection with Theorem 4.3 we obtain the logarithmic convergence rate (5.3).

(II) Let $\sigma(t) = t$, i.e., a structural condition (3.9) is satisfied, which is the strongest in our sense. However, the canonical source condition (3.2) is strongly violated, which is expressed by a logarithmic decay

$$d(R) = (\log R)^{-\nu}$$

of the corresponding distance function for some $\nu > 0$ and all sufficiently large $R > \overline{R} > 0$. However, since we have for all such R and for $\varepsilon > 0$ a constant $K > 0$ with

$$\Psi(R) = \frac{1}{R(\log R)^{\nu q^*}} \geq \frac{K}{R^{1+\varepsilon}}\,,$$

this implies $\Psi^{-1}(t) \geq \hat{K} t^{-1/(1+\varepsilon)}$ for some constant $\hat{K} > 0$ and sufficiently small $t > 0$. Hence, by Theorem 5.2 the function φ in (4.4) attains the form (5.2) with $\mu = \nu q^*$.

Acknowledgements

Both authors express their thanks to three anonymous referees for important hints that improved the paper in some essential points. The research of R.I. Boţ was partially supported by Deutsche Forschungsgemeinschaft (DFG) under Grant WA922/1-3. The research of B. Hofmann was partially supported by Deutsche Forschungsgemeinschaft (DFG) under Grant HO1454/7-2.

References

[1] Bonesky, T.; Kazimierski, K.S.; Maass, P.; Schöpfer, F.; Schuster, T. (2008): Minimization of Tikhonov functionals in Banach spaces. *Abstract and Applied Analysis* **2008**. Article ID 192679 (19 pp), DOI:10.1155/2008/192679.

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

[3] Düvelmeyer, D.; Hofmann, B.; Yamamoto, M. (2007): Range inclusions and approximate source conditions with general benchmark functions. *Numerical Functional Analysis and Optimization* **28**, 1245–1261.

[4] Engl, H. W.; Hanke, M.; Neubauer, A. (1996): *Regularization of Inverse Problems*. Dordrecht: Kluwer.

[5] Engl, H. W.; Kunisch, K.; Neubauer, A. (1989): Convergence rates for Tikhonov regularization of nonlinear ill-posed problems. *Inverse Problems* **5**, 523–540.

[6] Geissler, J.; Hofmann, B. (2009): A new approach to source conditions in regularization with general residual term. Paper published electronically as arXiv:0906.3438v1[math.NA] under http://arxiv.org/abs/0906.3438v1

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

[8] Groetsch, C. W. (1984): *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind.* Boston: Pitman.

[9] Groetsch, C. W. (2007): *Stable Approximate Evaluation of Unbounded Operators.* Berlin: Springer.

[10] Hegland, M. (1995): Variable Hilbert scales and their interpolation inequalities with applications to Tikhonov regularization. *Appl. Anal.* **59**, 207–223.

[11] Hein, T. (2009): Tikhonov regularization in Banach spaces – improved convergence rates results. *Inverse Problems* **25**, 035002 (18pp).

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

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

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

[15] Hofmann, B.; Scherzer, O. (1994): Factors influencing the ill-posedness of nonlinear problems. *Inverse Problems* **10**, 1277–1297.

[16] Hofmann, B.; Yamamoto, M. (2010): On the interplay of source conditions and variational inequalities for nonlinear ill-posed problems. *Applicable Analysis* **89**. To appear.

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

[18] Kaltenbacher, B. (2008): A note on logarithmic convergence rates for nonlinear Tikhonov regularization. *J. Inv. Ill-Posed Problems* **16**, 79–88.

[19] Losonczi, L. (1984): Inequalities of Young-type. *Mh. Math.* **97**, 125–132.

[20] Neubauer, A. (2009): On enhanced convergence rates for Tikhonov regularization of nonlinear ill-posed problems in Banach spaces. *Inverse Problems* **25**, 065009 (10 pp).

[21] Neubauer, A.; Hein, T.; Hofmann, B.; Kindermann, S.; Tautenhahn, U. (2009): Enhanced convergence rates for Tikhonov regularization revisited: improved results. RICAM Report 2009-14. Linz: Johann Radon Institute for Computational and Applied Mathematics,

[22] PÖSCHL, C. (2008): *Tikhonov Regularization with General Residual Term*. PhD Thesis (Dissertation). Innsbruck: Leopold Franzens Universität.

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

[24] RESMERITA, E.; SCHERZER, O. (2006): Error estimates for non-quadratic regularization and the relation to enhancement. *Inverse Problems* **22**, 801–814.

[25] SCHERZER, O.; GRASMAIR, M.; GROSSAUER, H.; HALTMEINER, M.; LENZEN, F. (2009): *Variational Methods in Imaging*. New York: Springer.

[26] SCHOCK, E. (1985): Approximate solution of ill-posed equations: arbitrarily slow convergence vs. superconvergence. In: *Constructive Methods for the Practical Treatment of Integral Equations* (Eds.: G. Hämmerlin and K. H. Hoffmann). Basel: Birkhäuser, 234–243.

[27] SCHÖPFER, F.; LOUIS, A.K.; SCHUSTER, T. (2006): Nonlinear iterative methods for linear ill-posed problems in Banach spaces. *Inverse Problems* **22**, 311–329.

[28] TIKHONOV, A. N.; LEONOV, A. S.; YAGOLA, A. G. (1998): *Nonlinear Ill-Posed Problems*, Vol. 1. Series Applied Mathematics and Mathematical Computation, Vol. 14. Translated from the Russian. London: Chapman & Hall.

[29] ZĂLINESCU, C. (2002): *Convex Analysis in General Vector Spaces*. River Edge: World Scientific