A new approach to source conditions in regularization with general residual term

Jens Flemming* and Bernd Hofmann*

March 3, 2010

Abstract

This paper addresses Tikhonov like regularization methods with convex penalty functionals for solving nonlinear ill-posed operator equations formulated in Banach or, more general, topological spaces. We present an approach for proving convergence rates which combines advantages of approximate source conditions and variational inequalities. Precisely, our technique provides both a wide range of convergence rates and the capability to handle general and not necessarily convex residual terms as well as nonsmooth operators. Initially formulated for topological spaces, the approach is extensively discussed for Banach and Hilbert space situations, showing that it generalizes some well-known convergence rates results.

1 Introduction

In recent years because of numerous applications which occurred in imaging, natural sciences, engineering, and mathematical finance a growing interest in different forms of regularization methods for solving nonlinear ill-posed inverse problems in a Banach space setting could be observed. This also led to new ideas for proving convergence rates of such methods in Banach spaces (see, e.g., [2, 5, 11–13, 15, 18, 20–28]). The main problem of handling ill-posed problems in Banach spaces is the absence of spectral theoretic tools including generalized source conditions with arbitrary index functions applied to the forward operator, which were essential for proving results in the Hilbert space setting.

One way for obtaining convergence rates similar to the well-known Hilbert space results is the idea of so-called approximate source conditions, which was originally developed for linear ill-posed problems in [14] (see also [6, 17]) and extended to nonlinear problems in Banach spaces in [13]. Approximate source conditions, however, rely heavily on the traditional residual structure being a p-th power of the discrepancy norm. Therefore they are not suited for investigating convergence rates of variational regularization methods with general residual terms using appropriate similarity measures. Such progressive variants of variational regularization were suggested, comprehensively analyzed, and motivated by means of concrete examples in [23].

A second approach, which uses variational inequalities for proving convergence rates, was first formulated in [18] and has also been extended to general residual terms in [23]. The drawback of this second approach in its original form is its limitation to the standard convergence rate $\mathcal{O}(\delta)$ for noise level $\delta > 0$ when the reconstruction error is measured by a Bregman distance.

In this paper, which is mainly based on the thesis [8], we present an alternative concept that allows both a wide range of different convergence rates and the use of general residual terms.

*Department of Mathematics, Chemnitz University of Technology, D-09107 Chemnitz, Germany. E-mail addresses: jens.flemming@mathematik.tu-chemnitz.de ; hofmannb@mathematik.tu-chemnitz.de.

Moreover, we give some new insight into the interplay of source conditions and variational inequalities by extending the ideas of [15] and [19] to a more general setting. Furthermore, we address the question concerning the role and admissible intervals of an exponent $p > 0$ imposed on the residual term in Tikhonov type regularization (see also [16]).

The paper is organized as follows: In Section 2 we introduce a Tikhonov type regularization method for the stable approximate solution of nonlinear ill-posed operator equations in topological vector spaces with focus on Banach spaces. We formulate basic assumptions ensuring well-definedness, stability and convergence of the method and we give a short discussion on fundamental differences between the classical source conditions and variational inequalities. In Section 3 we extend the concept of variational inequalities introduced in [18]. Moreover, we formulate a first convergence rate result in that section. Based on Section 3 in Section 4, which is the main section of this paper, we present the new approach of approximate variational inequalities for proving convergence rates of variational regularization methods with general residual term. In Section 5 we restrict our investigations to Banach spaces to clarify the interplay of approximate source conditions and approximate variational inequalities. The final Section 6 is devoted to some concluding remarks where also open questions are formulated. Some technical proofs will be postponed to the appendix.

2 Problem, notation, and basic assumptions

Let $F : D(F) \subseteq U \to V$ be an in general nonlinear operator possessing the domain $D(F)$ and mapping between a real topological vector space U and a topological space V with topologies τ_U and τ_V. We are going to study operator equations

$$F(u) = v^0 \tag{2.1}$$

expressing inverse problems with exact data $v^0 \in V$ on the right-hand side.

To ensure mathematical correctness we assume that U and V are Hausdorff spaces, which implies that the limit of any convergent sequence is uniquely determined. The topologies τ_U und τ_V should be regarded as "weak" topologies because as we will see later in infinite dimensional Banach space settings they have to be weaker than the norm topologies. For this reason we denote convergence with respect to the topology τ_U or τ_V by "$\rightharpoonup$".

Instead of the exact right-hand side v^0 in (2.1) only noisy data v^δ for some noise level $\delta > 0$ are available. To clarify the meaning of δ we introduce a non-negative similarity functional $\mathcal{S} : V \times V \to [0, \infty]$, which not necessarily has to have metric properties, and demand

$$\mathcal{S}(v^\delta, v^0) \le \delta \qquad \text{and} \qquad \mathcal{S}(v^0, v^\delta) \le \delta. \tag{2.2}$$

As approximate solutions of (2.1) we consider minimizers u_α^δ over $D(F)$ of the Tikhonov type functional

$$\mathcal{T}_\alpha^\delta(u) := \mathcal{S}(F(u), v^\delta)^p + \alpha\Omega(u) \tag{2.3}$$

with a stabilizing functional $\Omega : U \to [0, \infty]$, a regularization parameter $\alpha > 0$, and a prescribed exponent $0 < p < \infty$. We set

$$D(\Omega) := \{u \in U : \Omega(u) < \infty\} \qquad \text{and} \qquad D := D(F) \cap D(\Omega).$$

The setting described above can be modified in different ways: One is to set $\tilde{\mathcal{S}}(v_1, v_2) := \mathcal{S}(v_1, v_2)^p$ and to bound $\tilde{\mathcal{S}}(v^\delta, v^0)$ and $\tilde{\mathcal{S}}(v^0, v^\delta)$ by δ. In this case all the results of this paper would remain true with $p = 1$, but convergence rates would look different since the noise level δ then has a different meaning. In more detail, we then have

$$\mathcal{S}(v^\delta, v^0) \le \delta,\ \mathcal{S}(v^0, v^\delta) \le \delta \quad \Leftrightarrow \quad \tilde{\mathcal{S}}(v^\delta, v^0) \le \delta^p,\ \tilde{\mathcal{S}}(v^0, v^\delta) \le \delta^p.$$

To avoid difficulties in interpreting the convergence rates results of this paper we do not use the $\tilde{\mathcal{S}}$-setting, though some constants would become less complex. Another modification of our setting is to replace the exponent p by a more general function $\psi : [0, \infty] \to [0, \infty]$, i.e. to write $\psi(\mathcal{S}(F(u), v^\delta))$ instead of $\mathcal{S}(F(u), v^\delta)^p$ in the Tikhonov functional; this approach is carried out in [3].

Throughout this paper we make the following assumptions.

Assumption 2.1.

(i) $F : D(F) \subseteq U \to V$ is sequentially continuous with respect to τ_U and τ_V, i.e. $u_k \rightharpoonup u$ with $u, u_k \in D(F)$ implies $F(u_k) \rightharpoonup F(u)$.

(ii) $D(F)$ is sequentially closed with respect to τ_U, i.e $u_k \rightharpoonup u$ with $u_k \in D(F)$ and $u \in U$ implies $u \in D(F)$.

(iii) There exists a $u \in D$ with $F(u) = v^0$, in particular $D \neq \emptyset$.

(iv) The following assertions are fulfilled by $\mathcal{S}$ (for all sequences $(v_k)_{k\in\mathbb{N}}$ and $(\tilde{v}_k)_{k\in\mathbb{N}}$ in V and $v, \tilde{v} \in V$):

  (a) $\mathcal{S}(v, \tilde{v}) = 0$ if and only if $v = \tilde{v}$.

  (b) There exists a value $s \ge 1$ with
$$\mathcal{S}(v_1, v_2) \le s\mathcal{S}(v_1, v_3) + s\mathcal{S}(v_3, v_2) \quad \text{for all} \quad v_1, v_2, v_3 \in V. \tag{2.4}$$

  (c) $\mathcal{S}$ is sequentially lower semi-continuous with respect to $\tau_V \times \tau_V$, i.e. if $v_k \rightharpoonup v$ and $\tilde{v}_k \rightharpoonup \tilde{v}$ then $\mathcal{S}(v, \tilde{v}) \le \liminf_{k\to\infty} \mathcal{S}(v_k, \tilde{v}_k)$.

  (d) $\mathcal{S}(v_k, v) \to 0$ implies $v_k \rightharpoonup v$.

  (e) If $\mathcal{S}(v_k, v) \to 0$, $\mathcal{S}(v, v_k) \to 0$, and $\mathcal{S}(\tilde{v}, v) < \infty$ then $\mathcal{S}(\tilde{v}, v_k) \to \mathcal{S}(\tilde{v}, v)$.

(v) Ω is convex.

(vi) Ω is sequentially lower semi-continuous with respect to τ_U, i.e. $u_k \rightharpoonup u$ implies $\Omega(u) \le \liminf_{k\to\infty} \Omega(u_k)$.

(vii) For each $\alpha > 0$ and each $c > 0$ the level sets
$$M_\alpha(c) := \{u \in D : \mathcal{T}_\alpha^0(u) \le c\} \tag{2.5}$$
are sequentially pre-compact with respect to τ_U, i.e. each sequence $(u_k)_{k\in\mathbb{N}}$ in $M_\alpha(c)$ has a subsequence which converges with respect to τ_U.

In the sequel for simplicity we will use the terms "continuous", "closed", and so on instead of "sequentially continuous", "sequentially closed", and so on if no confusion is to be expected. By U^* we denote the dual space of U, i.e. U^* is the set of all τ_U-continuous linear functionals on U. For $\xi \in U^*$ and $u \in U$ we write $\xi(u)$ if we evaluate the functional ξ at the point u. If U is a Banach space and τ_U is the weak topology on U, then we exploit the usual notation $\langle \xi, u\rangle_{U^*,U} := \xi(u)$.

Example 2.2. Let U and V be Banach spaces and let τ_U and τ_V be the corresponding weak topologies, i.e.

$$u_k \rightharpoonup u \quad \Leftrightarrow \quad \langle \xi, u_k \rangle_{U^*,U} \to \langle \xi, u \rangle_{U^*,U} \ \forall \xi \in U^*. \tag{2.6}$$

Then the similarity functional

$$\mathcal{S}(v_1, v_2) := \|v_1 - v_2\|_V \tag{2.7}$$

fulfills (iv) in Assumption 2.1 with $s = 1$. For a further discussion of this example see Section 5 below.

The next example, which is taken from [23], shows that next to norms also other similarity functionals are of interest.

Example 2.3. Let (X, ρ) be a complete, separable metric space. By $B(X)$ we denote the family of all Borel subsets of X, i.e. $B(X)$ is the σ-algebra generated by the ρ-open sets in X, and by $P(X)$ we denote the family of all Borel probability measures on X, i.e. $P(X)$ is the family of all measures $\mu : B(X) \to [0, \infty)$ satisfying $\mu(X) = 1$. For $1 \leq q < \infty$ and some $x_0 \in X$ (the concrete choice has no influence on the definition) we set

$$V := \left\{ \mu \in P(X) : \int_X \rho(\bullet, x_0)^q \,\mathrm{d}\mu < \infty \right\} \tag{2.8}$$

and as topology τ_V we choose the *narrow topology* on V, i.e. a series $(\mu_k)_{k\in\mathbb{N}}$ in V converges to $\mu \in V$ with respect to τ_V if and only if

$$\int_X f \,\mathrm{d}\mu_k \to \int_X f \,\mathrm{d}\mu \tag{2.9}$$

for all continuous and bounded real functions f defined on X.

