Solution smoothness of ill-posed equations in Hilbert spaces: four concepts and their cross connections

Jens Flemming*

January 14, 2011

Key words: ill-posed problem, source condition, variational inequality
MSC: 47A52, 65J20, 65F22

Abstract

Numerical solution of ill-posed operator equations requires regularization techniques. The convergence of regularized solutions to the exact solution usally can be guaranteed, but to obtain also estimates for the speed of convergence one has to exploit some kind of smoothness of the exact solution. We consider four such smoothness concepts in a Hilbert space setting: source conditions, approximate source conditions, variational inequalities, and approximate variational inequalities. Besides some new auxiliary results on variational inequalities the equivalence of the last three concepts is shown. In addition, it turns out that the classical concept of source conditions and the modern concept of variational inequalities are connected via Fenchel duality.

1 Introduction

The most challenging aspect in the analysis of regularization methods for solving ill-posed operator equations is the derivation of convergence rates. Such convergence rates describe the speed with which the regularized solutions approximate the exact solution of an equation if the noise level of the data goes to zero. In general this convergence can be arbitrarily slow

*Chemnitz University of Technology, Department of Mathematics, 09107 Chemnitz, Germany, email: jens.flemming@mathematik.tu-chemnitz.de. Research supported by the DFG under grant HO 1454/8-1

for ill-posed problems (see [6, Section 3.2]). The classical concept of *source conditions* as sufficient conditions for convergence rates has several disadvantages. The major drawback in recent years was their strong connection to Hilbert spaces; apart from very few special cases, source conditions are based on spectral theory. Another problem has been pointed out in [17, 21]: convergence rates obtained from source conditions cannot be optimal.

Therefore different new concepts have been developed in recent years. The first of interest to us was the idea of *approximate source conditions* introduced in [14]. Later so called *variational inequalities* were introduced to obtain convergence rates for nonlinear and nonsmooth operators in Banach spaces (see [15, 26]). Variants of such inequalities are also used in [3]. Extensions to Tikhonov regularization with general fitting functionals are given in [7,8,10,25]. An extension of variational inequalities are *approximate variational inequalities* introduced in [8].

This article tries to bring more structure into the growing set of smoothness assumptions. Especially the role and interpretation of variational inequalities has to be clarified.

We restrict the study to the classical Hilbert space setting for linear ill-posed problems: Let X and Y be Hilbert spaces and let $A : X \to Y$ be an injective bounded linear operator with non-closed range $\mathcal{R}(A) \neq \overline{\mathcal{R}(A)}$. The injectivity allows for a more clear exposition and the non-closed range is an equivalent formulation of the assumption that the inverse of A is unbounded. We aim to solve the ill-posed equation

$$Ax = y^0, \quad x \in X, \tag{1.1}$$

for $y^0 \in \mathcal{R}(A)$ approximately by minimizing the Tikhonov functional

$$T_\alpha^\delta(x) := \|Ax - y^\delta\|^2 + \alpha\|x\|^2 \tag{1.2}$$

over $x \in X$. Here, $\alpha > 0$ is the regularization parameter controlling the trade-off between stability and quality of approximation, and $y^\delta \in Y$ is some noisy version of the exact right-hand side y^0. The noise level $\delta > 0$ bounds the noise, that is, we assume $\|y^\delta - y\| \leq \delta$. We could also consider general linear regularization methods in Hilbert spaces (see [6, Chapter 4]), but this would not give additional insights and would make the exposition more complex.

We denote the exact solution of (1.1) by $x^\dagger \in X$, that is, $Ax^\dagger = y^0$. The unique minimizer of T_α^δ will be denoted by $x_\alpha^\delta \in X$ and we set $x_\alpha := x_\alpha^0$. Note that we drop the dependence of x_α^δ on the concrete choice of y^δ because we are only interested in the influence of the noise level δ (and not of y^δ itself) on the regularized solutions.

The question to be answered is how fast $\|x_\alpha^\delta - x^\dagger\|$ decays to zero if the noise level δ goes to zero and the regularization parameter α is suitably chosen depending on δ. The well-known estimate

$$\|x_\alpha^\delta - x^\dagger\| \leq \|x_\alpha^\delta - x_\alpha\| + \|x_\alpha - x^\dagger\| \leq \frac{\delta}{2\sqrt{\alpha}} + \|x_\alpha - x^\dagger\| \tag{1.3}$$

shows that the speed of convergence can be estimated by the decay of $\|x_\alpha - x^\dagger\|$ if α goes to zero. A similar estimate holds for general linear regularization methods as described in [22]. Upper bounds for $\|x_\alpha - x^\dagger\|$ which go to zero if α goes to zero are called *profile functions* in [16].

We will use the following type of functions several times.

Definition 1.1. A function $f : [0, \infty) \to [0, \infty)$ is called *index function* if it is continuous and strictly monotonically increasing and if $f(0) = 0$.

In addition we need the notion of *conjugate functions*: Given a convex function $f : \mathbb{R} \to (-\infty, \infty]$ the conjugate function $f^* : \mathbb{R} \to (-\infty, \infty]$ of f is defined by $f^*(s) := \sup_{t\in\mathbb{R}}(st - f(t))$ (see [1, Section 2.3]). Most functions occuring in this article will be defined only for nonnegative arguments. If we calculate the conjugate of such a function then we implicitly assume that it is infinite for negative arguments.

The structure of the remaining part of this article is as follows: The four smoothness concepts under consideration are presented in Section 2 and some results from literatur describing their interplay are collected in Section 3. The main sections of this article are Section 4 and Section 5. In Section 4 we prove new connections among the three more modern smoothness assumptions and in Section 5 we give an alternative proof for one result from Section 4. The main results are formulated in Theorem 4.1 and Theorem 4.5. Finally, in Section 6 we draw some conclusions and make suggestions for further investigations.

2 Four smoothness concepts

In this section we present four concepts expressing the smoothness of the exact solution $x^\dagger$ of (1.1) and yielding convergence rates for the regularized solutions x_α. These approaches are termed

- source condition,

- approximate source condition,
- variational inequality,
- approximate variational inequality.

