A note on convergence rates for variational regularization with non-convex residual term

Bernd Hofmann *

Linz/Austria – May 26, 2009

Abstract

This note formulates some assertions and conjectures concerning an up to now missing case of convergence rates results for variational regularization of nonlinear ill-posed problems in Banach spaces. If the residual term is the p-th power of a Banach space norm, then the use of powers $0 < p < 1$ instead of the common values $1 \le p < \infty$ leads to an artificial limitation of convergence rates. This effect also occurs for general residual terms when they represent concave monomials of that distance which is bounded by the noise level and expresses some kind of qualification for the regularization method.

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

Keywords: Nonlinear ill-posed problems, variational regularization, variational inequalities, non-convex residual term, convergence rates, qualification.

1 Introduction

We will consider ill-posed operator equations

$$F(u) = v \tag{1.1}$$

with an in general nonlinear operator $F : \mathcal{D}(F) \subseteq U \to V$ possessing the domain $\mathcal{D}(F)$ and mapping between normed real linear spaces U and V with norms $\|\cdot\|_U$ and $\|\cdot\|_V$, respectively. Based on noisy data v^δ of the exact right-hand side $v = v^0 \in F(\mathcal{D}(F))$ with

$$\|v^\delta - v\|_V \le \delta \tag{1.2}$$

*Department of Mathematics, Chemnitz University of Technology, 09107 Chemnitz, Germany.
This paper was written during the research stay of the author at Johann Radon Institute for Computational and Applied Mathematics (RICAM), 4040 Linz, Austria, April–July 2009.
Email: `hofmannb@mathematik.tu-chemnitz.de`.

and noise level $\delta \geq 0$ we analyze stable approximate solutions u_α^δ as minimizers of the Tikhonov type functional

$$T_\alpha^{v^\delta}(u) := \|F(u) - v^\delta\|_V^p + \alpha\,\Omega(u) \tag{1.3}$$

with regularization parameters $\alpha > 0$ and a penalty functional $\Omega : U \to [0, +\infty]$ with proper domain

$$\mathcal{D}(\Omega) := \{u \in U : \Omega(u) \neq +\infty\}\ .$$

Hence we have to minimize in the functional $T_\alpha^{v^\delta}$ in (1.3) over the intersection

$$\mathcal{D} := \mathcal{D}(F) \cap \mathcal{D}(\Omega)$$

of the domains of F and Ω which is assumed to be non-empty.

In the literature the variational regularization (1.3) is studied comprehensively for convex penalty functionals Ω (see, e.g., [2, 3, 4, 6]), and in the context of sparsity also for non-convex ones (see, e.g., [7] and recent papers of O. Scherzer's Innsbruck research group). The exponent p, however, is always chosen from the interval $[1, \infty)$, since the background of $\|F(u) - v^\delta\|_V^p$ is mostly $\|F(u) - v^\delta\|_{L^p}^p$ and L^p is only a Banach space for $p \geq 1$. On the other hand, the theory allows a wide range of Banach space norms $\|\cdot\|_V$. In this note we therefore outline the consequences for convergence rates occurring when small exponents $0 < p < 1$ are connected with convex penalties Ω. Then the weak point is that one uses a concave function of the norm as residual term, which leads to an artificial limitation of convergence rates. This is an analytical argument that complements the abhorrence of the case $0 < p < 1$ coming from numerical difficulties in finding minimizers of (1.3). On the other hand, this expresses some kind of qualification for the regularization method depending on p in analogy to the qualification concepts in linear regularization theory.

Throughout this note we make the following assumptions:

Assumption 1.1

1. *U and V are reflexive Banach spaces with duals U^* and V^*, respectively. In U and V we consider in addition to the norm convergence the associated weak convergence. That means in U*

$$u_k \rightharpoonup u \quad \Longleftrightarrow \quad \langle f, u_k\rangle_{U^*,U} \to \langle f, u\rangle_{U^*,U} \;\; \forall f \in U^*$$

for the dual pairing $\langle\cdot,\cdot\rangle_{U^*,U}$ with respect to U^* and U. The weak convergence in V is defined in an analog manner.