For defining the similarity functional $\mathcal{S}$ we introduce the set $\Gamma(\mu_1, \mu_2) \subseteq P(X \times X)$ (with $\mu_1, \mu_2 \in V$) consisting of all measures $\underline{\mu} \in P(X \times X)$ satisfying $\underline{\mu}((\pi_i)^{-1}(A)) = \mu_i(A)$ for all $A \in B(X)$ and $i = 1, 2$, where $\pi_1(x_1, x_2) := x_1$ and $\pi_2(x_1, x_2) := x_2$. Then the similarity functional $\mathcal{S}$ defined by

$$\mathcal{S}(\mu_1, \mu_2) := \left(\inf_{\underline{\mu} \in \Gamma(\mu_1, \mu_2)} \int_{X \times X} \rho^q \,\mathrm{d}\underline{\mu} \right)^{\frac{1}{q}}, \quad \mu_1, \mu_2 \in V, \tag{2.10}$$

is a metric, the *Wasserstein metric*. Thus, items (iv)(a) and (iv)(b) of Assumption 2.1 are satisfied with $s = 1$. Items (c) and (d) follow from [9, Lemma 1] and [1, Proposition 7.1.5], respectively. By (b) we have

$$-\mathcal{S}(v_k, v) \leq \mathcal{S}(\tilde{v}, v_k) - \mathcal{S}(\tilde{v}, v) \leq \mathcal{S}(v, v_k)$$

and thus item (e) is satisfied, too.

This similarity functional has been applied to flow, mass transport, and image registration problems. For details on applications and some references we refer to [23].

In connection with the exponent p in (2.3) from time to time we will make use of the inequality

$$(a + b)^p \leq c_p(a^p + b^p) \tag{2.11}$$

for $a \geq 0$ and $b \geq 0$ with

$$c_p := \begin{cases} 1 & \text{if } 0 < p < 1, \\ 2^{p-1} & \text{if } p \geq 1. \end{cases} \tag{2.12}$$

Under Assumption 2.1 one can show that there exist minimizers of the Tikhonov functional (2.3) for all $p > 0$ and that these minimizers are stable with respect to perturbations of the data v^δ. The ideas of corresponding proofs given in Section 2 of [18] and in [23] can be applied to our general setting, and we note that hence existence and stability of minimizers can be ensured also in the case of exponents $0 < p < 1$ in Example 2.2. With the exception of [23] and [4] that case was mostly faded out in the literature.

To formulate assertions about convergence of a series of minimizers as δ tends to zero we need the concept of Ω-minimizing solutions: An element $u^\dagger \in D$ is called *Ω-minimizing solution* if

$$F(u^\dagger) = v^0 \quad \text{and} \quad \Omega(u^\dagger) = \inf\{\Omega(u) : u \in D,\, F(u) = v^0\}. \tag{2.13}$$

Under Assumption 2.1 one can show that there exists an Ω-minimizing solution. The following theorem was proven in [23].

Theorem 2.4. *Assume that Assumption 2.1 is satisfied. Let $(\delta_k)_{k\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ monotonically decreasing to zero, let $\alpha : (0, \delta_1] \to (0, \infty)$ be a parameter choice with $\alpha(\delta) \to 0$ and $\frac{\delta^p}{\alpha(\delta)} \to 0$ as $\delta \to 0$, and let $\alpha_k := \alpha(\delta_k)$ and $v_k := v^{\delta_k}$. Then every sequence $(u_k)_{k\in\mathbb{N}}$ in U with $u_k \in \operatorname{argmin}\{\mathcal{S}(F(u), v_k)^p + \alpha\Omega(u) : u \in D\}$ has a τ_U-convergent subsequence and the limit of each τ_U-convergent subsequence is an Ω-minimizing solution. If the Ω-minimizing solution is unique then (u_k) converges to this Ω-minimizing solution.*

To express convergence rates we use Bregman distances, which have become quite popular in recent years for this purpose. In this context, let

$$\tilde{u} \in D_B := \{u \in U : \partial\Omega(u) \neq \emptyset\} \quad \text{and} \quad \xi \in \partial\Omega(\tilde{u}) \subseteq U^*,$$

where $\partial\Omega(u)$ denotes the subdifferential of Ω at u. Then the functional

$$\mathcal{B}_\xi(u, \tilde{u}) := \Omega(u) - \Omega(\tilde{u}) - \xi(u - \tilde{u}), \quad u \in U,$$

is called *Bregman distance* with respect to Ω, $\tilde{u}$, and ξ. In the sequel we always assume that there exists an Ω-minimizing solution $u^\dagger \in D_B$.

At the end of this section we want to mention the two basic concepts occurring in the literature for proving convergence rates of Tikhonov type regularization of ill-posed equations. Classical source conditions, as, e.g., in a Banach space setting $\xi = F'(u^\dagger)^*\,\eta$, $\eta \in V^*$, and in Hilbert spaces $u^\dagger = \varphi(F'(u^\dagger)^* F'(u^\dagger))\,w$, $w \in U$, i.e. sourcewise representations of an element ξ of the subdifferential of Ω for an Ω-minimizing solution or of an Ω-minimizing solution itself, are the main ingredient for proving convergence rates. Note that Hilbert space type source conditions are also available if U is a Banach space and V is a Hilbert space (see [25]). Such classical kinds of source conditions express the smoothness of the solution with respect to the operator and they alone are responsible for possible convergence rates in common linear ill-posed equations; exceptions, where also other factors influence the rates, are, e.g., sparsity constraint settings (see [4, 10]). If we are concerned with nonlinear operators F, then in addition we have to take into account structural conditions which express the nonlinearity. For nonlinear ill-posed equations classical source conditions and nonlinearity conditions together control convergence rates. Their interplay, however, is rather complicated.

Originally in [18] (see also [27]) an extended concept of source conditions was presented for obtaining convergence rates for the Banach space situation of Example 2.2. It is based on variational inequalities, which have to hold on appropriate level sets $M_\alpha(c)$ of the Tikhonov type functional (2.3). In [13, 15, 19] an additional exponent $\kappa \in (0,1]$ was introduced and motivated such that the variational inequalities attain the form

$$\langle \xi, u^\dagger - u \rangle_{U^*,U} \le \beta_1 \mathcal{B}_\xi(u, u^\dagger) + \beta_2 \|F(u) - F(u^\dagger)\|_V^\kappa. \tag{2.14}$$

If such a variational inequality holds, then a convergence rate $\mathcal{B}_\xi(u^\delta_{\alpha(\delta)}, u^\dagger) = \mathcal{O}(\delta^\kappa)$ as $\delta \to 0$ can immediately be derived without additional knowledge (i.e. without nonlinearity conditions or assumptions on solution smoothness) when appropriate a priori parameter choices are used. Both the classical source conditions and the structural conditions of nonlinearity result into one parameter, namely the exponent κ that alone controls the rate.

3 Variational inequalities and convergence rates

In this section we extend the concept of variational inequalities introduced in [18]. At first we state some simple properties of the level sets defined in (2.5).

Proposition 3.1. *Let $u^\dagger$ be an Ω-minimizing solution and let $\varrho > 0$ be an arbitrary constant. Then for $0 < \alpha_1 \le \alpha_2$ we have*

(i) $M_{\alpha_1}(\varrho) \supseteq M_{\alpha_2}(\varrho)$,

(ii) $M_{\alpha_1}(\varrho\alpha_1) \subseteq M_{\alpha_2}(\varrho\alpha_2)$,

(iii) $\bigcap_{\alpha > 0} M_\alpha(\varrho\alpha) = \{u \in D : F(u) = v^0,\, \Omega(u) \le \varrho\}$.

Proof. Item (i) is trivial. Item (ii) follows from

$$\begin{aligned} \mathcal{T}^0_{\alpha_2}(u) &= \mathcal{S}(F(u), v^0)^p + \alpha_1 \Omega(u) - (\alpha_1 - \alpha_2)\Omega(u) \\ &\le \varrho\alpha_1 - (\alpha_1 - \alpha_2)\Omega(u) = \varrho\alpha_2 + (\alpha_1 - \alpha_2)(\varrho - \Omega(u)) \le \varrho\alpha_2 \end{aligned}$$

for all $u \in M_{\alpha_1}(\varrho\alpha_1)$ and (iii) from $\mathcal{S}(F(u), v^0)^p \le \alpha(\varrho - \Omega(u))$ for all $\alpha > 0$ and for $u \in \bigcap_{\alpha>0} M_\alpha(\varrho\alpha)$. □

The next proposition shows the importance of the level sets $M_\alpha(\varrho\alpha)$.

Proposition 3.2. *Let $u^\dagger$ be an Ω-minimizing solution, $\bar\alpha > 0$, and*

$$\varrho > c_p\, s^p\, \Omega(u^\dagger)\,. \tag{3.1}$$

Further let $\delta \mapsto \alpha(\delta)$ be an a priori parameter choice satisfying

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

and let $u^\delta_{\alpha(\delta)} \in \operatorname{argmin}\{\mathcal{T}^\delta_{\alpha(\delta)}(u) : u \in D\}$ for $\delta > 0$. Then there exists some $\bar\delta > 0$, such that $u^\delta_{\alpha(\delta)} \in M_{\bar\alpha}(\varrho\bar\alpha)$ holds for all $\delta \in (0, \bar\delta]$.

Proof. Because $\alpha(\delta) \to 0$ and $\frac{\delta^p}{\alpha(\delta)} \to 0$ as $\delta \to 0$ there exists some $\bar{\delta} > 0$ with $\alpha(\delta) \le \bar{\alpha}$ and $\frac{\delta^p}{\alpha(\delta)} \le \frac{\varrho}{2c_p s^p} - \frac{1}{2}\Omega(u^\dagger)$ for all $\delta \in (0, \bar{\delta}]$. For the sake of brevity we write α instead of $\alpha(\delta)$. For $\delta \in (0, \bar{\delta}]$ we now have (in analogy to [13, p. 5])

$$\begin{aligned}
\mathcal{T}^0_{\bar{\alpha}}(u^\delta_\alpha) &\le \big(s\mathcal{S}(F(u^\delta_\alpha), v^\delta) + s\delta\big)^p + \bar{\alpha}\Omega(u^\delta_\alpha) \\
&\le c_p s^p \big(\mathcal{S}(F(u^\delta_\alpha), v^\delta)^p + \alpha\Omega(u^\delta_\alpha) + \delta^p + (\bar{\alpha} - \alpha)\Omega(u^\delta_\alpha)\big) \\
&\le c_p s^p \Big(\delta^p + \alpha\Omega(u^\dagger) + \delta^p + \tfrac{\bar{\alpha}-\alpha}{\alpha}\alpha\Omega(u^\delta_\alpha)\Big) \\
&\le c_p s^p \Big(2\delta^p + \alpha\Omega(u^\dagger) + \tfrac{\bar{\alpha}-\alpha}{\alpha}\big(\mathcal{S}(F(u^\delta_\alpha), v^\delta)^p + \alpha\Omega(u^\delta_\alpha)\big)\Big) \\
&\le c_p s^p \big(\delta^p + \tfrac{\bar{\alpha}}{\alpha}\delta^p + \bar{\alpha}\Omega(u^\dagger)\big) \le c_p s^p \bar{\alpha}\big(2\tfrac{\delta^p}{\alpha} + \Omega(u^\dagger)\big) \le \varrho\bar{\alpha}. \qquad \square
\end{aligned}$$

We now give the basic definition of a variational inequality in a stronger sense.

Definition 3.3. An Ω-minimizing solution $u^\dagger$ *satisfies a variational inequality*, if there exist a $\xi \in \partial\Omega(u^\dagger)$ and constants ϱ fulfilling inequality (3.1), $\bar{\alpha} > 0$, $\beta_1 \in [0, 1)$, $\beta_2 \ge 0$, and $\kappa > 0$, such that

$$-\xi(u - u^\dagger) \le \beta_1 \mathcal{B}_\xi(u, u^\dagger) + \beta_2 \mathcal{S}(F(u), F(u^\dagger))^\kappa \tag{3.3}$$

holds for all $u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$.

As one would expect, a variational inequality with $\kappa = \kappa_0$ implies a variational inequality with $\kappa = \kappa_1$ for each $\kappa_1 \in (0, \kappa_0)$. The only changing constant in Definition 3.3 is the factor $\beta_2 = \beta_2(\kappa)$. This follows immediately from

$$\begin{aligned}
\beta_2(\kappa_0)\mathcal{S}(F(u), F(u^\dagger))^{\kappa_0} &= \beta_2(\kappa_0)\mathcal{S}(F(u), F(u^\dagger))^{\kappa_0 - \kappa_1}\mathcal{S}(F(u), F(u^\dagger))^{\kappa_1} \\
&\le \beta_2(\kappa_0)(\varrho\bar{\alpha})^{\frac{\kappa_0 - \kappa_1}{p}}\mathcal{S}(F(u), F(u^\dagger))^{\kappa_1}
\end{aligned}$$

because $u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ implies $\mathcal{S}(F(u), F(u^\dagger))^p \le \varrho\bar{\alpha}$.