For each approach we state the corresponding convergence rate result in a separate subsection. In part these convergence rates look quite complex. Therefore in the fifth and last subsection we summarize them for the important special case of power-type rates.

2.1 Source conditions

Source conditions represent the classical concept for expressing solution smoothness in Hilbert spaces.

Definition 2.1. The exact solution $x^\dagger$ satisfies a *(general) source condition* with respect to an index function ϑ if

$$x^\dagger = \vartheta(A^*A)w \quad \text{for some } w \in X. \tag{2.1}$$

The proof of the following convergence rates theorem can be found in [22].

Theorem 2.2. *Let $x^\dagger$ satisfy a source condition with respect to a concave index function ϑ. Then*

$$\|x_\alpha - x^\dagger\| = \mathcal{O}(\vartheta(\alpha)) \quad \text{if } \alpha \to 0.$$

The concavity assumption on ϑ implies that ϑ is a qualification for the Tikhonov method in the sense of [22]. The same convergence rate holds for general linear regularization methods if ϑ is a qualification of the method.

In [21] it was shown that for each $x \in X$ there is an index function ϑ and a source element $w \in X$ such that a source condition $x = \vartheta(A^*A)w$ is satisfied. Since w can be written as $w = \tilde{\vartheta}(A^*A)\tilde{w}$ for some index function $\tilde{\vartheta}$ and some $\tilde{w} \in X$, too, we have $x = (\vartheta \circ \tilde{\vartheta})(A^*A)\tilde{w}$ and $\vartheta \circ \tilde{\vartheta}$ decays faster to zero than ϑ. Thus, convergence rates derived from (general) source conditions always can be improved somewhat, that is, they are not optimal.

2.2 Approximate source conditions

An extension of source conditions is the concept of approximate source conditions. The idea is to measure the violation of a prescribed benchmark source condition by considering the distance of the exact solution $x^\dagger$ to certain subsets of the space X.

Definition 2.3. For a given index function ψ, the *benchmark function*, we define the *distance function* $d_\psi : [0,\infty) \to [0,\infty)$ of the exact solution $x^\dagger$ by

$$d_\psi(R) := \inf \{\|\psi(A^*A)w - x^\dagger\| : w \in X, \|w\| \le R\}. \tag{2.2}$$

The term distance function is used because $d_\psi(R)$ is the distance of $x^\dagger$ to the image $\psi(A^*A)B_R(0)$ of the closed ball $B_R(0) \subseteq X$ with radius R centered at zero.

One can show that the distance function d_ψ is monotonically decreasing and tends to zero if $R \to \infty$. If $x^\dagger \notin \mathcal{R}(\psi(A^*A))$ then d_ψ is positive and strictly monotonically decreasing. These results and further details on approximate source conditions can be found in [14], the paper where the concept has been introduced, and [16].

The proof of the following theorem is given in [16].

Theorem 2.4. *Let ψ be a concave index function such that $x^\dagger \notin \mathcal{R}(\psi(A^*A))$ and let d_ψ be the distance function* (2.2) *for $x^\dagger$. Then*

$$\|x_\alpha - x^\dagger\| = \mathcal{O}\big(\psi(\alpha)\Phi^{-1}(\psi(\alpha))\big) \quad \text{if } \alpha \to 0,$$

where $\Phi(R) := \frac{d_\psi(R)}{R}$.

Note that also majorants of the distance function d_ψ yield convergence rates if d_ψ is replaced by such a majorant. Again the convergence rates result can be extended to general linear regularization methods if instead of concavity of ψ we assume that ψ is a qualification of the chosen method.

2.3 Variational inequalities

The concepts of source conditions and of approximate source conditions were originally developed for linear ill-posed problems in Hilbert spaces because they rely on spectral theory for selfadjoint operators. Only for the two source conditions $x^\dagger \in \mathcal{R}((A^*A)^{1/2}) = \mathcal{R}(A^*)$ and $x^\dagger \in \mathcal{R}(A^*A)$ and for approximate source conditions with benchmark function $\psi(t) = t^{1/2}$ or $\psi(t) = t$ generalizations to nonlinear problems in Banach spaces are known (see [4, 13, 24]). But to obtain convergence rates for nonlinear problems source conditions have to be combined with assumptions on the structure of nonlinearity and on properties of the involved spaces.

A new smoothness concept combining all assumptions needed to obtain convergence rates for nonlinear and even for nonsmooth problems in Banach spaces has been introduced in [15] in form of variational inequalities. Extensions can be found in [7, 8, 13, 25]. Such variational inequalities also can be formulated for general stabilizing functionals and non-metric fitting terms in Tikhonov regularization as well as for arbitrary measures expressing the distance between exact and regularized solutions. Besides their wide field of applicability, the main advantage is that no additional assumptions, e.g. assumptions on the structure of nonlinearity, have to be posed to obtain convergence rates.

A drawback is that by now variational inequalities provide convergence rates primarily for Tikhonov regularization (exception can be found [11, 12, 20]). But as we will see below, at least for linear problems in Hilbert spaces they can be used to obtain rates for any linear regularization method.

To understand the content of information in a variational inequality we consider them here in the standard Hilbert space setting for Tikhonov regularization. Thanks to spectral theory we can extend the concept somewhat, which will bring more insight into the nature of variational inequalities.

Definition 2.5. The exact solution $x^\dagger$ satisfies a *variational inequality* with constant $\beta > 0$, *modifier function* φ, and *benchmark function* ψ if

$$\beta\|x - x^\dagger\|^2 \leq \|x\|^2 - \|x^\dagger\|^2 + \varphi(\|\psi(A^*A)(x - x^\dagger)\|) \tag{2.3}$$

holds for all $x \in X$. The functions φ and ψ are assumed to be index functions.