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

$$u_k \rightharpoonup u \quad \text{in} \quad U \quad \text{with} \quad u_k \in \mathcal{D}(F) \quad \Longrightarrow \quad u \in \mathcal{D}(F) \quad \text{and} \quad F(u_k) \rightharpoonup F(u) \quad \text{in} \quad V.$$

3. *The functional Ω is convex and weakly sequentially lower semi-continuous.*

4. *We have $0 < p < 1$ for the exponent in the residual term of (1.3).*

5. *For every $\alpha > 0$, $c \geq 0$, and for the exact right-hand side $v = v^0$ of (1.1), the sets*

$$\mathcal{M}_\alpha^v(c) := \{u \in \mathcal{D} : T_\alpha^v(u) \leq c\} \tag{1.4}$$

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

We believe that assertions on existence and stability of regularized solutions u_α^δ, well-known for $p \geq 1$, can be extended to the case $0 < p < 1$, but we do not verify that here.

As obvious in Banach space theory of variational regularization errors will be measured for the convex functional Ω with subdifferential $\partial\Omega$ by means of Bregman distances

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

at $u \in \mathcal{D}(\Omega) \subseteq U$ and $\xi \in \partial\Omega(u) \subseteq U^*$. We denote the Bregman domain by

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

and consider Ω-minimizing solutions $u^\dagger \in \mathcal{D}_B(\Omega) \cap \mathcal{D}(F)$ to (1.1), which we assume to exist and for which we have

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

2 Convergence and variational inequalities

As the following proposition shows, all regularized solutions associated with data possessing a sufficiently small noise level δ belong to a common weakly pre-compact level set of type $\mathcal{M}_\alpha^v(c)$ whenever the regularization parameters $\alpha = \alpha(\delta)$ are chosen such that weak convergence to Ω-minimizing solutions $u^\dagger$ is enforced.

Proposition 2.1 *Consider an a priori choice $\alpha = \alpha(\delta) > 0$, $0 < \delta < \infty$, for the regularization parameter in (1.3) depending on the noise level δ such that*

$$\alpha(\delta) \to 0 \quad \text{and} \quad \frac{\delta^p}{\alpha(\delta)} \to 0, \quad \text{as} \quad \delta \to 0. \tag{2.1}$$

Provided that (1.1) has a solution $u \in \mathcal{D}$ then under Assumption 1.1 every sequence $\{u_n\}_{n=1}^\infty := \{u_{\alpha(\delta_n)}^{\delta_n}\}_{n=1}^\infty$ of regularized solutions corresponding to a sequence $\{v^{\delta_n}\}_{n=1}^\infty$ of data with $\lim\limits_{n\to\infty} \delta_n = 0$ has a subsequence $\{u_{n_k}\}_{k=1}^\infty$, which is weakly convergent in U, i.e. $u_{n_k} \rightharpoonup u^\dagger$, and its limit $u^\dagger$ is an Ω-minimizing solution of (1.1) with $\Omega(u^\dagger) = \lim\limits_{k\to\infty} \Omega(u_{n_k})$.

For given $\alpha_{max} > 0$ let $u^\dagger$ denote an Ω-minimizing solution of (1.1). If we set

$$\rho := \alpha_{max}(1 + \Omega(u^\dagger)) , \tag{2.2}$$

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

$$u_{\alpha(\delta)}^\delta \in \mathcal{M}_{\alpha_{max}}^v(\rho) \qquad \text{for all} \qquad 0 \leq \delta \leq \delta_{max} . \tag{2.3}$$

Proof: The proof of the first part of the proposition concerning convergence can be done as in [6, Theorem 3.26]. The second part can be proven as follows: Owing to (2.1) there exists some $\delta_{max} > 0$ such that $\alpha(\delta) \le \alpha_{max}$ and $\frac{\delta^p}{\alpha(\delta)} \le \frac{1}{2}$ for all $0 < \delta \le \delta_{max}$. Then for such δ, by writing for simplicity α instead of $\alpha(\delta)$, we have with $(a+b)^p \le a^p + b^p$ $(a, b \ge 0,\ 0 < p < 1)$ the estimate