For $\kappa = 1$ and for the case of topological spaces with general similarity functional $\mathcal{S}$ Definition 3.3 was introduced in [23]. The definition was already presented earlier in [18] for the Banach space situation with norm as similarity functional $\mathcal{S}$. For that situation and $\kappa \in (0, 1]$ this variational inequality (3.3) appeared also in [13, proof of Theorem 3.3].

The connection between classical source conditions and variational inequalities will be discussed in Section 5.

We now give a first convergence rate result, which will be proven later in a more general context.

Theorem 3.4. *Let $u^\dagger$ be an Ω-minimizing solution which satisfies a variational inequality in the sense of Definition 3.3 with $0 < \kappa < p$ and let $\delta \mapsto \alpha(\delta)$ be an a priori parameter choice with $\underline{c}\delta^{p-\kappa} \le \alpha(\delta) \le \overline{c}\delta^{p-\kappa}$ for sufficiently small δ and constants $\underline{c} > 0$, $\overline{c} > 0$. Then*

$$\mathcal{B}_\xi(u^\delta_{\alpha(\delta)}, u^\dagger) = \mathcal{O}(\delta^\kappa) \quad \text{as } \delta \to 0. \tag{3.4}$$

Note that the a priori parameter choice in Theorem 3.4 restricts the admissible values for the exponent κ to the interval $(0, p)$. As we will see this restriction is due to the proof technique using Young's inequality. On the other hand, the following proposition provides an upper bound for κ in a variational inequality (3.3) if Ω, $\mathcal{S}$, and F satisfy a quite weak smoothness assumption. A special case of this proposition was already formulated as Proposition 4.3 in the paper [19] (cf. also [15]).

Proposition 3.5. *Let $u^\dagger$ be an Ω-minimizing solution which satisfies a variational inequality in the sense of Definition 3.3. If there exist a $q > 0$, a $u \in U$ with $\xi(u) < 0$, and a $t_0 > 0$, such that $u^\dagger + tu \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ holds for all $t \in [0, t_0]$, the limits*

$$L_\Omega := \lim_{t\to+0} \frac{\Omega(u^\dagger + tu) - \Omega(u^\dagger)}{t}, \quad L_\mathcal{S} := \lim_{t\to+0} \frac{\mathcal{S}(F(u^\dagger + tu), F(u^\dagger))^q}{t},$$

i.e. the directional derivatives in $u^\dagger$ in direction u of Ω and $\mathcal{S}(F(\bullet), F(u^\dagger))^q$, exist, and $L_\Omega = \xi(u)$ is valid, then $\kappa \leq q$ must hold.

Proof. Let $\kappa > q$. For each $t \in (0, t_0]$ inequality (3.3) then implies

$$-\xi(tu) \leq \beta_1\big(\Omega(u^\dagger + tu) - \Omega(u^\dagger) - \xi(tu)\big) + \beta_2\mathcal{S}(F(u^\dagger + tu), F(u^\dagger))^\kappa$$

and thus

$$-\xi(u) \leq \beta_1\left(\tfrac{\Omega(u^\dagger+tu)-\Omega(u^\dagger)}{t} - \xi(u)\right) + \beta_2\left(\tfrac{\mathcal{S}(F(u^\dagger+tu),F(u^\dagger))^q}{t}\right)^{\frac{\kappa}{q}} t^{\frac{\kappa}{q}-1}.$$

Passage to the limit $t \to +0$ gives $\xi(u) \geq 0$, which is a contradiction to $\xi(u) < 0$. □

Remark 3.6. Under the standing assumptions of this paper on F, $D(F)$, Ω, and $u^\dagger$ one can easily show that for Banach spaces U and V and $\mathcal{S}(v_1, v_2) := \|v_1 - v_2\|_V$ the Proposition 3.5 applies for $q = 1$ when F and Ω are Gâteaux differentiable in $u^\dagger$. So in this case only variational inequalities (2.14) with $\kappa \leq 1$ can be satisfied if the singular case $\xi = \Omega'(u^\dagger) = 0$ is excluded.

4 Approximate variational inequalities

The aim of this section is to formulate convergence rates results without assuming that a variational inequality is satisfied. As in the method of approximate source conditions (see [6] and [13]) we use distance functions $d : [0, \infty) \to [0, \infty)$ measuring the violation of a prescribed benchmark condition. However, here we have a variational inequality (3.3) as benchmark condition and the distance functions are defined in a completely different manner.

If a benchmark inequality of type (3.3) is not satisfied then there exists at least one $u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ with

$$-\xi(u - u^\dagger) > \beta_1\mathcal{B}_\xi(u, u^\dagger) + \beta_2\mathcal{S}(F(u), F(u^\dagger))^\kappa.$$

Thus the "maximum violation" of a variational inequality (3.3) may be expressed by

$$\sup_{u\in M_{\bar{\alpha}}(\varrho\bar{\alpha})} \big(-\xi(u - u^\dagger) - \beta_1\mathcal{B}_\xi(u, u^\dagger) - \beta_2\mathcal{S}(F(u), F(u^\dagger))^\kappa\big). \tag{4.1}$$

The question whether the satisfaction of the benchmark inequality can be forced by increasing the factor β_2 leads to the definition of an approximate variational inequality.

Definition 4.1. An Ω-minimizing solution $u^\dagger$ *satisfies an approximate variational inequality* (approximate inequality for short) if there exist a $\xi \in \partial\Omega(u^\dagger)$ and constants ϱ fulfilling (3.1), $\bar{\alpha} > 0$, $\beta_1 \in [0, 1)$, and $\kappa > 0$, such that the function $d : [0, \infty) \to \mathbb{R}$ defined by

$$d(r) := -\min_{u\in M_{\bar{\alpha}}(\varrho\bar{\alpha})} \big(\xi(u - u^\dagger) + \beta_1\mathcal{B}_\xi(u, u^\dagger) + r\mathcal{S}(F(u), F(u^\dagger))^\kappa\big)$$

satisfies $d(r) \to 0$ as $r \to \infty$.

At first we formulate some basic properties of the distance function d, the proofs will be postponed to the appendix.

Proposition 4.2. *Let $u^\dagger$ be an Ω-minimizing solution which satisfies an approximate inequality in the sense of Definition 4.1. Then we have:*

(i) $0 \le d(r) < \infty$ *holds for all* $r \ge 0$.

(ii) The minimum in the definition of d is attained.

(iii) d is continuous.

(iv) d is monotonically decreasing.

(v) If $d(r) > 0$ holds for all $r \ge 0$, then d is strictly monotonically decreasing.

Obviously an Ω-minimizing solution satisfies a variational inequality in the sense of Definition 3.3 if and only if it satisfies an approximate inequality in the sense of Definition 4.1 and there exists an $r_0 \ge 0$ with $d(r_0) = 0$.

If $u^\dagger$ satisfies an approximate inequality with constant $\bar{\alpha} = \alpha_0$ then it satisfies an approximate inequality with $\bar{\alpha} = \alpha_1$ for all $\alpha_1 \in (0, \alpha_0]$ and with the same other constants. Later we will see that the constant $\bar{\alpha}$ from Definition 4.1 does not appear explicitly in the formulation of convergence rates. So for the sake of plausibility of Definition 4.1 the distance function d should be independent of $\bar{\alpha}$. The next two propositions give some insight into this problem, where the first one is quite technical but is needed as a preparation for the second. The proofs will be postponed to the appendix.

Proposition 4.3. *Let $u^\dagger$ be an Ω-minimizing solution which satisfies an approximate inequality in the sense of Definition 4.1 with $d(r) > 0$ for all $r \ge 0$. Further let $(r_k)_{k\in\mathbb{N}}$ be a sequence in $(0, \infty)$ with $r_k \to \infty$ and let $(u_k)_{k\in\mathbb{N}}$ be a sequence of elements $u_k \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ which realize the minimum in the definition of d, such that $u_k \rightharpoonup \tilde{u}$ holds for some $\tilde{u} \in D$. Then we get*

$$F(\tilde{u}) = v^0, \quad \Omega(\tilde{u}) \le \varrho, \quad \text{and} \quad \xi(\tilde{u} - u^\dagger) = \tfrac{-\beta_1}{1-\beta_1}(\Omega(\tilde{u}) - \Omega(u^\dagger)).$$

Proposition 4.4. *Let $u^\dagger$ be an Ω-minimizing solution which satisfies an approximate inequality in the sense of Definition 4.1 with $d(r) > 0$ for all $r \ge 0$ and let d_α for $\alpha \in (0, \bar{\alpha}]$ be the function defined in analogy to d with $\bar{\alpha}$ replaced by α. If there exists no $u \in U$ with $F(u) = v^0$, $\Omega(u) = \varrho$ and $\xi(u - u^\dagger) = \frac{-\beta_1}{1-\beta_1}(\varrho - \Omega(u^\dagger))$ then the following assertions are true:*

(i) For all $\alpha \in (0, \bar{\alpha}]$ there exists an $r_\alpha \ge 0$, such that $d(r) = d_\alpha(r)$ holds for all $r \ge r_\alpha$.

(ii) For all $\alpha \in (0, \bar{\alpha}]$ there exists an $r_\alpha \ge 0$, such that for all $r \ge r_\alpha$ all elements of $M_{\bar{\alpha}}(\varrho\bar{\alpha})$ which realize the minimum in the definition of $d(r)$ lie in $M_\alpha(\varrho\alpha)$.

Proposition 4.4 (i) states that the distance function d, or in more detail its behaviour at infinity, is independent of $\bar{\alpha}$ if all solutions of $F(u) = v^0$ are regular in the sense that they do not coincide with certain boundary points of the set $\{u \in U : \Omega(u) \le \varrho\}$.

The following Lemma prepares the main theorem of this paper.

Lemma 4.5. *Let $u^\dagger$ be an Ω-minimizing solution which satisfies an approximate inequality in the sense of Definition 4.1 with $0 < \kappa < p$. Further let $\alpha \mapsto \alpha(\delta)$ be a parameter choice fulfilling the condition (3.2) from Proposition 3.2 and let $\bar{\delta}$ be the corresponding constant from that proposition. Then there exist constants $K_1 > 0$, $K_2 > 0$, and $K_3 > 0$, such that*

$$\mathcal{B}_\xi(u^\delta_{\alpha(\delta)}, u^\dagger) \le K_1 \tfrac{\delta^p}{\alpha(\delta)} + K_2 \alpha(\delta)^{\frac{\kappa}{p-\kappa}} r^{\frac{p}{p-\kappa}} + K_3 d(r) \tag{4.2}$$

holds for all $r \ge 0$ and all $\delta \in (0, \bar{\delta}]$.