Remark 2.6. The original variational inequality introduced for Tikhonov regularization

$$\|F(x) - y^\delta\|^p + \alpha\Omega(x) \to \min_x$$

in [15] reads

$$-\langle \xi, x - x^\dagger\rangle \leq \beta_1 B_\xi(x, x^\dagger) + \beta_2\|F(x) - F(x^\dagger)\|$$

and has to hold for all x in a certain subset of X, where X and Y are Banach spaces. The functional B_ξ is the Bregman distance with respect to a subgradient ξ of the Tikhonov stabilizing functional Ω at $x^\dagger$ and F is a nonlinear operator. Equivalently one can write

$$(1 - \beta_1)B_\xi(x, x^\dagger) \leq \Omega(x) - \Omega(x^\dagger) + \beta_2\|F(x) - F(x^\dagger)\|,$$

which becomes

$$(1 - \beta_1)\|x - x^\dagger\|^2 \leq \|x\|^2 - \|x^\dagger\|^2 + \beta_2\|A(x - x^\dagger)\|$$

for $\Omega = \|\bullet\|^2$ and a linear operator $A = F$ in Hilbert spaces, that is, we obtain (2.3) with modifier function $\varphi(t) = \beta_2 t$ and benchmark function

$\psi(t) = t^{1/2}$. The additional modifier function φ has been introduced in [2, 13]. General benchmark functions ψ are an extension available in Hilbert spaces.

Before we come to the convergence rates result we give some additional information on variational inequalities in Hilbert spaces. At first we prove some restrictions on the constant β and the modifier function φ.

Proposition 2.7. *If $x^\dagger \neq 0$ satisfies a variational inequality with constant β, modifier function φ, and benchmark function ψ, then the following assertions are true:*

(i) $\beta \leq 1$,

(ii) $\beta = 1$ *implies* $\varphi(t) \geq ct$ *for some* $c > 0$ *and all* $t \geq 0$.

(iii) $t = \mathcal{O}(\varphi(t)) \quad$ *if* $t \to 0$,

Proof. The variational inequality (2.3) for $x := x^\dagger - t\tilde{x}$ with $t > 0$ and $\|\tilde{x}\| = 1$ reads

$$\beta t^2 \leq -2t\langle x^\dagger, \tilde{x}\rangle + t^2 + \varphi\big(\|\psi(A^*A)\tilde{x}\|t\big),$$

which by multiplication with $1/t$, substitution of t for $\|\psi(A^*A)\tilde{x}\|t$, and minor rearrangements is equivalent to

$$\frac{\beta - 1}{\|\psi(A^*A)\tilde{x}\|^2}t + \frac{2\langle x^\dagger, \tilde{x}\rangle}{\|\psi(A^*A)\tilde{x}\|} \leq \frac{\varphi(t)}{t}. \tag{2.4}$$

Assume that $\beta > 1$. Since the range of $\psi(A^*A)$ is not closed we find a sequence $(\tilde{x}_n)_{n\in\mathbb{N}}$ in X with $\|\tilde{x}_n\| = 1$ and $\psi(A^*A)\tilde{x}_n \to 0$. In addition we can choose this sequence in such a way that $\langle x^\dagger, \tilde{x}_n\rangle \geq 0$. For each fixed t from (2.4) we now get

$$\frac{\varphi(t)}{t^2} \geq \frac{\beta - 1}{\|\psi(A^*A)\tilde{x}_n\|^2}$$

and the right-hand side goes to infinity if $n \to \infty$. Thus, φ has to be infinite at each $t > 0$, which is a contradiction. Therefore $\beta \leq 1$.

Assuming $\beta = 1$, inequality (2.4) reads

$$\frac{2\langle x^\dagger, \tilde{x}\rangle}{\|\psi(A^*A)\tilde{x}\|} \leq \frac{\varphi(t)}{t}$$

and therefore gives

$$c := \sup\left\{\frac{2\langle x^\dagger, \tilde{x}\rangle}{\|\psi(A^*A)\tilde{x}\|} : \tilde{x} \in X, \|\tilde{x}\| = 1\right\} < \infty$$

and $\frac{\varphi(t)}{t} \geq c$ for all $t > 0$.

For general $\beta \leq 1$ inequality (2.4) with fixed $\tilde{x}$ such that $\langle x^\dagger, \tilde{x}\rangle > 0$ provides $\frac{\varphi(t)}{t} \geq \tilde{c}$ for some $\tilde{c} > 0$ and all sufficiently small $t > 0$. □

Example 2.8. If the modifier function φ in Proposition 2.7 is of power-type, that is $\varphi(t) := at^\kappa$ with $a > 0$ and $\kappa > 0$, then $t = \mathcal{O}(\varphi(t))$ (if $t \to 0$) implies $\kappa \leq 1$. This bound for κ is already known from [19]. In the case $\beta = 1$ a variational inequality with a power-type modifier function can only hold if $\kappa = 1$ and $a \geq c$, where c is from (ii) in Proposition 2.7.

The next two propositions show that, at least for modifier functions φ of power-type, the constant β in a variational inequality can be changed without corrupting the inequality. Only the modifier function φ has to be adjusted slightly if it is not linear.

Proposition 2.9. *An element $x^\dagger \in X$ satisfies a variational inequality with constant $\beta \leq 1$, benchmark function ψ, and modifier function $\varphi(t) = at$, $a > 0$, if and only if*

$$|\langle x^\dagger, \tilde{x}\rangle| \leq \frac{a}{2}\|\psi(A^*A)\tilde{x}\|$$

for all $\tilde{x} \in X$ with $\|\tilde{x}\| = 1$.

Note that this equivalent formulation is independent of β.