$$T^v_{\alpha_{max}}(u^\delta_\alpha) \le \left[\left\|F(u^\delta_\alpha) - v^\delta\right\|^p_V + \delta^p + \alpha_{max}\Omega(u^\delta_\alpha)\right]$$
$$= \left[\left\|F(u^\delta_\alpha) - v^\delta\right\|^p_V + \alpha\Omega(u^\delta_\alpha) + (\alpha_{max} - \alpha)\Omega(u^\delta_\alpha) + \delta^p\right]$$
$$\le \left[T^{v^\delta}_\alpha(u^\dagger) + (\alpha_{max} - \alpha)\Omega(u^\delta_\alpha) + \delta^p\right] \le \left[\delta^p + \alpha\Omega(u^\dagger) + (\alpha_{max} - \alpha)\Omega(u^\delta_\alpha) + \delta^p\right].$$

Because of $\Omega(u^\delta_\alpha) \le \frac{\delta^p}{\alpha} + \Omega(u^\dagger)$, which is a consequence of $T^{v^\delta}_\alpha(u^\delta_\alpha) \le T^{v^\delta}_\alpha(u^\dagger)$ $(\alpha > 0)$, and with $\frac{\alpha_{max}}{\alpha} \ge 1$, $\frac{\delta^p}{\alpha} \le \frac{1}{2}$ this yields

$$T^v_{\alpha_{max}}(u^\delta_\alpha) \le \left[\delta^p + \alpha_{max}\frac{\delta^p}{\alpha} + \alpha_{max}\Omega(u^\dagger)\right] \le \left[2\alpha_{max}\frac{\delta^p}{\alpha} + \alpha_{max}\Omega(u^\dagger)\right] \le \rho$$

and hence proves (2.3). Evidently, it holds $T^v_{\alpha_{max}}(u^\dagger) = \alpha_{max}\Omega(u^\dagger) \le \alpha_{max}\Omega(u^\dagger)$ for all p under consideration. This, however, implies $u^\dagger \in \mathcal{M}^v_{\alpha_{max}}(\rho)$ and completes the proof. □

As recent publications show, for variational regularization the variational inequalities that have to hold on level sets $\mathcal{M}^v_{\alpha_{max}}(\rho)$ play an important role for obtaining convergence rates. This is also the case for $0 < p < 1$. In this context, we fix an a priori parameter choice $\alpha = \alpha(\delta)$ yielding convergence along the lines of Proposition 2.1 and use the notation around that proposition. Then we can consider variational inequalities of the form

$$\left\langle \xi, u^\dagger - u \right\rangle_{U^*,U} \le \beta_1 D_\xi(u, u^\dagger) + \beta_2 \left\|F(u) - F(u^\dagger)\right\|^\kappa_V \qquad \text{for all} \quad u \in \mathcal{M}^v_{\alpha_{max}}(\rho) \tag{2.4}$$

with some $\xi \in \partial\Omega(u^\dagger)$, two multipliers $0 \le \beta_1 < 1$, $\beta_2 \ge 0$ and an exponent $\kappa > 0$. In [3, Proposition 4.3] it was shown that only the interval $0 < \kappa \le 1$ is of real interest. Now we have the following assertion concerning the interplay of different exponents κ.

Lemma 2.2 *Let $u^\dagger \in \mathcal{D}_B(\Omega) \cap \mathcal{D}(F)$ with $\xi \in \partial\Omega(u^\dagger)$ be an Ω-minimizing solution of (1.1). If a variational inequality (2.4) is valid for some exponent $\kappa = \kappa_0 \in (0, 1]$ and two multipliers $0 \le \beta_1 < 1$, $\beta_2 = \beta_2(\kappa_0) \ge 0$, then for any smaller exponent $0 < \kappa = \kappa_1 < \kappa_0$ the inequality (2.4) also holds on the same level set, with the same β_1, but with another $\beta_2 = \beta_2(\kappa_1) \ge 0$.*

Proof: The proof is simple, since $u \in \mathcal{M}^v_{\alpha_{max}}(\rho)$ means that

$$\|F(u) - F(u^\dagger)\|^p_V + \alpha_{max}\,\Omega(u) \le \rho$$

and implies $\|F(u) - F(u^\dagger)\|_V \le \rho^{\frac{1}{p}}$ as well as

$$\beta_2(\kappa_0)\left\|F(u) - F(u^\dagger)\right\|^{\kappa_0}_V \le \beta_2(\kappa_0)\,\rho^{\frac{\kappa_0 - \kappa_1}{p}}\left\|F(u) - F(u^\dagger)\right\|^{\kappa_1}_V.$$

