Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

Enhanced convergence rates for Tikhonov regularization revisited: improved results

A. Neubauer, T. Hein, B. Hofmann, S. Kindermann, U. Tautenhahn

RICAM-Report 2009-14

Enhanced convergence rates for Tikhonov regularization revisited: improved results

Andreas Neubauer[1], Torsten Hein[2], Bernd Hofmann[2], Stefan Kindermann[1], and Ulrich Tautenhahn[3]

[1]: Industrial Mathematics Institute, Johannes Kepler University Linz, A-4040 Linz, Austria,

[2]: Department of Mathematics, Chemnitz University of Technology, D-09107 Chemnitz, Germany

[3]: Department of Mathematics, University of Applied Sciences Zittau/Görlitz, P.O.Box 1454, D-02754 Zittau, Germany

Abstract. In this paper, we are going to improve the enhanced convergence rates for Tikhonov regularization of nonlinear ill-posed problems in Banach spaces presented by Neubauer in [14]. The new message is that rates are shown to be independent of the residual norm exponents $1 \le p < \infty$ in the functional to be minimized for obtaining regularized solutions. However, on the one hand the smoothness of the image space influences the rates, and on the other hand best possible rates require specific choices of the regularization parameters $\alpha > 0$. In the limiting case $p = 1$, the α-values must not tend to zero as the noise level decreases, but has to converge to a fixed positive value characterized by properties of the solution.

1. Introduction

We consider nonlinear ill-posed problems

$$F(x) = y\,, \tag{1}$$

where $F : \mathcal{D}(F) \subset X \to Y$ is a nonlinear operator mapping between Banach or Hilbert spaces X and Y. For the practical treatment of noisy data y^δ of y with

$$\|y - y^\delta\| \le \delta$$

and noise level $\delta > 0$ the stable approximate solution of (1) requires regularization methods. Our focus is on variational regularization, where regularized solutions are obtained by minimizing the Tikhonov type functional

$$\mathcal{T}_\alpha^\delta := \tfrac{1}{p}\|F(x) - y^\delta\|^p + \alpha R(x)\,, \qquad \alpha > 0\,, \tag{2}$$

with minimizers x_α^δ.

In Hilbert spaces and for $p := 2$, $R(x) := \|x - x^*\|^2$ rates

$$\|x_\alpha^\delta - x^\dagger\| = \mathcal{O}(\delta^{\frac{1}{2}}) \quad \text{as} \quad \delta \to 0$$

for the convergence of regularized solutions x_α^δ to the exact solution $x^\dagger$ of (1) were already proven in [5] (cf. [4, Chapter 10]) under a source condition

$$x^\dagger - x^* = F'(x^\dagger)^* w, \quad w \in Y\,,$$

and some additional assumptions. Convergence rates up to such order characterize the *low rate world* and corresponding rate proofs can be performed by using the ansatz

$$\mathcal{T}_\alpha^\delta(x_\alpha^\delta) \le \mathcal{T}_\alpha^\delta(x^\dagger)\,. \tag{3}$$

In recent papers progress in this world including extensions to Banach spaces was achieved by using variational inequalities (see [16] and references therein).

On the other hand, the *enhanced rate world* showing higher convergence rates up to

$$\|x_\alpha^\delta - x^\dagger\| = \mathcal{O}(\delta^{\frac{2}{3}}) \quad \text{as} \quad \delta \to 0$$

under stronger source conditions up to

$$x^\dagger - x^* = F'(x^\dagger)^* F'(x^\dagger)\, v\,, \quad v \in X\,,$$

was entered for nonlinear problems with the paper [12], where rate proofs were obained using ansatz (3) and the first order optimality conditions for a minimizer of the Tikhonov functional. Some appropriate alternative ansatz for obtaining error estimates in the enhanced rate world is

$$\mathcal{T}_\alpha^\delta(x_\alpha^\delta) \le \mathcal{T}_\alpha^\delta(x^\dagger - z)$$

with appropriately chosen element $z \in X$. This idea can already be found in [11] and has been used in different works, see, e.g., [4, 17, 18] for the Hilbert space and [8, 14] for the Banach space setting. In the present paper, recent results from [14] on enhanced convergence rates are revisited and improved.

2. Preliminaries

The existence of minimizers x_α^δ of (2) acting as regularized solutions to (1) as well as their stability maybe guaranteed under the following conditions on X, Y, F, $\mathcal{D}(F)$, and R (see also [14] and [10, 16]) that we will assume throughout this paper. If the condition

$$\alpha(\delta) \to 0 \quad \text{and} \quad \frac{\delta^p}{\alpha(\delta)} \to 0 \qquad \text{as} \quad \delta \to 0 \tag{4}$$

is satisfied, then we even have (weak) convergence of regularized solutions (compare [16, Theorem 3.26]).

(A1) X and Y are reflexive Banach spaces with norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively, which define the strong convergence in that spaces. We will omit space indices whenever it is clear from the context what norm is meant. X^* and Y^* denote the dual spaces of X and Y with dual forms $\langle\cdot,\cdot\rangle_{Y^*,Y}$ and $\langle\cdot,\cdot\rangle_{X^*,X}$, respectively, that allow us to define the corresponding weak convergence. Again we omit space indices.