Proof. As at the beginning of the proof of Proposition 2.7 we see that the variational inequality is equivalent to

$$\frac{\beta - 1}{\|\psi(A^*A)\tilde{x}\|^2}t + \frac{2\langle x^\dagger, \tilde{x}\rangle}{\|\psi(A^*A)\tilde{x}\|} - a \leq 0$$

for all $\tilde{x}$ with $\|\tilde{x}\| = 1$ and all $t > 0$. For $t \to 0$ we get $\frac{2\langle x^\dagger, \tilde{x}\rangle}{\|\psi(A^*A)\tilde{x}\|} - a \leq 0$ and if $\frac{2\langle x^\dagger, \tilde{x}\rangle}{\|\psi(A^*A)\tilde{x}\|} - a \leq 0$ then the above inequality is obviously satisfied. □

Proposition 2.10. *An element $x^\dagger \in X$ satisfies a variational inequality with constant $\beta < 1$, benchmark function ψ, and modifier function $\varphi(t) = at^\kappa$, where $a > 0$ and $\kappa \in (0, 1)$, if and only if*

$$|\langle x^\dagger, \tilde{x}\rangle| \leq \tfrac{2-\kappa}{2}\left(\tfrac{1-\beta}{1-\kappa}\right)^{\frac{1-\kappa}{2-\kappa}} a^{\frac{1}{2-\kappa}}\|\psi(A^*A)\tilde{x}\|^{\frac{\kappa}{2-\kappa}} \tag{2.5}$$

for all $\tilde{x} \in X$ with $\|\tilde{x}\| = 1$.

Proof. As at the beginning of the proof of Proposition 2.7 we see that the variational inequality is equivalent to

$$f(t) := \frac{\beta - 1}{\|\psi(A^*A)\tilde{x}\|^2} t + \frac{2\langle x^\dagger, \tilde{x}\rangle}{\|\psi(A^*A)\tilde{x}\|} - a t^{-(1-\kappa)} \leq 0$$

for all $\tilde{x}$ with $\|\tilde{x}\| = 1$ and all $t > 0$. Since f is twice continuously differentiable and strictly concave, seeking for zeros of the derivative we find the global maximum at

$$t^* = \left(\frac{(1-\kappa)a\|\psi(A^*A)\tilde{x}\|^2}{1-\beta}\right)^{\frac{1}{2-\kappa}}$$

with

$$f(t^*) = \frac{2\langle x^\dagger, \tilde{x}\rangle}{\|\psi(A^*A)\tilde{x}\|} - (2-\kappa)\left(\frac{1-\beta}{(1-\kappa)\|\psi(A^*A)\tilde{x}\|^2}\right)^{\frac{1-\kappa}{2-\kappa}} a^{\frac{1}{2-\kappa}}.$$

Simple rearrangements show that $f(t^*) \leq 0$ is equivalent to the asserted inequality. □

Corollary 2.11. *A variational inequality with constant* $\beta_1 < 1$*, benchmark function* ψ*, and modifier function* $\varphi(t) = at^\kappa$*, where* $a > 0$ *and* $\kappa \in (0,1)$*, holds if and only if a variational inequality with constant* $\beta_2 < 1$*, benchmark function* ψ*, and modifier function* $\varphi(t) = bt^\kappa$ *is satisfied, where* $b = \left(\frac{1-\beta_1}{1-\beta_2}\right)^{1-\kappa} a$*.*

Proof. The equivalent formulation (2.5) is the same for both variational inequalities. □

Since a variational inequality with $\beta = 1$ always implies a variational inequality with arbitrary $\beta < 1$ we will focus on the case $\beta < 1$ in the sequel.

The following convergence rates theorem is a special case of the main theorems in [10] and [7], from which one can deduce the proof. As one sees in [7], the assumptions on φ can be weakend slightly; but the price would be a more complex formulation of the theorem.

Theorem 2.12. *Let the exact solution* $x^\dagger$ *satisfy a variational inequality with a modifier function* φ *such that* $\tilde{\varphi} := \varphi(\sqrt{\bullet})$ *is concave and with the benchmark function* $\psi(t) = t^{1/2}$*. Then*

$$\|x_\alpha - x^\dagger\| = \mathcal{O}\left(\sqrt{\tfrac{1}{\alpha}\left(\tilde{\varphi}^{-1}\right)^*(\alpha)}\right) \quad \text{if } \alpha \to 0,$$

where $\left(\tilde{\varphi}^{-1}\right)^*$ *denotes the conjugate function of* $\tilde{\varphi}^{-1}$ *(see end of Section 1).*

Proof. Using the variational inequality (2.3) with $x = x_\alpha$ and exploiting that x_α is a minimizer of the Tikhonov functional with exact data y^0 we obtain

$$\begin{aligned}
\beta\|x_\alpha - x^\dagger\|^2 &\leq \|x_\alpha\|^2 - \|x^\dagger\|^2 + \varphi(\|A(x_\alpha - x^\dagger)\|) \\
&= \tfrac{1}{\alpha}(\|A(x_\alpha - x^\dagger)\|^2 + \alpha\|x_\alpha\|^2 - \alpha\|x^\dagger\|^2) \\
&\qquad + \varphi(\|A(x_\alpha - x^\dagger)\|) - \tfrac{1}{\alpha}\|A(x_\alpha - x^\dagger)\|^2 \\
&\leq \varphi(\|A(x_\alpha - x^\dagger)\|) - \tfrac{1}{\alpha}\|A(x_\alpha - x^\dagger)\|^2 \\
&\leq \sup_{t>0}\big(\varphi(t) - \tfrac{1}{\alpha}t^2\big) = \sup_{t>0}\big(\tilde{\varphi}(t) - \tfrac{1}{\alpha}t\big) \\
&= \sup_{t>0}\big(t - \tfrac{1}{\alpha}\tilde{\varphi}^{-1}(t)\big) = \tfrac{1}{\alpha}\big(\tilde{\varphi}^{-1}\big)^*(\alpha).
\end{aligned}$$

□

Since the $\mathcal{O}$-expression in Theorem 2.12 looks somewhat abstract we should mention that there exists an a priori parameter choice $\delta \mapsto \alpha(\delta)$ such that

$$\|x^\delta_{\alpha(\delta)} - x^\dagger\| = \mathcal{O}\big(\sqrt{\varphi(\delta)}\big) \quad \text{if } \delta \to 0$$

under the assumptions of Theorem 2.12 (see [2, 7, 10] for details).