Proof. For the sake of brevity we write α instead of $\alpha(\delta)$. Proposition 3.2 and the definition of $d(r)$ give us the inequality

$$-\xi(u_\alpha^\delta - u^\dagger) \leq \beta_1 \mathcal{B}_\xi(u_\alpha^\delta, u^\dagger) + r\mathcal{S}(F(u_\alpha^\delta), F(u^\dagger))^\kappa + d(r) \tag{4.3}$$

for all $\delta \in (0, \bar{\delta}]$. Using $\mathcal{S}(v^\delta, v^0) \leq \delta$, $\mathcal{S}(v^0, v^\delta) \leq \delta$, and $\mathcal{T}_\alpha^\delta(u_\alpha^\delta) \leq \mathcal{T}_\alpha^\delta(u^\dagger)$, from this inequality we now get

$$\begin{aligned}
\alpha\mathcal{B}_\xi(u_\alpha^\delta, u^\dagger) &= \mathcal{S}(F(u_\alpha^\delta), v^\delta)^p + \alpha\Omega(u_\alpha^\delta) - \alpha\Omega(u^\dagger) - \alpha\xi(u_\alpha^\delta - u^\dagger) - \mathcal{S}(F(u_\alpha^\delta), v^\delta)^p \\
&\leq \delta^p - \alpha\xi(u_\alpha^\delta - u^\dagger) - \mathcal{S}(F(u_\alpha^\delta), v^\delta)^p \\
&\leq \delta^p + \alpha\beta_1\mathcal{B}_\xi(u_\alpha^\delta, u^\dagger) + \alpha r\mathcal{S}(F(u_\alpha^\delta), F(u^\dagger))^\kappa + \alpha d(r) - \mathcal{S}(F(u_\alpha^\delta), v^\delta)^p \\
&\leq \delta^p + \alpha\beta_1\mathcal{B}_\xi(u_\alpha^\delta, u^\dagger) + \alpha r s^\kappa c_\kappa\big(\mathcal{S}(F(u_\alpha^\delta), v^\delta)^\kappa + \delta^\kappa\big) \\
&\quad\ + \alpha d(r) - \mathcal{S}(F(u_\alpha^\delta), v^\delta)^p,
\end{aligned}$$

where c_κ in analogy to c_p is given by

$$c_\kappa := \begin{cases} 1 & \text{if } 0 < \kappa < 1, \\ 2^{\kappa-1} & \text{if } \kappa \geq 1. \end{cases}$$

Thus we have

$$\begin{aligned}
\mathcal{B}_\xi(u_\alpha^\delta, u^\dagger) \leq \tfrac{1}{\alpha(1-\beta_1)}\Big(&2\delta^p + \alpha c_\kappa r s^\kappa \delta^\kappa - \delta^p + \alpha c_\kappa r s^\kappa \mathcal{S}(F(u_\alpha^\delta), v^\delta)^\kappa \\
&-\mathcal{S}(F(u_\alpha^\delta), v^\delta)^p + \alpha d(r)\Big).
\end{aligned} \tag{4.4}$$

Now we apply the inequality (a modification of Young's inequality)

$$ab - \varepsilon a^{p_1} \leq \frac{b^{p_2}}{(\varepsilon p_1)^{p_2/p_1} p_2}, \tag{4.5}$$

where $a, b \geq 0$, $\varepsilon > 0$, $p_1, p_2 > 1$ and $\frac{1}{p_1} + \frac{1}{p_2} = 1$ have to hold, once with

$$a := \delta^\kappa, \quad b := \alpha c_\kappa r s^\kappa, \quad \varepsilon := 1, \quad p_1 := \tfrac{p}{\kappa}, \quad p_2 := \tfrac{p}{p-\kappa}$$

and once with $\mathcal{S}(F(u_\alpha^\delta), v^\delta)$ instead of δ. We get

$$\begin{aligned}
\mathcal{B}_\xi(u_\alpha^\delta, u^\dagger) &\leq \tfrac{1}{\alpha(1-\beta_1)}\left(2\delta^p + 2(c_\kappa s^\kappa)^{\frac{p}{p-\kappa}}\left(\tfrac{\kappa}{p}\right)^{\frac{\kappa}{p-\kappa}}\tfrac{p-\kappa}{p}\alpha^{\frac{p}{p-\kappa}} r^{\frac{p}{p-\kappa}} + \alpha d(r)\right) \\
&= \tfrac{2}{1-\beta_1}\tfrac{\delta^p}{\alpha} + 2(c_\kappa s^\kappa)^{\frac{p}{p-\kappa}}\left(\tfrac{\kappa}{p}\right)^{\frac{\kappa}{p-\kappa}}\tfrac{p-\kappa}{p(1-\beta_1)}\alpha^{\frac{\kappa}{p-\kappa}} r^{\frac{p}{p-\kappa}} + \tfrac{1}{1-\beta_1} d(r).
\end{aligned}$$

□

Now we can prove the convergence rate theorem from Section 3.

Proof of Theorem 3.4. Because $u^\dagger$ satisfies a variational inequality it also satisfies an approximate inequality with a distance function d for which there exists an $r_0 \geq 0$ with $d(r) = 0$ for all $r \geq r_0$. So the assertion follows immediately from Lemma 4.5 with $r := r_0$. □

Theorem 4.6. *Let $u^\dagger$ be an Ω-minimizing solution which satisfies for some $0 < \kappa < p$ an approximate inequality in the sense of Definition 4.1 with $d(r) > 0$ for all $r \geq 0$. For $r > 0$ we define*

$$\Psi(r) := d(r)^{\frac{p-\kappa}{\kappa}} r^{-\frac{p}{\kappa}} \quad \text{and} \quad \Phi(r) := d(r)^{\frac{1}{\kappa}} r^{-\frac{1}{\kappa}}. \tag{4.6}$$

Further let $\alpha \mapsto \alpha(\delta)$ be a parameter choice with $\delta^p = \alpha(\delta) d\big(\Psi^{-1}(\alpha(\delta))\big)$ for sufficiently small $\delta > 0$. Then we have the convergence rate

$$\mathcal{B}_\xi(u^\delta_{\alpha(\delta)}, u^\dagger) = \mathcal{O}\big(d(\Phi^{-1}(\delta))\big) \quad \text{as } \delta \to 0\,. \tag{4.7}$$

Proof. For the sake of brevity we write $\alpha(\delta)$ instead of α. Because d is strictly monotonically decreasing Ψ and Φ are strictly monotonically decreasing, too. Thus the inverse functions Ψ^{-1} and Φ^{-1} exist and are strictly monotonically decreasing.

Lemma 4.5 with $r := \Psi^{-1}(\alpha)$, i.e.

$$\alpha^{\frac{\kappa}{p-\kappa}} r^{\frac{p}{p-\kappa}} = \Psi(r)^{\frac{\kappa}{p-\kappa}} r^{\frac{p}{p-\kappa}} = d(r),$$

implies

$$\mathcal{B}_\xi(u^\delta_\alpha, u^\dagger) \leq K_1 \tfrac{\delta^p}{\alpha} + (K_2 + K_3) d(\Psi^{-1}(\alpha)) = (K_1 + K_2 + K_3) d(\Psi^{-1}(\alpha))$$

for sufficiently small $\delta \leq \bar{\delta}$ and from

$$\begin{aligned}
\Phi(\Psi^{-1}(\alpha)) &= d(\Psi^{-1}(\alpha))^{\frac{1}{\kappa}} \Psi^{-1}(\alpha)^{-\frac{1}{\kappa}} = \big(\tfrac{\delta^p}{\alpha}\big)^{\frac{1}{\kappa}} \Psi^{-1}(\alpha)^{-\frac{1}{\kappa}} \\
&= \delta^{\frac{p}{\kappa}} \left(\alpha^{\frac{\kappa}{p-\kappa}} \Psi^{-1}(\alpha)^{\frac{p}{p-\kappa}}\right)^{\frac{\kappa-p}{\kappa p}} \alpha^{\frac{1}{p}-\frac{1}{\kappa}} = \delta^{\frac{p}{\kappa}} \left(d(\Psi^{-1}(\alpha))\right)^{\frac{\kappa-p}{\kappa p}} \alpha^{\frac{1}{p}-\frac{1}{\kappa}} \\
&= \delta^{\frac{p}{\kappa}} \big(\tfrac{\delta^p}{\alpha}\big)^{\frac{\kappa-p}{\kappa p}} \alpha^{\frac{1}{p}-\frac{1}{\kappa}} = \delta
\end{aligned}$$

we conclude $\Psi^{-1}(\alpha) = \Phi^{-1}(\delta)$, which proves the assertion. □

Remark 4.7. If instead of d only a strictly monotonically decreasing majorant $\bar{d}$ of d is available then Lemma 4.5 and Theorem 4.6 also hold with d replaced by $\bar{d}$.

The following proposition that uses the notation of Theorem 4.6 gives some further insight into the convergence rates results of this paper. The proof will be postponed to the appendix.

Proposition 4.8. *Under the assumptions of Theorem 4.6 we have the limit condition*

$$\lim_{\delta \to 0} \zeta(\delta) = 0 \qquad \text{for} \qquad \zeta(\delta) := \frac{\delta^\kappa}{d(\Phi^{-1}(\delta))}, \tag{4.8}$$

which is equivalent to $\delta^\kappa = o(d(\Phi^{-1}(\delta)))$ as $\delta \to 0$.

Hence there is a deficit in the convergence rate expressed by the function ζ coming from the fact that a variational inequality in the sense of Definition 3.3 holds only in an approximate manner. The slower the distance function $d(r) > 0$ declines to zero as $r \to \infty$ the greater is the deficit.

Proposition 3.5 told us that under weak assumptions there is an upper bound $q > 0$ for κ in a variational inequality. Now the question arises, whether there is also an upper bound for κ in an approximate inequality. The next proposition does not answer this specific question, but it shows that the maximal rate which can be obtained with the approach of approximate inequalities as described in this paper is bounded by δ^q. The proof will be postponed to the appendix.

Proposition 4.9. *Let the assumptions of Proposition 3.5 be satisfied with some $q > 0$, but let $u^\dagger$ satisfy here an approximate inequality in the sense of Definition 4.1 with $d(r) > 0$ for all $r \geq 0$. Then, with the notation of Theorem 4.6,*

$$\delta^q = \mathcal{O}(d(\Phi^{-1}(\delta))) \quad \text{as } \delta \to 0$$

is valid.

Example 4.10. If we assume that a distance function d has a majorant $\bar{d}$ of the form $\bar{d}(r) = ar^{-b}$ with $a > 0$ and $b > 0$ then the auxiliary functions in Theorem 4.6 become

$$\Psi(r) = a_1 r^{\frac{-b(p-\kappa)-p}{\kappa}} \quad \text{and} \quad \Phi(r) = a_2 r^{\frac{-b-1}{\kappa}} \tag{4.9}$$

with constants $a_1, a_2 > 0$ and thus the theorem provides the convergence rate

$$\mathcal{B}_\xi(u^\delta_{\alpha(\delta)}, u^\dagger) = \mathcal{O}\big(\delta^{\frac{b}{b+1}\kappa}\big) \qquad \text{if } \alpha(\delta) = a_3 \delta^{p-\frac{b}{b+1}\kappa}, \tag{4.10}$$

with a constant $a_3 > 0$.

If an Ω-minimizing solution satisfies a variational inequality, then Theorem 3.4 gives us the corresponding convergence rate. Now an interesting question is whether in this case also an approximate inequality with higher κ is satisfied and, if so, does Theorem 4.6 provide the same rates as Theorem 3.4? The next proposition answers this question. The proof will be postponed to the appendix.

Proposition 4.11. *Let $u^\dagger$ be an Ω-minimizing solution which satisfies a variational inequality in the sense of Definition 3.3 with $0 < \kappa < p$ and let $\mu \in (\kappa, p)$ be such that $u^\dagger$ does not satisfy a variational inequality with κ replaced by μ. Then $u^\dagger$ satisfies an approximate inequality in the sense of Definition 4.1 with κ replaced by μ and the rate obtained from the approximate inequality with μ by Theorem 4.6 is not lower than the rate obtained from the variational inequality with κ by Theorem 3.4.*

Remark 4.12. The idea to introduce approximate variational inequalities in the way it was done in Definition 4.1 arose from the question, whether one can force a variational inequality in the sense of Definition 3.3 to hold by increasing the constant β_2. Thus, we replaced β_2 by r and examined the corresponding distance function d. The question which remains to answer is as follows: what happens if we replace β_2 by any continuous and strictly monotonically increasing function $r \mapsto f(r)$ satisfying $f(0) = 0$ and $f(r) \to \infty$ if $r \to \infty$? Then we would have to examine the corresponding distance function d_f.

Obviously we have $d_f(r) = d(f(r))$ and thus, $d_f(r) \to 0$ as $r \to \infty$ if and only if $d(r) \to 0$ as $r \to \infty$. In other words: An approximate variational inequality with r replaced by $f(r)$ holds if and only if it holds in its original form. One easily checks that after minor modifications all assertions of this section also hold for d replaced by d_f, but at some points we have to look carefully. For example, the auxiliary functions in Theorem 4.6 would read

$$\Psi_f(r) := d_f(r)^{\frac{p-\kappa}{\kappa}} f(r)^{-\frac{p}{\kappa}} \quad \text{and} \quad \Phi_f(r) := d_f(r)^{\frac{1}{\kappa}} f(r)^{-\frac{1}{\kappa}}.$$

Thus, the idea that the convergence rate depends on f seems to be obvious. But noting $\Phi_f(r) = \Phi(f(r))$ and thus $\Phi_f^{-1}(\delta) = f^{-1}(\Phi^{-1}(\delta))$, we see $d_f\big(\Phi_f^{-1}(\delta)\big) = d\big(\Phi^{-1}(\delta)\big)$, i.e. the convergence rate (and analogously the parameter choice) is independent of f. This observation shows that it suffices to consider $f(r) = r$.