Setting $\beta_2(\kappa_1) := \beta_2(\kappa_0)\,\rho^{\frac{\kappa_0-\kappa_1}{p}}$ this proves the lemma. □

3 A new convergence rates result for $0 < p < 1$

Now we are ready to formulate and prove the main assertion of this note concerning convergence rates for the new exponent interval $0 < p < 1$ in variational regularization. We are based on and connect to the well-known recent results on convergence rates results for $p \geq 1$ based on the technique of variational inequalities published in [4], [6] and [2, 3]. The theorem assumes that a variational inequality (2.4) is satisfied with some exponent κ. Sufficient conditions for that assumptions are formulated in [4] and [6, Section 3.2] for $\kappa = 1$ and in [2, 3] for $0 < \kappa \leq 1$.

Theorem 3.1 *Let $u^\dagger \in \mathcal{D}_B(\Omega)$ be an Ω-minimizing solution of (1.1) with $\xi \in \partial\Omega(u^\dagger)$ and let exist constants $0 \leq \beta_1 < 1$, $\beta_2 \geq 0$, and $0 < \kappa \leq 1$ such that the variational inequality (2.4) holds with ρ from (2.2). Then under the assumptions stated above we have for the Tikhonov type regularization method (1.3) with norm exponent $0 < p < 1$ in the residual term the convergence rate*

$$D_\xi(u^\delta_{\alpha(\delta)}, u^\dagger) \;=\; \mathcal{O}\left(\delta^\mu\right) \quad as \quad \delta \to 0 \tag{3.1}$$

for all rate exponents $0 < \mu < 1$ satisfying the condition

$$\mu < p \quad if \quad p \leq \kappa \qquad and \qquad \mu = \kappa \quad if \quad \kappa < p \tag{3.2}$$

whenever we use an a priori parameter choice $\alpha(\delta) \asymp \delta^{p-\mu}$.

Proof: By Lemma 2.2 we obtain from (2.4) that such a variational inequality with some $0 \leq \beta < 1$ and $\beta_2 \geq 0$ is also valid for all exponents $0 < \mu < 1$ satisfying the condition (3.2) which implies $\mu < p$. We write again for simplicity α instead of $\alpha(\delta)$ and note that the parameter choice rule $\alpha \asymp \delta^{p-\mu}$ satisfies the condition (2.1) with the consequence that Proposition 2.1 is applicable. Then by using $T^{v^\delta}_\alpha(u^\delta_\alpha) \leq T^{v^\delta}_\alpha(u^\dagger)$ and (1.2) we can estimate as follows:

$$\left\|F(u^\delta_\alpha) - v^\delta\right\|^p_V + \alpha D_\xi(u^\delta_\alpha, u^\dagger) \leq \delta^p + \alpha\left(\Omega(u^\dagger) - \Omega(u^\delta_\alpha) + D_\xi(u^\delta_\alpha, u^\dagger)\right) . \tag{3.3}$$

Moreover, by exploiting the inequality $(a+b)^\mu \leq a^\mu + b^\mu$ $(a, b > 0,\ 0 < \mu \leq 1)$ because of (2.3) we obtain from the variational inequality (2.4) that

$$\begin{aligned} \Omega(u^\dagger) - \Omega(u^\delta_\alpha) + D_\xi(u^\delta_\alpha, u^\dagger) &= -\left\langle \xi, u^\delta_\alpha - u^\dagger \right\rangle_{U^*,U} \\ \leq \beta_1\, D_\xi(u^\delta_\alpha, u^\dagger) \;+\; \beta_2 &\left\|F(u^\delta_\alpha) - F(u^\dagger)\right\|^\mu_V \\ \leq \beta_1\, D_\xi(u^\delta_\alpha, u^\dagger) \;+\; \beta_2 &\left(\left\|F(u^\delta_\alpha) - v^\delta\right\|^\mu_V + \delta^\mu\right) . \end{aligned}$$