This theorem will be extended in Remark 4.6 to cover general benchmark functions ψ and general linear regularization methods. In the recent literature variational inequalities were only used in connection with Tikhonov regularization because the technique of the proof heavily relies on the structure of the Tikhonov functional.

2.4 Approximate variational inequalities

Before variational inequalities have been extended from power-type to general modifier functions in [2] the concept of approximate variational inequalities was introduced in [8]. The idea is to measure the violation of a variational inequality (2.3) with the modifier function $\varphi(t) = ct$, $c \geq 0$, since this is the modifier function with fastest decay to zero if $t \to 0$.

Definition 2.13. The exact solution $x^\dagger$ satisfies an *approximate variational inequality* with constant $\beta \in (0, 1)$ and *benchmark function* ψ if the *distance function* $D_\psi : [0, \infty) \to [0, \infty)$ defined by

$$D_\psi(r) := \sup_{x \in X}\big(\beta\|x - x^\dagger\|^2 - \|x\|^2 + \|x^\dagger\|^2 - r\|\psi(A^*A)(x - x^\dagger)\|\big)$$

goes to zero for $r \to \infty$. The benchmark function ψ is assumed to be an index function.

The distance function D_ψ is nonnegative, convex, and monotonically decreasing. If $D_\psi(r) > 0$ for all $r \geq 0$ then D_ψ is strictly decreasing. Further details on approximate variational inequalities can be found in [8]. Also the proof of the following theorem is given there, but for a more general setting.

Theorem 2.14. *Let the exact solution $x^\dagger$ satisfy an approximate variational inequality with the benchmark function $t \mapsto t^{1/2}$ and with a distance function D_ψ. Then*

$$\|x_\alpha - x^\dagger\| = \mathcal{O}\big(\sqrt{\alpha}\Psi^{-1}(\sqrt{\alpha})\big) \quad \text{if } \alpha \to 0,$$

where $\Psi(r) := \frac{\sqrt{D_\psi(r)}}{r}$.

As for approximate source conditions the distance function D_ψ can be replaced by some majorant. An extension to general benchmark functions ψ will be given in Remark 4.6.

2.5 Summary for power-type smoothness assumptions

Because the above convergence rates results are quite general we summarize them for situations where all the appearing index functions are monomials. Note that a constant factor in any index function does not influence the convergence rate. The above convergence rate theorems provide the rate

$$\|x_\alpha - x^\dagger\| = \mathcal{O}(\alpha^\mu) \quad \text{if } \alpha \to 0$$

if one of the following four conditions is satisfied:

- a source condition with $\vartheta(t) = t^\mu$, $\mu \in (0, 1]$;
- an approximate source condition with benchmark function $\psi(t) = t^\eta$, $0 < \mu < \eta \leq 1$, and distance function $d_\psi(R) = \mathcal{O}\big(R^{\frac{-\mu}{\eta-\mu}}\big)$ if $R \to \infty$;
- a variational inequality with modifier function $\varphi(t) = t^{\frac{4\mu}{1+2\mu}}$, $\mu \in (0, \frac{1}{2}]$, and benchmark function $\psi(t) = \sqrt{t}$;
- an approximate variational inequality with benchmark function $\psi(t) = \sqrt{t}$ and distance function $D_\psi(r) = \mathcal{O}\big(r^{\frac{-4\mu}{1-2\mu}}\big)$ if $r \to \infty$, $\mu \in (0, \frac{1}{2})$.

3 Known results on interconnections

First we look at equivalent formulations for source conditions. It is obvious that a source conditions with index function ψ is satisfied if and only if the distance function d_ψ from the corresponding approximate source condition is zero for large arguments. In addition from [26, Section 3.2] we know that a source condition with ψ holds if and only if a variational inequality with benchmark ψ and modifier function $\varphi(t) = at$, $a > 0$, is satisfied. The following proposition repeats this result with our notation and with a different proof.

Proposition 3.1. *An element $x^\dagger$ satisfies a variational inequality with constant $\beta \leq 1$, benchmark function ψ, and modifier function $\varphi(t) = at$, $a \geq 0$, if and only if there exists some $w \in X$ with $\|w\| \leq \frac{a}{2}$ such that $x^\dagger = \psi(A^*A)w$.*

Proof. The variational inequality is satisfied if and only if

$$f(x) := \beta\|x - x^\dagger\|^2 - \|x\|^2 + \|x^\dagger\|^2 - a\|\psi(A^*A)(x - x^\dagger)\| \leq 0$$

for all $x \in X$. Since $f(x^\dagger) = 0$ and f is concave, this happens if and only if f has a global maximum at $x^\dagger$. The subdifferential of $-f$ at $x^\dagger$ is given by

$$\partial(-f)(x^\dagger) = \{2x^\dagger + a\psi(A^*A)w : w \in X, \|w\| \leq 1\}$$

and thus, $0 \in \partial(-f)(x^\dagger)$ if and only if there exists some $w \in X$ with $\|w\| \leq 1$ and $x^\dagger = \psi(A^*A)\big(-\frac{a}{2}w\big)$. □

As for source conditions, the distance function D_ψ from the approximate variational inequality is zero for large arguments if and only if a variational inequality with benchmark ψ and modifier function $\varphi(t) = at$, $a > 0$, is satisfied. Therefore we have the following relations:

$d_\psi(R) = 0$ for large R	$\Leftrightarrow$	source condition with ψ	$\Leftrightarrow$	variational inequality with ψ and $\varphi(t) = at$	$\Leftrightarrow$	$D_\psi(r) = 0$ for large r

More interesting is the situation if the benchmark smoothness is higher then the smoothness of $x^\dagger$. So assume that a source condition with index function $\vartheta(t) = t^\mu$, $\mu \in (0, 1)$, is satisfied, but not with index function $\psi(t) = t^\nu$, $\nu \in (\mu, 1]$. Then the following is true:

- The distance function d_ψ from the approximate source conditions is bounded by
$$d_\psi(R) = \mathcal{O}\big(R^{\frac{-\mu}{\nu-\mu}}\big) \quad \text{if } R \to \infty$$
(see [5] and for more general functions ϑ and ψ also [16]).