5 Source conditions and variational inequalities

An important question which remains to be answered is the interplay of approximate source conditions and approximate inequalities. Note that we discussed the relationships between classical source conditions and variational inequalities in the last paragraph of Section 2 (see also [15]).

At first we want to show that the concept of approximate variational inequalities described in this paper is a generalization of the concept of approximate source conditions in Banach spaces as introduced in [13]. So in this section our focus is on the situation of Example 2.2 and we let U and V be reflexive Banach spaces with τ_U and τ_V describing the corresponding weak topologies. We set $\mathcal{S}(v_1, v_2) := \|v_1 - v_2\|_V$ for $v_1, v_2 \in V$, i.e. we are concerned with the Tikhonov functional

$$\mathcal{T}_\alpha^\delta(u) = \|F(u) - v^\delta\|_V^p + \alpha\Omega(u) \tag{5.1}$$

with $\|v^\delta - v^0\|_V \leq \delta$. In Example 2.2 we mentioned that item (iv) of Assumption 2.1 is satisfied with $s = 1$. We moreover assume that F, $D(F)$ and Ω are chosen such that the other items of Assumption 2.1 are satisfied, too.

For the remaining part of this section let $u^\dagger \in D_B$ (defined below Theorem 2.4) be an Ω-minimizing solution. Because sequentially weakly pre-compact subsets of reflexive Banach spaces are bounded, for all α there is a constant $K_\alpha > 0$ such that

$$\|u - u^\dagger\|_U \leq K_\alpha \quad \text{for all } u \in M_\alpha(\varrho\alpha) \tag{5.2}$$

holds.

We make the following additional assumptions.

Assumption 5.1. Let us assume that:

(i) $D(F)$ is starlike with respect to $u^\dagger$, i.e. for all $u \in D(F)$ there is a $t_0 > 0$, such that $u^\dagger + t(u - u^\dagger) \in D(F)$ holds for all $t \in [0, t_0]$.

(ii) There is a bounded linear operator $F'(u^\dagger) : U \to V$, such that

$$\left\| \frac{F(u^\dagger + t(u - u^\dagger)) - F(u^\dagger)}{t} - F'(u^\dagger)(u - u^\dagger) \right\|_V \to 0 \quad \text{as } t \to +0$$

holds for all $u \in D$.

The convexity of Ω and Assumption 5.1 (i) imply that D is then also starlike with respect to $u^\dagger$. In the sequel we denote by $F'(u^\dagger)^* : V^* \to U^*$ the adjoint operator of $F'(u^\dagger)$, where U^* and V^* are the dual spaces of U and V with respect to the norm topologies. The handling of weakly continuous linear functionals becomes much simpler by the fact that a linear functional on a Banach space is weakly continuous if and only if it is continuous with respect to the norm topology.

We now define what we understand under source conditions.

Definition 5.2. The Ω-minimizing solution $u^\dagger$ *satisfies a source condition* if there exists an element $\xi \in \partial\Omega(u^\dagger)$ with $\xi \in \mathcal{R}(F'(u^\dagger)^*)$. The Ω-minimizing solution $u^\dagger$ *satisfies an approximate source condition* if there exists an element $\xi \in \partial\Omega(u^\dagger)$ with $\xi \in \overline{\mathcal{R}(F'(u^\dagger)^*)}$ and we define the corresponding *distance function* $\tilde{d} : [0, \infty) \to [0, \infty)$ by

$$\tilde{d}(r) := \min\{\|\xi - F'(u^\dagger)^*\eta\|_{U^*} : \eta \in V^*,\ \|\eta\|_{V^*} \leq r\}.$$

As mentioned in [13] the distance function $\tilde{d}$ is well-defined, non-negative, finite and monotonically decreasing. If $u^\dagger$ satisfies a source condition, then it obviously also satisfies an approximate source condition and there is an $r_0 \geq 0$ with $\tilde{d}(r) = 0$ for all $r \geq r_0$. If $u^\dagger$ satisfies an approximate source condition with $\xi \in \overline{\mathcal{R}(F'(u^\dagger)^*)} \setminus \mathcal{R}(F'(u^\dagger)^*)$ then $\tilde{d}(r) > 0$ holds for all $r \geq 0$ and $\tilde{d}$ is strictly monotonically decreasing.

The following definition was used in [13] and [15, 19].

Definition 5.3. Let $c_1, c_2 \geq 0$. The operator F is said to be *nonlinear of degree* (c_1, c_2) with respect to Ω, $u^\dagger$ and $\xi \in \partial\Omega(u^\dagger)$ if there exist constants ϱ fulfilling (3.1), $\bar{\alpha} > 0$, and $K > 0$, such that

$$\|F(u) - F(u^\dagger) - F'(u^\dagger)(u - u^\dagger)\|_V \leq K \|F(u) - F(u^\dagger)\|_V^{c_1} \mathcal{B}_\xi(u, u^\dagger)^{c_2}$$

holds for all $u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$.

The following lemma is an adaption of results in [13]. The proofs of this lemma will be given in the appendix.

Lemma 5.4. *Let the Ω-minimizing solution $u^\dagger$ satisfy an approximate source condition and let F be nonlinear of degree (c_1, c_2) with respect to Ω, $u^\dagger$, and ξ with $c_2 \in [0, 1)$ and $c_1 \in (0, 1 - c_2]$. Then for any $r_0 > 0$ there exists a constant $\beta \geq 0$ such that*

$$-\langle \xi, u - u^\dagger \rangle_{U^*, U} \leq c_2 \mathcal{B}_\xi(u, u^\dagger) + \beta r^{\frac{1}{1-c_2}} \|F(u) - F(u^\dagger)\|_V^{\frac{c_1}{1-c_2}} + K_{\bar{\alpha}} \tilde{d}(r)$$

holds for all $u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ and all $r \geq r_0$.

If $u^\dagger$ satisfies an approximate source condition and F is nonlinear of degree (c_1, c_2) with respect to Ω, $u^\dagger$, and ξ where $c_2 \in [0, 1)$ and $c_1 \in (0, 1 - c_2]$, then the following convergence rates were obtained in [13] (with $\kappa = \frac{c_1}{1-c_2}$ and $p > 1$): In case $\tilde{d}(r) = 0$ for all sufficiently large r the rate

$$\mathcal{B}_\xi(u_{\alpha(\delta)}^\delta, u^\dagger) = \mathcal{O}(\delta^\kappa) \quad \text{as } \delta \to 0 \qquad \text{if } c_1 \delta^{p-\kappa} \leq \alpha(\delta) \leq c_2 \delta^{p-\kappa} \tag{5.3}$$

was shown and in case $\tilde{d}(r) > 0$ for all r the rate

$$\mathcal{B}_\xi(u_{\alpha(\delta)}^\delta, u^\dagger) = \mathcal{O}\big(\tilde{d}(\tilde{\Phi}^{-1}(\delta))\big) \quad \text{as } \delta \to 0 \qquad \text{if } \delta^p = \alpha(\delta) \tilde{d}\big(\tilde{\Psi}^{-1}(\alpha(\delta))\big) \tag{5.4}$$

was stated, where

$$\tilde{\Psi}(r) := \tilde{d}(r)^{\frac{p-\kappa}{\kappa}} r^{-\frac{p}{c_1}} \qquad \text{and} \qquad \tilde{\Phi}(r) := \tilde{d}(r)^{\frac{1}{\kappa}} r^{-\frac{1}{c_1}}.$$

The next theorem compares these rates with the rates obtained in Section 4. The proof will be given in the appendix.

Theorem 5.5. *Let the Ω-minimizing solution $u^\dagger$ satisfy an approximate source condition and let F be nonlinear of degree (c_1, c_2) with respect to Ω, $u^\dagger$ and ξ with $c_2 \in [0, 1)$, $c_1 \in (0, 1 - c_2]$, and $\frac{c_1}{1-c_2} < p$. Then for any $r_0 > 0$ the Ω-minimizing solution $u^\dagger$ satisfies an approximate inequality in the sense of Definition 4.1 with $0 < \kappa = \frac{c_1}{1-c_2} < p$ and $\beta_1 = c_2$. The convergence rate obtained from the approximate variational inequality is not lower than the corresponding rate (5.3) or (5.4) obtained from the approximate source condition.*

If $u^\dagger$ satisfies a source condition then by the proof of Theorem 5.5 $u^\dagger$ also satisfies a variational inequality and Theorem 3.4 and [13, Theorem 3.3] provide the same convergence rate. In analogy we have: If $u^\dagger$ satisfies an approximate source condition then $u^\dagger$ also satisfies an approximate inequality and the rates obtained in Theorem 4.6 are not lower than the rates in [13, Theorem 4.3].

Remark 5.6. In [18] and [13] it has been shown that in the case $c_1 = 0$ and $c_2 = 1$ a source condition $\xi = F'(u^\dagger)^*\eta$ with $K\|\eta\|_{V^*} < 1$ implies a variational inequality with $\kappa = 1$. The converse result that a variational inequality with $\kappa = 1$ implies the source condition is true if F and Ω are Gâteaux differentiable in $u^\dagger$ (see [27, Proposition 3.38]). However, the authors note that convergence rates results are missing up to now in the case $c_1 = 0$ and $c_2 = 1$ when $u^\dagger$ only satisfies an approximate source condition in the sense of Definition 5.2 with $\tilde{d}(r) > 0$ for all $r \geq 0$. Some progress concerning that point will be shown in the forthcoming paper [3].

Now that we know about a basic relationship between approximate source conditions and variational inequalities we conclude this section by repeating from [15] the interplay of classical Hölder type source conditions and variational inequalities in Hilbert spaces. So let U and V be Hilbert spaces and let $F = A$ be a bounded linear operator. Taking the standard Tikhonov functional

$$\mathcal{T}_\alpha^\delta(u) = \|Au - v^\delta\|_V^2 + \alpha\|u\|_U^2$$

the subdifferential of $\Omega = \|\bullet\|_U^2$ at $u \in U$ is the singleton $\{\langle \bullet, 2u\rangle_U\}$ (where $\langle \bullet, \bullet\rangle_U$ denotes the inner product), i.e. we set $\xi = \langle \bullet, 2u^\dagger\rangle_U$, and the corresponding Bregman distance is $\mathcal{B}_\xi(\bullet, u^\dagger) = \|\bullet - u^\dagger\|_U^2$. To legitimize the extended concept of variational inequalities for $\kappa \neq 1$ in [15] the following is stated:

If $u^\dagger$ satisfies a source condition of type $u^\dagger \in \mathcal{R}((A^*A)^{\frac{\mu}{2}})$ with $\mu \in (0, 1)$ then $u^\dagger$ satisfies a variational inequality

$$\langle u^\dagger - u, 2u^\dagger\rangle_U \leq \beta_1\|u - u^\dagger\|_U^2 + \beta_2\|A(u - u^\dagger)\|_V^\kappa \tag{5.5}$$

with $\kappa = \frac{2\mu}{1+\mu}$. For $\mu = 1$ this holds too, as we saw in the Banach space setting above. Because of Proposition 3.5 such a relationship cannot hold for $\mu > 1$. In [15, Proposition 5.7] also the following converse result is formulated: If $u^\dagger$ satisfies a variational inequality (5.5) with exponent κ then it satisfies a source condition of type $u^\dagger \in \mathcal{R}((A^*A)^{\frac{\mu}{2}})$ for all $\mu \in (0, \frac{\kappa}{2-\kappa})$.