Therefore from (3.3) it follows that

$$\left\|F(u^\delta_\alpha) - v^\delta\right\|^p_V + \alpha D_\xi(u^\delta_\alpha, u^\dagger) \leq \delta^p + \alpha\left(\beta_1 D_\xi(u^\delta_\alpha, u^\dagger) \;+\; \beta_2\left(\left\|F(u^\delta_\alpha) - v^\delta\right\|^\mu_V + \delta^\mu\right)\right). \tag{3.4}$$

Using the variant

$$a\, b \;\leq \varepsilon\, a^{p_1} + \frac{b^{p_2}}{(\varepsilon\, p_1)^{p_2/p_1} p_2} \qquad (a, b \geq 0,\ \ \varepsilon > 0\ \ p_1, p_2 > 1 \ \text{ with } \ \frac{1}{p_1} + \frac{1}{p_1} = 1) \tag{3.5}$$

of Young's inequality twice with $\varepsilon := 1$, $p_1 := p/\mu$, $p_2 := p/(p-\mu)$ and $b := \alpha\beta_2$, on the one hand with $a := \|F(u_\alpha^\delta) - u^\dagger\|_V^\mu$ and on the other hand with $a := \delta^\mu$, the inequality

$$\alpha D_\xi(u_\alpha^\delta, u^\dagger) \le 2\delta^p + \alpha\beta_1 D_\xi(u_\alpha^\delta, u^\dagger) + \frac{2(p-\mu)}{(p/\mu)^{\mu/(p-\mu)}\,p}(\alpha\beta_2)^{p/(p-\mu)}$$

follows from (3.4). Because of $0 \le \beta_1 < 1$ this provides us with the estimate

$$D_\xi(u_\alpha^\delta, u^\dagger) \le \frac{2\delta^p + \frac{2(p-\mu)}{(p/\mu)^{\mu/(p-\mu)}\,p}\,(\alpha\,\beta_2)^{p/(p-\mu)}}{\alpha\,(1-\beta_1)} \tag{3.6}$$

for sufficiently small $\delta > 0$, which yields (3.1) for the a priori parameter choice $\alpha \asymp \delta^{p-\mu}$ and proves the theorem. □

The theorem and its proof indicate that there is for the variational inequality approach of the Tikhonov type method a caesura at exponent $p = 1$. For $p \ge 1$ there occur Hölder convergence rates mit rate exponent κ when (2.4) is fulfilled. For $p < 1$, however, the approach produces an artificial limitation of the convergence rate when $p \le \kappa$. This seems to be a drawback for the method. This caesura can also be seen by considering the a priori parameter choice $\alpha \asymp \delta^{p-\mu}$. If μ approaches p from below the decay of quotient $\frac{\delta}{\alpha(\delta)} = \delta^{p-\mu} \to 0$ gets very slow and in the limit situation $\mu = p$ the required convergence condition (2.1) is violated. On the other hand, Young's inequality (3.5) can only be applied with exponents $p_1, p_2 > 1$ which requires $\mu < p$ at all.

In [1] J. Geissler shows that the same ceasura also occurs in the variant

$$T_\alpha^{v^\delta}(u) := \mathcal{S}(F(u), v^\delta)^p + \alpha\,\Omega(u) \quad \to \min$$

of variational regularization with general residual term suggested by C. Pöschl in [5], where $\mathcal{S}(F(u), v^\delta)$ replaces the norm $\|F(u) - v^\delta\|_V$ and the inequality $\mathcal{S}(F(u^\dagger), v^\delta) \le \delta$ defines the noise model.

We conjecture that such effects can appear whenever the residual term has a non-convex structure, for example for terms $f(\|F(u) - v^\delta\|_V)$ with $f(t)$ $(t \ge 0)$ a concave increasing function and $f(0) = 0$.

References

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

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

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

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

[5] Pöschl, C. (2008): *Tikhonov Regularization with General Residual Term*. Dissertation. Innsbruck: Leopold Franzens Universität.

[6] Scherzer, O.; Grasmair, M.; Grossauer, H.; Haltmeiner, M.; Lenzen, F. (2009): *Variational Methods in Imaging*. New York: Springer.

[7] Zarzer, C.A. (2009): On Tikhonov regularization with non-convex sparsity constraints. *Inverse Problems* **25**, 025006 (13pp).