- The exact solution $x^\dagger$ satisfies a variational inequality with benchmark ψ and modifier function $\varphi(t) = at^{\frac{2\mu}{\mu+\nu}}$ for some $a > 0$ (see [19]).

In addition the following relations between smoothness concepts are given in the literature:

- A variational inequality with benchmark function ψ and modifier function $\varphi(t) = t^\kappa$, $\kappa \in (0,1)$, implies an approximate variational inequality with benchmark function ψ and distance function
$$D_\psi(r) = \mathcal{O}\big(r^{\frac{-\kappa}{1-\kappa}}\big) \quad \text{if } r \to \infty$$
(see [8]).

- An approximate source condition with benchmark function ψ and distance function d_ψ implies an approximate variational inequality with benchmark function ψ and distance function $D_\psi(r) = \mathcal{O}\big(d_\psi^2(r)\big)$ if $r \to \infty$. This can be proven analogously to Lemma 5.4 in [8].

- An approximate source condition with benchmark function ψ and distance function $d_\psi(R) > 0$ for all $R \geq 0$ implies a variational inequality with benchmark function ψ and modifier function $\varphi(t) = ct\Theta^{-1}(t)$ for some $c > 0$, where $\Theta(R) := \frac{1}{R} d_\psi^2(R)$. This can be proven analogously to Theorem 5.2 in [2].

In the light of the known cross connections the idea of approximate variational inequalities seems to be the most general concept.

4 New results on interconnections

In this section we prove two new results which completely clarify the connections between variational inequalities, approximate variational inequalities, and approximate source conditions in Hilbert spaces. In essence we show that all three contain exactly the same information about the exact solution $x^\dagger$.

The first main theorem shows that variational inequalities can carry the same information as approximate variational inequalities.

Theorem 4.1. *By M we denote the family of all modifier functions φ for which $x^\dagger$ satisfies a variational inequality with constant β and benchmark function ψ. Then $x^\dagger$ satisfies an approximate variational inequality with distance function*

$$D_\psi = \min_{\varphi \in M} (-\varphi)^*(-\bullet) \quad \textit{(pointwise minimum)}$$

and the minimum is attained for $\varphi(t) = -D_\psi^*(-t)$.

The *-notation was introduced at the end of Section 1. Note that $(-\varphi)^*(-r) = r(\varphi^{-1})^*(\frac{1}{r})$.

Proof. Let $x^\dagger$ satisfy a variational inequality with modifier function φ. Then

$$\begin{aligned}
&\beta\|x - x^\dagger\|^2 - \|x\|^2 + \|x^\dagger\|^2 - r\|\psi(A^*A)(x - x^\dagger)\| \\
&\quad \leq \varphi(\|\psi(A^*A)(x - x^\dagger)\|) - r\|\psi(A^*A)(x - x^\dagger)\| \\
&\quad \leq \sup_{t \geq 0}(\varphi(t) - rt) = \sup_{t \geq 0}\bigl(-rt - (-\varphi)(t)\bigr) = (-\varphi)^*(-r)
\end{aligned}$$

for all $r \geq 0$ and all $x \in X$, that is, $D_\psi(r) \leq (-\varphi)^*(-r)$ for all $r \geq 0$.

On the other hand, given D_ψ, we have

$$\beta\|x - x^\dagger\|^2 - \|x\|^2 + \|x^\dagger\|^2 \leq r\|\psi(A^*A)(x - x^\dagger)\| + D_\psi(r)$$

for all $r \geq 0$ and all $x \in X$. Therefore

$$\begin{aligned}
&\beta\|x - x^\dagger\|^2 - \|x\|^2 + \|x^\dagger\|^2 \\
&\quad \leq \inf_{r \geq 0}\bigl(r\|\psi(A^*A)(x - x^\dagger)\| + D_\psi(r)\bigr) \\
&\quad = -\sup_{r \geq 0}\bigl(-\|\psi(A^*A)(x - x^\dagger)\|r - D_\psi(r)\bigr) \\
&\quad = -D_\psi^*(-\|\psi(A^*A)(x - x^\dagger)\|),
\end{aligned}$$

that is, $x^\dagger$ satisfies a variational inequality with modifier function $\varphi(t) = -D_\psi^*(-t)$. □

Remark 4.2. Since the proof of Theorem 4.1 does not use any tools restricted to Hilbert spaces, the theorem also holds for variational inequalities in Banach spaces settings as introduced in Remark 2.6.

In preparation of the second main theorem we prove two lemmas. The first one provides a simplified expression for D_ψ.

Lemma 4.3. *Let $x^\dagger$ satisfy an approximate variational inequality with constant $\beta < 1$, benchmark function ψ, and distance function D_ψ and let $x_r \in X$ be a maximizer in the definition of $D_\psi(r)$, that is,*

$$D_\psi(r) = \beta\|x_r - x^\dagger\|^2 - \|x_r\|^2 + \|x^\dagger\|^2 - r\|\psi(A^*A)(x_r - x^\dagger)\|.$$

Then $D_\psi(r) = (1-\beta)\|x_r - x^\dagger\|^2$.

Proof. If $x_r = x^\dagger$ then $D_\psi(r) = 0$ by the definition of $D_\psi(r)$. So assume that $x_r \neq x^\dagger$. Then the gradient of

$$x \mapsto \beta\|x - x^\dagger\|^2 - \|x\|^2 + \|x^\dagger\|^2 - r\|\psi(A^*A)(x - x^\dagger)\|$$

at x_r has to be zero, that is,

$$2\beta(x_r - x^\dagger) - 2x_r - r\frac{\psi^2(A^*A)(x_r - x^\dagger)}{\|\psi(A^*A)(x_r - x^\dagger)\|} = 0. \tag{4.1}$$

Applying $\langle \bullet, x_r - x^\dagger\rangle$ at both sides we get

$$-r\|\psi(A^*A)(x_r - x^\dagger)\| = -2\beta\|x_r - x^\dagger\|^2 + 2\langle x_r, x_r - x^\dagger\rangle$$

and therefore

$$\begin{aligned} D_\psi(r) &= \beta\|x_r - x^\dagger\|^2 - \|x_r\|^2 + \|x^\dagger\|^2 - 2\beta\|x_r - x^\dagger\|^2 + 2\langle x_r, x_r - x^\dagger\rangle \\ &= (1-\beta)\|x_r - x^\dagger\|^2. \end{aligned}$$

□

The second lemma connects the minimizers from the definition of $d_\psi(\frac{r}{2})$ (approximate source condition) with the maximizers of $D_\psi(r)$ (approximate variational inequality).