6 Conclusions and open questions

The following diagram should help to understand the cross-connections between the different approaches for obtaining convergence rates. In this context, $\Rightarrow$ stands for an implication and $\rightarrow$ stands for "as good or better as". This, however, is only a very rough characterization of the interplay which the reader can find in detail in the corresponding theorems, propositions and remarks.

$$\begin{array}{ccccccc}
\text{rates} & \Leftarrow & \begin{array}{c}\text{source}\\ \text{condition}\end{array} & \Rightarrow & \begin{array}{c}\text{approximate}\\ \text{source condition}\end{array} & \Rightarrow & \text{rates} \\
\updownarrow & & \Downarrow & & \Downarrow & & \uparrow \\
\text{rates} & \Leftarrow & \begin{array}{c}\text{variational}\\ \text{inequality}\end{array} & \Rightarrow & \begin{array}{c}\textbf{approximate}\\ \textbf{variational}\\ \textbf{inequality}\end{array} & \Rightarrow & \text{rates}
\end{array}$$

As we have seen from Proposition 3.5, Remark 3.6, and Proposition 4.9 for the Banach space setting when (2.7) and (5.1) are under consideration the proven convergence rates of Section 3 and Section 4 are because of the occurring limitation $\kappa \le 1$ by construction not faster than $\mathcal{B}_\xi(u^\delta_{\alpha(\delta)}, u^\dagger) = \mathcal{O}(\delta)$ as $\delta \to 0$. Therefore with the technique of variational inequalities (2.14) and also with the corresponding approximate inequalities we are captured in the *low rate world*. A *higher rate world* for that Banach space setting was structured, for example, by the recent papers [12, 22], where under higher source conditions, for $p > 1$, and provided that the space V is smooth enough rates up to $\mathcal{B}_\xi(u^\delta_{\alpha(\delta)}, u^\dagger) = \mathcal{O}(\delta^{4/3})$ can be proven.

In our low rate world the rates are additionally limited by the inequality $\kappa < p$. Up to now the literature considered preferably the case $p > 1$, where this inequality gives no restriction. In the case $0 < p \le 1$, however, for which our approach also applies, this gives a serious restriction. One can interpret the condition $\kappa < p$ then as follows: The exponent $0 < p < 1$ seems to be a *qualification* of the chosen method (similar to the qualification of linear regularization methods, see [7]) which itself defines an upper bound for convergence rates. If the *smoothness* of the solution $u^\dagger$ grows further, i.e. $p < \kappa \le 1$, then the convergence rate does not follow. Note that the boundary situation $0 < \kappa = p \le 1$ shows the so-called *exact penalization* effect studied in [5] for $p = 1$, where the rate $\mathcal{B}_\xi(u^\delta_{\alpha_{\text{fix}}}, u^\dagger) = \mathcal{O}(\delta)$ was proven under the source condition $\xi = F'(u^\dagger)^*\eta$ whenever the regularization parameter $\alpha_{\text{fix}} > 0$ was chosen fixed but small enough. From the proof of Lemma 4.5 yielding the estimate (4.4) we immediately obtain the corresponding rate $\mathcal{B}_\xi(u^\delta_{\alpha_{\text{fix}}}, u^\dagger) = \mathcal{O}(\delta^p)$ whenever a variational inequality is satisfied with exponent $0 < p = \kappa \le 1$ and the regularization parameter $\alpha_{\text{fix}} > 0$ is fixed and small enough. However, it is an open problem to answer the question whether the rates $\mathcal{O}(\delta^{\min\{\kappa,p\}})$ for $0 < p < 1$ can be improved or not.

An advantage of our new approach for the low rate world is the fact that the items (ii) and (iii) of Proposition 3.1 tell us that $\{M_\alpha(\varrho\alpha) : \alpha > 0\}$ in some sense is a family of neighbourhoods of solutions $u^\dagger$ to the equation $F(u) = v^0$. We recall that if a variational inequality holds on a level set $M_{\bar\alpha}(\varrho\bar\alpha)$ then it holds on each level set $M_\alpha(\varrho\alpha)$ with $0 < \alpha < \bar\alpha$. Hence, satisfying a variational inequality means that there exists an arbitrarily small neighbourhood of $u^\dagger$ such that a variational inequality holds on this neighbourhood. Or, in other words, convergence rates depend only on the behaviour of the three functionals $\xi(\bullet - u^\dagger)$, $\mathcal{B}_\xi(\bullet, u^\dagger)$ and $\mathcal{S}(F(\bullet), F(u^\dagger))$ in an arbitrarily small neighborhood of the set of solutions. Looking at the problem from such *functional* point of view this suggests the conjecture that some kind of variational inequality like tools may exist, which is able to integrate higher source conditions and would lead us to the higher rate world. For example, we see that $\mathcal{S}$ and F themselves are not important, only their combination $\mathcal{S}(F(\bullet), F(u^\dagger))$ is of interest. Hence one could ask in this context how the mentioned functionals reflect the combination of source conditions and structure of nonlinearity in case of higher smoothness. This should be forthcoming work.

Acknowledgements

Research was supported by Deutsche Forschungsgemeinschaft (DFG) under Grant HO1454/7-2. The work was partly conducted during the Mini Special Semester on Inverse Problems, May 18 – July 15, 2009, organized by RICAM (Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences) Linz, Austria.

The authors thank Radu Ioan Boţ from Chemnitz University of Technology for reading and commenting on an early version of the manuscript. The two referees made valuable suggestions, which improved the quality of the paper significantly. Therefore, we are very grateful to them.

Appendix

Proof of Proposition 4.2.

(i) Because $\mathcal{T}^0_{\bar\alpha}(u^\dagger) = \bar\alpha\Omega(u^\dagger) \le \varrho\bar\alpha$ we have $u^\dagger \in M_{\bar\alpha}(\varrho\bar\alpha)$ and therefore $d(r) \ge 0$. For $r \ge 0$ we get the estimate

$$d(r) \le C(\xi(u^\dagger), \Omega(u^\dagger)) + \sup_{u\in M_{\bar\alpha}(\varrho\bar\alpha)} |\xi(u)| \tag{A.1}$$

with a constant $C < \infty$ depending on $\xi(u^\dagger)$ and $\Omega(u^\dagger)$ only. Assume there exists a sequence $(u_k)_{k\in\mathbb{N}}$ in $M_{\bar\alpha}(\varrho\bar\alpha)$ with $|\xi(u_k)| \to \infty$. Then from Assumption 2.1 (vii) the existence of a τ_U-convergent subsequence $(u_{k_l})_{l\in\mathbb{N}}$ follows; let $\tilde u \in U$ be its limit. The continuity of ξ implies $|\xi(u_{k_l})| \to |\xi(\tilde u)|$ and therefore the boundedness of the sequence $(|\xi(u_{k_l})|)$. This contradicts $|\xi(u_k)| \to \infty$. Thus

$$\sup_{u\in M_{\bar\alpha}(\varrho\bar\alpha)} |\xi(u)| < \infty,$$

i.e. $d(r) < \infty$.

(ii) We define $g_r : U \to \mathbb{R} \cup \{+\infty\}$ by

$$g_r(u) := \xi(u - u^\dagger) + \beta_1 \mathcal{B}_\xi(u, u^\dagger) + r\mathcal{S}(F(u), F(u^\dagger))^\kappa. \tag{A.2}$$

The continuity of ξ and F and the lower semi-continuity of Ω and $\mathcal{S}$ together imply the lower semi-continuity of g_r. Now let $(u_k)_{k\in\mathbb{N}}$ be a sequence in $M_{\bar\alpha}(\varrho\bar\alpha)$ satisfying $g_r(u_k) \to \inf_{u\in M_{\bar\alpha}(\varrho\bar\alpha)} g_r(u)$. Then there exists a τ_U-convergent subsequence $(u_{k_l})_{l\in\mathbb{N}}$ with limit $\tilde u \in U$, especially we get $\tilde u \in M_{\bar\alpha}(\varrho\bar\alpha)$ (because $\mathcal{T}^0_{\bar\alpha}$ is lower semi-continuous), and

$$g_r(\tilde u) \le \liminf_{l\to\infty} g_r(u_{k_l}) = \lim_{l\to\infty} g_r(u_{k_l}) = \inf_{u\in M_{\bar\alpha}(\varrho\bar\alpha)} g_r(u)$$

holds. Thus, $g_r(\tilde u) = \inf_{u\in M_{\bar\alpha}(\varrho\bar\alpha)} g_r(u)$.

(iii) For $r \ge 0$ let g_r be defined as in the proof of item (ii) and let $u_r \in M_{\bar\alpha}(\varrho\bar\alpha)$ be a minimizer of g_r. Further, define g_t and u_t analogously to g_r and u_r. Then for all $r, t \ge 0$ we have

$$d(r) - d(t) = \min g_t - \min g_r \le g_t(u_r) - g_r(u_r) = (t - r)\mathcal{S}(F(u_r), F(u^\dagger))^\kappa$$

and

$$-(d(r) - d(t)) = \min g_r - \min g_t \le g_r(u_t) - g_t(u_t) = (r - t)\mathcal{S}(F(u_t), F(u^\dagger))^\kappa,$$

and together with $u_r, u_t \in M_{\bar\alpha}(\varrho\bar\alpha)$ this implies

$$|d(r) - d(t)| \le (\varrho\bar\alpha)^{\frac{\kappa}{p}} |r - t|, \tag{A.3}$$

i.e. d is continuous.

(iv) The assertion follows directly from the definition of d.

(v) We assume that there is an $r \geq 0$ for which g_r (set as in the proof of (ii)) has a minimizer $\tilde{u} \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ satisfying $F(\tilde{u}) = v^0$. Then for each $t \geq 0$ we get

$$\begin{aligned}
&\xi(\tilde{u} - u^\dagger) + \beta_1 \mathcal{B}_\xi(\tilde{u}, u^\dagger) \\
&\quad = \xi(\tilde{u} - u^\dagger) + \beta_1 \mathcal{B}_\xi(\tilde{u}, u^\dagger) + t\mathcal{S}(F(\tilde{u}), F(u^\dagger))^\kappa \\
&\quad \geq \min_{u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})} \big(\xi(u - u^\dagger) + \beta_1 \mathcal{B}_\xi(u, u^\dagger) + t\mathcal{S}(F(u), F(u^\dagger))^\kappa\big) \\
&\quad = -d(t)
\end{aligned}$$

and thus $d(t) \to 0$ as $t \to \infty$ implies $\xi(\tilde{u} - u^\dagger) + \beta_1 \mathcal{B}_\xi(\tilde{u}, u^\dagger) \geq 0$. But this contradicts

$$\begin{aligned}
&\xi(\tilde{u} - u^\dagger) + \beta_1 \mathcal{B}_\xi(\tilde{u}, u^\dagger) \\
&\quad = \xi(\tilde{u} - u^\dagger) + \beta_1 \mathcal{B}_\xi(\tilde{u}, u^\dagger) + r\mathcal{S}(F(\tilde{u}), F(u^\dagger))^\kappa \\
&\quad = \min_{u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})} \big(\xi(u - u^\dagger) + \beta_1 \mathcal{B}_\xi(u, u^\dagger) + r\mathcal{S}(F(u), F(u^\dagger))^\kappa\big) \\
&\quad = -d(r) < 0.
\end{aligned}$$

So for each $r \geq 0$ each minimizer $\tilde{u} \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ of g_r satisfies the inequality $\mathcal{S}(F(\tilde{u}), F(u^\dagger)) > 0$. Now for $0 \leq t < r$ we have (g_t defined analogously to g_r)

$$\begin{aligned}
d(r) &= -\min_{u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})} g_r(u) = -g_r(\tilde{u}) = -g_t(\tilde{u}) - (r-t)\mathcal{S}(F(\tilde{u}), F(u^\dagger))^\kappa \\
&< -g_t(\tilde{u}) \leq -\min_{u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})} g_t(u) = d(t),
\end{aligned}$$

i.e. d is strictly monotonically decreasing. □

Proof of Proposition 4.3. The definitions of u_k and $d(r_k)$ imply

$$-r_k \mathcal{S}(F(u_k), F(u^\dagger))^\kappa = d(r_k) + \xi(u_k - u^\dagger) + \beta_1 \mathcal{B}_\xi(u_k, u^\dagger).$$

From the continuity of ξ and the lower semi-continuity of Ω for $\varepsilon > 0$ and sufficiently large $k \in \mathbb{N}$ we get

$$-r_k \mathcal{S}(F(u_k), F(u^\dagger))^\kappa \geq d(r_k) + \xi(\tilde{u} - u^\dagger) + \beta_1 \mathcal{B}_\xi(\tilde{u}, u^\dagger) - \varepsilon$$

and therefore

$$\mathcal{S}(F(u_k), F(u^\dagger))^\kappa \leq \tfrac{-1}{r_k}\big(d(r_k) + \xi(\tilde{u} - u^\dagger) + \beta_1 \mathcal{B}_\xi(\tilde{u}, u^\dagger) - \varepsilon\big).$$

Passage to the limit $k \to \infty$ gives $\mathcal{S}(F(u_k), F(u^\dagger))^\kappa \to 0$ and with Assumption 2.1 (iv)(d) this implies $F(u_k) \rightharpoonup v^0$. On the other hand Assumption 2.1 (i) implies $F(u_k) \rightharpoonup F(\tilde{u})$ and therefore $F(\tilde{u}) = v^0$.

The second assertion follows from

$$\Omega(\tilde{u}) \leq \liminf_{k \to \infty} \Omega(u_k) \leq \liminf_{k \to \infty} \tfrac{1}{\bar{\alpha}} \mathcal{T}^0_{\bar{\alpha}}(u_k) \leq \varrho.$$

To prove the third and last assertion we first observe

$$\begin{aligned}
-\xi(\tilde{u} - u^\dagger) - \beta_1 \mathcal{B}_\xi(\tilde{u}, u^\dagger) &= -\xi(\tilde{u} - u^\dagger) - \beta_1 \mathcal{B}_\xi(\tilde{u}, u^\dagger) - r_k \mathcal{S}(F(\tilde{u}), F(u^\dagger))^\kappa \\
&\leq d(r_k) \to 0,
\end{aligned}$$

which gives

$$-\xi(\tilde{u}-u^\dagger)-\beta_1\mathcal{B}_\xi(\tilde{u},u^\dagger)\le 0. \tag{A.4}$$

For $\varepsilon>0$ and $k\in\mathbb{N}$ sufficiently large the continuity of ξ and the lower semicontinuity of Ω imply

$$\begin{aligned}0\ge -r_k\mathcal{S}(F(u_k),F(u^\dagger))^\kappa &= d(r_k)+\xi(u_k-u^\dagger)+\beta_1\mathcal{B}_\xi(u_k,u^\dagger)\\ &\ge d(r_k)+\xi(\tilde{u}-u^\dagger)+\beta_1\mathcal{B}_\xi(\tilde{u},u^\dagger)-\varepsilon.\end{aligned}$$

By passage to the limit $k\to\infty$ we get $\xi(\tilde{u}-u^\dagger)+\beta_1\mathcal{B}_\xi(\tilde{u},u^\dagger)\le\varepsilon$ and from the arbitrariness of ε we obtain

$$-\xi(\tilde{u}-u^\dagger)-\beta_1\mathcal{B}_\xi(\tilde{u},u^\dagger)\ge 0. \tag{A.5}$$

Inequalities (A.4) and (A.5) together imply

$$-\xi(\tilde{u}-u^\dagger)=\beta_1\mathcal{B}_\xi(\tilde{u},u^\dagger)$$

and substituting the Bregman distance by its definition gives the assertion. □

Proof of Proposition 4.4. Assertion (i) is a direct consequence of (ii). We give an indirect proof of assertion (ii). We assume that there exist an $\alpha\in(0,\bar{\alpha}]$ and a sequence $(r_k)_{k\in\mathbb{N}}$ in $(0,\infty)$ with $r_k\to\infty$, such that for each r_k there exists an element $u_k\in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ which realizes the minimum in the definition of $d(r_k)$ and which satisfies $u_k\notin M_\alpha(\varrho\alpha)$. Because of Assumption 2.1 (vii) and the lower semi-continuity of $\mathcal{T}^0_{\bar{\alpha}}$ the sequence $(u_k)_{k\in\mathbb{N}}$ has a convergent subsequence, which we again denote by $(u_k)_{k\in\mathbb{N}}$, with limit $\tilde{u}\in M_{\bar{\alpha}}(\varrho\bar{\alpha})$.

Proposition 4.3 now implies

$$F(\tilde{u})=v^0,\quad \Omega(\tilde{u})\le\varrho\quad\text{and}\quad \xi(\tilde{u}-u^\dagger)=\tfrac{-\beta_1}{1-\beta_1}(\Omega(\tilde{u})-\Omega(u^\dagger)). \tag{A.6}$$

From $u_k\notin M_\alpha(\varrho\alpha)$ in addition we get

$$\Omega(u_k)>\varrho-\tfrac{1}{\alpha}\mathcal{S}(F(u_k),v^0)^p$$

for all $k\in\mathbb{N}$ and thus $\mathcal{S}(F(u_k),F(u^\dagger))\to 0$ (cf. proof of Proposition 4.3) implies $\Omega(u_k)>\varrho-\varepsilon$ for $\varepsilon>0$ and sufficiently large $k\in\mathbb{N}$. Together with $\Omega(u_k)\le\varrho$ this gives $\Omega(u_k)\to\varrho$. Therefore from

$$0\ge -r_k\mathcal{S}(F(u_k),F(u^\dagger))^\kappa=d(r_k)+\xi(u_k-u^\dagger)+\beta_1\big(\Omega(u_k)-\Omega(u^\dagger)-\xi(u_k-u^\dagger)\big)$$

by passage to the limit we conclude

$$0\ge(1-\beta_1)\xi(\tilde{u}-u^\dagger)+\beta_1(\varrho-\Omega(u^\dagger))$$

and together with (A.6) we get

$$\tfrac{-\beta_1}{1-\beta_1}(\Omega(\tilde{u})-\Omega(u^\dagger))=\xi(\tilde{u}-u^\dagger)\le\tfrac{-\beta_1}{1-\beta_1}(\varrho-\Omega(u^\dagger))\le\tfrac{-\beta_1}{1-\beta_1}(\Omega(\tilde{u})-\Omega(u^\dagger)),$$

i.e. especially $\Omega(\tilde{u})=\varrho$ is true. Substituting this equality into (A.6) gives a contradiction to the assumptions of the proposition. □

Proof of Proposition 4.8. With $r:=\Phi^{-1}(\delta)$, i.e. $\delta=\Phi(r)$, we have

$$\zeta(\delta)=\frac{\delta^\kappa}{d(\Phi^{-1}(\delta))}=\frac{\Phi(r)^\kappa}{d(r)}=\frac{d(r)r^{-1}}{d(r)}=\frac{1}{r}=\frac{1}{\Phi^{-1}(\delta)}. \tag{A.7}$$

From $\Phi^{-1}(\delta)\to\infty$ as $\delta\to 0$ we conclude $\big(\Phi^{-1}(\delta)\big)^{-1}\to 0$ as $\delta\to 0$. Therefore we have (4.8). □

Proof of Proposition 4.9. Assume that the assertion is not true, i.e. we have

$$\frac{d(\Phi^{-1}(\delta))}{\delta^q} \to 0 \quad \text{as } \delta \to 0\,. \tag{A.8}$$

As in the proof of Proposition 3.5, but starting with the inequality

$$-\xi(\tilde{u}-u^\dagger) \le \beta_1 \mathcal{B}_\xi(\tilde{u},u^\dagger) + r\mathcal{S}(F(\tilde{u}),F(u^\dagger))^\kappa + d(r)$$

for $\tilde{u} = u^\dagger + tu \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ and $r \ge 0$ instead of (3.3), for $t \in (0,t_0]$ and $r \ge 0$ we get

$$-\xi(u) \le \beta_1 \left(\tfrac{\Omega(u^\dagger+tu)-\Omega(u^\dagger)}{t} - \xi(u)\right) + r\left(\tfrac{\mathcal{S}(F(u^\dagger+tu),F(u^\dagger))^q}{t}\right)^{\frac{\kappa}{q}} t^{\frac{\kappa}{q}-1} + \tfrac{d(r)}{t}.$$

Now we choose $r(t) := \Phi^{-1}(t^{\frac{1}{q}})$, i.e. we have $t = \Phi(r(t))^q$. On the one hand this (together with (A.8)) implies

$$\frac{d(r(t))}{t} = \frac{d(\Phi^{-1}(t^{\frac{1}{q}}))}{t} \to 0 \quad \text{as } t \to +0$$

and on the other hand this implies

$$r(t)t^{\frac{\kappa}{q}-1} = \frac{r(t)\Phi(r(t))^\kappa}{t} = \frac{d(r(t))}{t} \to 0 \quad \text{as } t \to +0.$$

So all terms of the above inequality tend to zero as $t \to +0$ and thus $\xi(u) \ge 0$ is valid, which is a contradiction to the assumption $\xi(u) < 0$. $\square$

Proof of Proposition 4.11. Let ϱ, $\bar{\alpha}$, β_1, and β_2 be the constants from the variational inequality satisfied by $u^\dagger$. For all $u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ and all $r > 0$ then

$$\begin{aligned} &-\xi(u-u^\dagger) - \beta_1\mathcal{B}_\xi(u,u^\dagger) - r\mathcal{S}(F(u),F(u^\dagger))^\mu \\ &\qquad \le \beta_2\mathcal{S}(F(u),F(u^\dagger))^\kappa - r\mathcal{S}(F(u),F(u^\dagger))^\mu \end{aligned}$$

follows, and the Young-type inequality (4.5) with

$$\begin{aligned} &a := \left(r\mathcal{S}(F(u),F(u^\dagger))^\mu\right)^{\frac{\kappa}{\mu}}, \quad b := \beta_2 r^{-\frac{\kappa}{\mu}}, \\ &\varepsilon := 1, \quad p_1 := \frac{\mu}{\kappa}, \quad p_2 := \frac{\mu}{\mu-\kappa} \end{aligned}$$

implies

$$\begin{aligned} d(r) &\le \max_{M_{\bar{\alpha}}(\varrho\bar{\alpha})} \left(-\xi(u-u^\dagger) - \beta_1\mathcal{B}_\xi(u,u^\dagger) - r\mathcal{S}(F(u),F(u^\dagger))^\mu\right) \\ &\le \left(\tfrac{\kappa}{\mu}\right)^{\frac{\kappa}{\mu-\kappa}} \tfrac{\mu-\kappa}{\mu} \beta_2^{\frac{\mu}{\mu-\kappa}} r^{-\frac{\kappa}{\mu-\kappa}}, \end{aligned}$$

i.e. $u^\dagger$ satisfies an approximate inequality and the corresponding distance function d has a majorant $\bar{d}$ of the form $\bar{d}(r) = ar^{-b}$ with $a > 0$ and $b = \frac{\kappa}{\mu-\kappa}$.

Equation (4.10) with κ replaced by μ thus gives

$$\mathcal{B}_\xi(u^\delta_{\alpha(\delta)},u^\dagger) = \mathcal{O}(\delta^{\frac{b}{b+1}\mu}) = \mathcal{O}(\delta^{\frac{\kappa}{\mu}\mu}) = \mathcal{O}(\delta^\kappa) \tag{A.9}$$

for the parameter choice $\alpha(\delta) = c\delta^{p-\frac{b}{b+1}\mu} = c\delta^{p-\frac{\kappa}{\mu}\mu} = c\delta^{p-\kappa}$ with a constant $c > 0$. This is exactly the convergence rate which is stated by Theorem 3.4. $\square$

Proof of Lemma 5.4. For $r \geq 0$ let $\eta_r \in V^*$ be an element for which the minimum in the definition of $\tilde{d}(r)$ is attained. Then for $u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})$ we have

$$\begin{aligned}
&-\langle \xi, u - u^\dagger \rangle_{U^*,U} \\
&\quad \leq \left| \langle F'(u^\dagger)^* \eta_r + \xi - F'(u^\dagger)^* \eta_r, u - u^\dagger \rangle_{U^*,U} \right| \\
&\quad = \left| \langle \eta_r, F'(u^\dagger)(u - u^\dagger) \rangle_{V^*,V} + \langle \xi - F'(u^\dagger)^* \eta_r, u - u^\dagger \rangle_{U^*,U} \right| \\
&\quad \leq \|\eta_r\|_{V^*} \|F'(u^\dagger)(u - u^\dagger)\|_V + \|\xi - F'(u^\dagger)^* \eta_r\|_{U^*} \|u - u^\dagger\|_U \\
&\quad \leq r \|F(u) - F(u^\dagger) - F'(u^\dagger)(u - u^\dagger) + F(u^\dagger) - F(u)\|_V + K_{\bar{\alpha}} \tilde{d}(r) \\
&\quad \leq K r \|F(u) - F(u^\dagger)\|_V^{c_1} \mathcal{B}_\xi(u, u^\dagger)^{c_2} + r \|F(u) - F(u^\dagger)\|_V + K_{\bar{\alpha}} \tilde{d}(r).
\end{aligned}$$