Lemma 4.4. *Let $x^\dagger \notin \mathcal{R}(\psi(A^*A))$ satisfy an approximate variational inequality with constant $\beta < 1$, benchmark function ψ, and distance function D_ψ and let $w_R \in \operatorname{argmin}\{\|\psi(A^*A)w - x^\dagger\| : w \in X, \|w\| \leq R\}$. Then*

$$x_r := x^\dagger + \tfrac{1}{1-\beta}\big(\psi(A^*A)w_{r/2} - x^\dagger\big)$$

is a maximizer in the definition of $D_\psi(r)$, that is,

$$D_\psi(r) = \beta\|x_r - x^\dagger\|^2 - \|x_r\|^2 + \|x^\dagger\|^2 - r\|\psi(A^*A)(x_r - x^\dagger)\|.$$

Proof. By the definition of w_R there exists some Lagrange multiplier $\lambda \geq 0$ with

$$\psi(A^*A)\big(\psi(A^*A)w_R - x^\dagger\big) = -\frac{\lambda}{2} w_R. \tag{4.2}$$

For $\lambda = 0$ we would get $x^\dagger = \psi(A^*A)w_R$ which contradicts $x^\dagger \notin \mathcal{R}(\psi(A^*A))$. Thus $\lambda > 0$ and therefore $\|w_R\| = R$. Defining x_r as in the lemma and using (4.2) one easily verifies (4.1), which is equivalent to the fact that x_r is a maximizer in the definition of $D_\psi(r)$. $\square$

The second main theorem shows that approximate variational inequalities are equivalent to approximate source conditions.

Theorem 4.5. *Let $x^\dagger \notin \mathcal{R}(\psi(A^*A))$ satisfy an approximate variational inequality with constant $\beta < 1$, benchmark function ψ, and distance function D_ψ and let d_ψ be the distance function from the approximate source condition with respect to ψ. Then*

$$D_\psi(r) = \tfrac{1}{1-\beta} d_\psi^2\big(\tfrac{r}{2}\big) \quad \textit{for all } r \geq 0.$$

Proof. The assertion is an immediate consequence of the preceding two lemmas (apply the first lemma to x_r from the second lemma). $\square$

Remark 4.6. Combining the two theorems of this section, from a variational inequality with concave benchmark function ψ and modifier function $\varphi(t) = t^\kappa$, $\kappa \in (0,1)$, we get an approximate source condition with distance function

$$d_\psi(R) = \mathcal{O}\big(R^{\frac{-\kappa}{2(1-\kappa)}}\big) \quad \text{if } R \to \infty$$

and thus the convergence rate

$$\|x_\alpha - x^\dagger\| = \mathcal{O}\big(\psi(\alpha)^{\frac{\kappa}{2-\kappa}}\big) \quad \text{if } \alpha \to 0.$$

Analogously we can derive convergence rates from approximate variational inequalities with general benchmark function. Assuming that ψ is a qualification of a general linear regularization method variational inequalities also provide convergence rates for this method. Till now a direct (without detour via approximate source conditions) proof of convergence rates for general linear regularization methods given a variational inequality is missing.

It remains to investigate the relation between source conditions and the other three smoothness concepts. This question is addressed in [9] and the answer given there is that approximate source conditions provide also lower bounds for the regularization error $\|x_\alpha - x^\delta\|$. If the distance function d_ψ decays not too fast, then the lower bound obtained in [9] coincides with the well-known upper bound provided by the concept of approximate source conditions. In this sense approximate source conditions yield optimal rates. Since one knowns (see [23]) that power-type rates $\mathcal{O}(\alpha^\mu)$ do not imply the corresponding source condition $x^\dagger \in \mathcal{R}((A^*A)^\mu)$, source conditions are less powerful than approximate source conditions and therefore, as shown in this article, also less powerful than variational inequalities and approximate variational inequalities.

5 Fenchel duality between source conditions and variational inequalities

The equivalence of approximate variational inequalities and approximate source conditions is not by chance but comes from Fenchel duality. The idea to consider the Fenchel dual of the minimization problem in the definition of the distance function d_ψ of an approximate source condition is taken from [18].

We proceed the other way round, that is, we derive the Fenchel dual of the maximization problem in the definition of the distance function D_ψ of an approximate variational inequality.

Setting $f(x) := \beta\|x-x^\dagger\|^2-\|x\|^2+\|x^\dagger\|^2$ and $g(x) := -r\|x-\psi(A^*A)x^\dagger\|$ for $x \in X$ the Fenchel dual of

$$\beta\|x - x^\dagger\|^2 - \|x\|^2 + \|x^\dagger\|^2 - r\|\psi(A^*A)(x - x^\dagger)\| \to \max_{x\in X}$$

is given by

$$(-f)^*(-\psi(A^*A)u) + (-g)^*(u) \to \min_{u\in X}$$

with $(-f)^*$ and $(-g)^*$ being the conjugate functions of $-f$ and $-g$ (introduced at the end of Section 1); see [1] for details. The calculation of the conjugate functions leads to

$$(-f)^*(u) = \tfrac{1}{1-\beta}\big(\|\tfrac{1}{2}u\|^2 - \beta\langle u, x^\dagger\rangle + \|x^\dagger\|^2\big)$$