Now we have to distinguish between two cases:

- Case $c_2 = 0$. We get

$$\begin{aligned}
&-\langle \xi, u - u^\dagger \rangle_{U^*,U} \\
&\quad \leq K r \|F(u) - F(u^\dagger)\|_V^{c_1} + r \|F(u) - F(u^\dagger)\|_V + K_{\bar{\alpha}} \tilde{d}(r) \\
&\quad = \left(K r + r \|F(u) - F(u^\dagger)\|_V^{1-c_1} \right) \|F(u) - F(u^\dagger)\|_V^{c_1} + K_{\bar{\alpha}} \tilde{d}(r) \\
&\quad \leq \left(K + (\varrho\bar{\alpha})^{\frac{1-c_1}{p}} \right) r \|F(u) - F(u^\dagger)\|_V^{c_1} + K_{\bar{\alpha}} \tilde{d}(r).
\end{aligned}$$

- Case $c_2 \in (0,1)$. We apply the inequality

$$ab \leq \frac{a^{p_1}}{p_1} + \frac{b^{p_2}}{p_2} \quad \text{for} \quad a, b \geq 0, \quad \frac{1}{p_1} + \frac{1}{p_2} = 1, \quad p_1, p_2 > 1$$

  with

$$a := \mathcal{B}_\xi(u, u^\dagger)^{c_2}, \; b := K r \|F(u) - F(u^\dagger)\|_V^{c_1}, \; p_1 := \frac{1}{c_2}, \; p_2 := \frac{1}{1 - c_2}$$

  and get

$$\begin{aligned}
&-\langle \xi, u - u^\dagger \rangle_{U^*,U} \\
&\quad \leq K r \|F(u) - F(u^\dagger)\|_V^{c_1} \mathcal{B}_\xi(u, u^\dagger)^{c_2} + r \|F(u) - F(u^\dagger)\|_V + K_{\bar{\alpha}} \tilde{d}(r) \\
&\quad \leq c_2 \mathcal{B}_\xi(u, u^\dagger) + (1 - c_2) K^{\frac{1}{1-c_2}} r^{\frac{1}{1-c_2}} \|F(u) - F(u^\dagger)\|_V^{\frac{c_1}{1-c_2}} \\
&\qquad + r \|F(u) - F(u^\dagger)\|_V + K_{\bar{\alpha}} \tilde{d}(r) \\
&\quad = c_2 \mathcal{B}_\xi(u, u^\dagger) + K_{\bar{\alpha}} \tilde{d}(r) \\
&\qquad + \left((1 - c_2) K^{\frac{1}{1-c_2}} r^{\frac{1}{1-c_2}} + r \|F(u) - F(u^\dagger)\|_V^{\frac{1-c_1-c_2}{1-c_2}} \right) \|F(u) - F(u^\dagger)\|_V^{\frac{c_1}{1-c_2}} \\
&\quad \leq c_2 \mathcal{B}_\xi(u, u^\dagger) + K_{\bar{\alpha}} \tilde{d}(r) \\
&\qquad + \left((1 - c_2) K^{\frac{1}{1-c_2}} + (\varrho\bar{\alpha})^{\frac{1-c_1-c_2}{p(1-c_2)}} r_0^{\frac{-c_2}{1-c_2}} \right) r^{\frac{1}{1-c_2}} \|F(u) - F(u^\dagger)\|_V^{\frac{c_1}{1-c_2}}
\end{aligned}$$

  for $r \geq r_0$. □

Proof of Theorem 5.5. For $f(r) := \beta r^{\frac{1}{1-c_2}}$ by Lemma 5.4 we have

$$\begin{aligned}
d_f(r) := d(f(r)) &= \max_{u \in M_{\bar{\alpha}}(\varrho\bar{\alpha})} \left(-\langle \xi, u - u^\dagger \rangle_{U^*,U} - c_2 \mathcal{B}_\xi(u, u^\dagger) - f(r) \|F(u) - F(u^\dagger)\|_V^{\frac{c_1}{1-c_2}} \right) \\
&\leq K_{\bar{\alpha}} \tilde{d}(r)
\end{aligned}$$

for all $r \geq r_0$, i.e. $d_f(r) \to 0$ as $r \to \infty$ and therefore an approximate variational inequality is satisfied (cf. Remark 4.12). For the remaining part of the proof we assume $d(r) > 0$ and $\tilde{d}(r) > 0$ for all $r \geq 0$. Setting

$$\tilde{\Psi}(r) := \tilde{d}(r)^{\frac{p-\kappa}{\kappa}} r^{-\frac{p}{c_1}} \qquad \text{and} \qquad \tilde{\Phi}(r) := \tilde{d}(r)^{\frac{1}{\kappa}} r^{-\frac{1}{c_1}}$$

as in [13], and setting Ψ_f and Φ_f as in Remark 4.12, the same remark in connection with Theorem 4.6 provides the convergence rate $\mathcal{O}\big(d_f(\Phi_f^{-1}(\delta))\big)$ if $\delta^p = \alpha(\delta) d_f\big(\Psi_f^{-1}(\alpha(\delta))\big)$, and [13, Theorem 4.3] gives the rate $\mathcal{O}\big(\tilde{d}(\tilde{\Phi}^{-1}(\delta))\big)$ if $\delta^p = \alpha(\delta)\tilde{d}\big(\tilde{\Psi}^{-1}(\alpha(\delta))\big)$. Again by Remark 4.12, $d_f(\Psi_f^{-1}(\delta)) = d(\Psi^{-1}(\delta))$ and $d_f(\Phi_f^{-1}(\delta)) = d(\Phi^{-1}(\delta))$, where Ψ and Φ are defined as in Theorem 4.6.

Thus, it remains to show $d_f(\Phi_f^{-1}(\delta)) = \mathcal{O}\big(\tilde{d}(\tilde{\Phi}^{-1}(\delta))\big)$ as $\delta \to 0$. First we note

$$\Phi_f(r) \leq \beta^{-\frac{1}{\kappa}} K_{\bar{\alpha}}^{\frac{1}{\kappa}} \tilde{\Phi}(r) \qquad \text{for } r \geq r_0$$

and therefore, by calculating the inverse function of each side,

$$\Phi_f^{-1}(\delta) \leq \tilde{\Phi}^{-1}\Big(\big(\tfrac{\beta}{K_{\bar{\alpha}}}\big)^{\frac{1}{\kappa}}\delta\Big) \qquad \text{for } \delta \leq \Phi_f(r_0). \tag{A.10}$$

Setting $r = \Phi_f^{-1}(\delta)$ we get

$$d_f(\Phi_f^{-1}(\delta)) = \delta^\kappa \frac{d_f(\Phi_f^{-1}(\delta))}{\delta^\kappa} = \Phi_f(r)^\kappa \frac{d_f(r)}{d_f(r) f(r)^{-1}} = \Phi_f(r)^\kappa f(r) = \delta^\kappa f(\Phi_f^{-1}(\delta)),$$

and setting $r = \tilde{\Phi}^{-1}(\delta)$ we get

$$\beta\tilde{d}(\tilde{\Phi}^{-1}(\delta)) = \beta\delta^\kappa \frac{\tilde{d}(\tilde{\Phi}^{-1}(\delta))}{\delta^\kappa} = \beta\tilde{\Phi}(r)^\kappa \frac{\tilde{d}(r)}{\tilde{d}(r)\big(\frac{1}{\beta} f(r)\big)^{-1}} = \tilde{\Phi}(r)^\kappa f(r) = \delta^\kappa f(\tilde{\Phi}^{-1}(\delta)).$$

Without loss of generality we may assume $K_{\bar{\alpha}} \geq \beta$. Thus, with the help of inequality (A.10) we arrive at

$$d_f(\Phi_f^{-1}(\delta)) = \delta^\kappa f(\Phi_f^{-1}(\delta)) \leq \delta^\kappa f\Big(\tilde{\Phi}^{-1}\Big(\big(\tfrac{\beta}{K_{\bar{\alpha}}}\big)^{\frac{1}{\kappa}}\delta\Big)\Big) = \beta\tilde{d}\Big(\tilde{\Phi}^{-1}\Big(\big(\tfrac{\beta}{K_{\bar{\alpha}}}\big)^{\frac{1}{\kappa}}\delta\Big)\Big) \leq \beta\tilde{d}(\tilde{\Phi}^{-1}(\delta))$$

for all $\delta \leq \Phi_f(r_0)$. $\square$

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures.* Birkhäuser Verlag Basel, Boston, Berlin, 2008.

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

[3] R. I. Boţ and B. Hofmann. An extension of the variational inequality approach for nonlinear ill-posed problems. *Journal of Integral Equations and Applications*, 22, 2010. to appear.

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

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

[6] D. Düvelmeyer, B. Hofmann, and M. Yamamoto. Range inclusions and approximate source conditions with general benchmark functions. *Numer. Funct. Anal. Optim.*, 28(11-12):1245–1261, 2007.

[7] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[8] J. Geissler. *Studies on convergence rates for the Tikhonov regularization with general residual functionals*. Diploma theses (in German), Chemnitz University of Technology, Faculty of Mathematics, 2009. Diploma thesis (in German).

[9] C. R. Givens and R. M. Shortt. A class of Wasserstein metrics for probability distributions. *Michigan Mathematical Journal*, 31(2):231–240, 1984.

[10] M. Grasmair, M. Haltmeier, and O. Scherzer. Sparse regularization with l^q penalty term. *Inverse Problems*, 24(5):055020, 13pp, 2009.

[11] T. Hein. Convergence rates for regularization of ill-posed problems in Banach spaces by approximate source conditions. *Inverse Problems*, 24(4):045007, 10, 2008.

[12] T. Hein. Tikhonov regularization in Banach spaces – improved convergence rates results. *Inverse Problems*, 25(3):035002, 18, 2009.

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

[14] B. Hofmann. Approximate source conditions in Tikhonov-Phillips regularization and consequences for inverse problems with multiplication operators. *Mathematical Methods in the Applied Sciences*, 29:351–371, 2006.

[15] B. Hofmann. Interplay of source conditions and variational inequalities for nonlinear ill-posed problems. Preprint 2009-6, Preprint Series of the Faculty of Mathematics, Chemnitz University of Technology, 2009. www.mathematik.tu-chemnitz.de/preprint/quellen/2009/PREPRINT_06.pdf.

[16] B. Hofmann. A note on convergence rates for variational regularization with non-convex residual term. Preprint 2009-11, Preprint Series of the Faculty of Mathematics, Chemnitz University of Technology, 2009. www.mathematik.tu-chemnitz.de/preprint/quellen/2009/PREPRINT_11.pdf.

[17] B. Hofmann, D. Düvelmeyer, and K. Krumbiegel. Approximate source conditions in Tikhonov regularization—new analytical results and some numerical studies. *Math. Model. Anal.*, 11(1):41–56, 2006.

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

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

[20] B. Kaltenbacher, F. Schöpfer, and T. Schuster. Iterative methods for nonlinear ill-posed problems in Banach spaces: convergence and applications to parameter identification problems. *Inverse Problems*, 25(6):065003, 19, 2009.

[21] D. A. Lorenz and D. Trede. Optimal convergence rates for Tikhonov regularization in Besov scales. *Inverse Problems*, 24(5):055010, 14, 2008.

[22] A. Neubauer. On enhanced convergence rates for Tikhonov regularization of nonlinear ill-posed problems in Banach spaces. *Inverse Problems*, 25(6):065009, 10, 2009.

[23] C. Pöschl. *Tikhonov Regularization with General Residual Term.* PhD thesis, University of Innsbruck, Austria, 2008.

[24] R. Ramlau. Regularization properties of Tikhonov regularization with sparsity constraints. *Electron. Trans. Numer. Anal.*, 30:54–74, 2008.

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

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

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

[28] F. Schöpfer, A. K. Louis, and T. Schuster. Nonlinear iterative methods for linear ill-posed problems in Banach spaces. *Inverse Problems*, 22(1):311–329, 2006.