and

$$(-g)^*(u) = \langle\psi(A^*A)u, x^\dagger\rangle + r\chi_r(u)$$

with $\chi_r(u) = 0$ for $\|u\| \leq r$ and $\chi_r(u) = \infty$ outside the r-ball. Hence, the objective function of the dual problem becomes

$$
\begin{aligned}
&(-f)^*(-\psi(A^*A)u) + (-g)^*(u) \\
&\quad = \tfrac{1}{1-\beta}\big(\|\tfrac{1}{2}\psi(A^*A)u\|^2 + \beta\langle\psi(A^*A)u, x^\dagger\rangle + \|x^\dagger\|^2\big) \\
&\qquad + \langle\psi(A^*A)u, x^\dagger\rangle + r\chi_r(u) \\
&\quad = \tfrac{1}{1-\beta}\|\psi(A^*A)(\tfrac{1}{2}u) - x^\dagger\|^2 + r\chi_r(u).
\end{aligned}
$$

Thus, also via Fenchel duality we obtain $D_\psi(r) = \frac{1}{1-\beta} d_\psi^2\big(\frac{r}{2}\big)$.

6 Conclusions and open questions

We have seen that in the standard Hilbert space setting the three modern concepts for expressing solution smoothness (approximate source conditions, variational inequalities, approximate variational inequalities) coincide and, taking into account [9], that source conditions contain less accurate information about the exact solution $x^\dagger$ than the other three concepts. By the way it turned out that the modifier function in a variational inequality contains the same information as the distance function for approximate source conditions or approximate variational inequalities.

A consequence is that the power of variational inequalities shows not up until we look at nonlinear problems in Banach spaces. Though our results do not apply to such more general settings, at least extensions in this direction are possible and have to be investigated in future. Especially the expression obtained when calculating the Fenchel dual of approximate variational inequalities in Banach spaces (with linear operator) suggests to modify the definition of approximate source conditions by replacing the norm by a certain Bregman distance. This approach looks very promising because for norm based distance functions in Banach spaces one needs additional assumptions to connect the Bregman distance for which convergence rates shall be shown with the norm used for defining the distance function (cf. [13]). Thus, further investigation of the duality approach in Banach spaces could lead to new and simplified proofs of convergence rates.

Another important point for future work is the calculation of distance functions for concrete examples. Despite the simple structure of approximate source conditions a direct calculation of distance functions is very difficult. First attempts can be found in [9, 18]. In part the results there also take advantage of Fenchel duality.

Acknowledgement

The author thanks the referees for their valuable comments. Especially the hints concerning extended use of conjugate functions significantly improved the presentation and also motivated the extension of one result.

References

[1] R. I. Boţ, S.-M. Grad, and G. Wanka. *Duality in Vector Opimization*. Vector Optimization. Springer, Berlin, Heidelberg, 2009.

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

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

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

[5] D. Düvelmeyer, B. Hofmann, and M. Yamamoto. Range Inclusions and Approximate Source Conditions with General Benchmark Functions. *Numerical Functional Analysis and Optimization*, 28(111–12):1245–1261, 2007.

[6] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Mathematics and Its Applications. Kluwer Academic Publishers, Dordrecht, 1996.

[7] J. Flemming. Theory and examples of variational regularization with non-metric fitting functionals. Preprint series of the Department of Mathematics 2010-14, Chemnitz University of Technology, Chemnitz, Germany, July 2010.

[8] J. Flemming and B. Hofmann. A New Approach to Source Conditions in Regularization with General Residual Term. *Numerical Functional Analysis and Optimization*, 31(3):254–284, 2010.

[9] J. Flemming, B. Hofmann, and P. Mathé. Sharp converse results for the regularization error using distance functions. WIAS preprint No. 1535, Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany, September 2010.

[10] M. Grasmair. Generalized Bregman Distances and Convergence Rates for Non-convex Regularization Methods. Industrial Geometry report 104, University of Vienna, Vienna, Austria, May 2010.

[11] M. Grasmair, M. Haltmeier, and O. Scherzer. The Residual Method for Regularizing Ill-Posed Problems. *arXiv.org*, arXiv:cs/0905.1187v1 [math.OC], May 2009. http://arxiv.org/abs/0905.1187.

[12] T. Hein. Convergence rates for multi-parameter regularization in Banach spaces. *International Journal of Pure and Applied Mathematics*, 43(4):593–614, 2008.

[13] T. Hein and B. Hofmann. Approximate source conditions for nonlinear ill-posed problems—chances and limitations. *Inverse Problems*, 25(3):035033 (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(3):351–371, 2006.

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

[16] B. Hofmann and P. Mathé. Analysis of Profile Functions for General Linear Regularization Methods. *SIAM Journal on Numerical Analysis*, 45(3):1122–1141, 2007.

[17] B. Hofmann, P. Mathé, and H. von Weizsäcker. Regularization in Hilbert space under unbounded operators and general source conditions. *Inverse Problems*, 25(11):115013 (15pp), 2009.

[18] B. Hofmann, M. Schieck, and L. von Wolfersdorf. On the analysis of distance functions for linear ill-posed problems with an application to the integration operator in L^2. *Journal of Inverse and Ill-posed Problems*, 15(1):83–98, 2007.

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

[20] B. Kaltenbacher and B. Hofmann. Convergence rates for the iteratively regularized Gauss–Newton method in Banach spaces. *Inverse Problems*, 26(3):035007 (21 pp), 2010.

[21] P. Mathé and B. Hofmann. How general are general source conditions? *Inverse Problems*, 24(1):015009 (5 pp), 2008.

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

[23] A. Neubauer. On converse and saturation results for Tikhonov regularization of linear ill-posed problems. *SIAM Journal on Numerical Analysis*, 34(2):517–527, 1997.

[24] A. Neubauer, T. Hein, B. Hofmann, S. Kindermann, and U. Tautenhahn. Improved and extended results for enhanced convergence rates of Tikhonov regularization in Banach spaces. *Applicable Analysis*, 2010. to appear.

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

[26] 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.