(A2) Y is s-smooth for some $s > 1$, i.e., for the modulus of smoothness $\rho_Y : [0,\infty] \to \mathbb{R}$ the estimate

$$\rho_Y(\tau) := \tfrac{1}{2}\sup\{\,\|y+\bar{y}\| + \|y-\bar{y}\| - 2 : \|y\| = 1, \|\bar{y}\| \le \tau\} \le c_s\tau^s$$

holds for some $c_s > 0$ and all $\tau \ge 0$.

(A3) The exponent p in (2) is in the interval $[1,\infty)$.

(A4) The functional $R : \mathcal{D}(R) \subset X \to [0,\infty]$ is convex and weakly sequentially lower semi-continuous.

(A5) The operator $F : \mathcal{D}(F) \subset X \to Y$ is weakly sequentially closed and the domain $\mathcal{D}(F)$ is also weakly sequentially closed.

(A6) $\mathcal{D} := \mathcal{D}(F) \cap \mathcal{D}(R) \neq \emptyset$.

(A7) Let $x^\dagger \in \mathcal{D}$ be an R-minimizing solution of equation (1), i.e., $R(x^\dagger) = \min\{R(x) : F(x) = y\}$, which exists due to [16, Theorem 3.25], and let F be Gâteaux-differentiable in $x^\dagger$. Moreover, we assume that the subdifferential $\partial R(x^\dagger)$ consists of a single element $dR(x^\dagger) \in X^*$.

(A8) The level sets $\mathcal{M}_\alpha(M) := \{x \in \mathcal{D} : \frac{1}{p}\|F(x) - y\|_Y^p + \alpha R(x) \leq M\}$ are weakly sequentially compact for all $\alpha, M > 0$.

(A9) There is an exponent $r > 1$ and constants $c_r > 0, \rho_R > 0$ such that

$$D_R(x, x^\dagger) \leq c_r \|x - x^\dagger\|^r$$

for all $x \in \mathcal{D}(R)$ with $\|x - x^\dagger\| \leq \rho_R$. Here, D_R denotes the Bregman distance defined by (see, e.g., [16])

$$D_R(x, x^\dagger) := R(x) - R(x^\dagger) - \langle\, dR(x^\dagger), x - x^\dagger \,\rangle.$$

(A10) There are constants $c_F \geq 0$ and $\rho_F > 0$ such that

$$\|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\| \leq c_F D_R(x, x^\dagger)$$

for all $x \in \mathcal{D}$ with $R(x) \leq R(x^\dagger) + \rho_F$ and $\|F(x) - F(x^\dagger)\| \leq \rho_F$.

Note that Assumption (A10) (also exploited, e.g., in [15, Assumption 3.1], [10, formula (16)] and [9, Def. 2.5, case $c_1 = 0$, $c_2 = 1$]) expresses the nonlinearity behaviour of F in a neighbourhood of $x^\dagger$.

Due to (A2), the function $f_p(y) := \frac{1}{p}\|y\|^p$, $y \in Y$, $p > 1$, is strictly convex and Fréchet-differentiable. Its derivative $J_p := f_p'$ is the so called duality mapping of Y with gauge function $t \mapsto t^{p-1}$. J_p is continuous and surjective from $Y \to Y^*$ and

$$J_p(\lambda y) = |\lambda|^{p-1} \mathrm{sgn}\,(\lambda)\, J_p(y)\,, \qquad \lambda \in \mathbb{R}\,. \tag{5}$$

(see [3, Chap. I+II]). The Bregman distance of f_p is defined by

$$D_p(y, \bar{y}) := \tfrac{1}{p}\|y\|^p - \tfrac{1}{p}\|\bar{y}\|^p - \langle\, J_p(\bar{y}), y - \bar{y} \,\rangle\,, \quad y, \bar{y} \in Y\,.$$

It always holds that $D_p(y, \bar{y}) \geq 0$.

It was discovered independently by the first two authors and published in [7, 8, 14] that one obtains the convergence rate

$$D_R(x_\alpha^\delta, x^\dagger) = \mathcal{O}(\delta^{\frac{rs}{r+s-1}}) \qquad \text{if} \quad p \geq s \tag{6}$$

and

$$D_R(x_\alpha^\delta, x^\dagger) = \mathcal{O}(\delta^{\frac{rp}{r+p-1}}) \qquad \text{if} \quad 1 < p < s\,,$$

whenever $x^\dagger$ satisfies the strong source condition

$$dR(x^\dagger) = F'(x^\dagger)^{\#} J_p(F'(x^\dagger)v) \quad \text{for some} \quad v \in X\,, \tag{7}$$

and $\alpha(\delta)$ is chosen appropriately. Here $A^{\#} : Y^* \to X^*$ denotes the Banach space adjoint of a bounded linear operator $A : X \to Y$. Proofs of this result are essentially based on an estimate for $D_p(y, \bar{y})$ that was derived using results in [20]. It turns out that this estimate can be improved if $\bar{y} \neq 0$. Moreover, it can be extended to the case $p = 1$, which was already mentioned in [8]. With this improved estimate one can show that the rate in (6) also holds uniformly for all $1 \leq p < s$ provided that $F'(x^\dagger)v \neq 0$, i.e., *the rate is independent of* p.

For the extension to the case $p = 1$ we need the following consideration: noting that

$$J_p(y) = \|y\|^{p-2} J_2(y) \tag{8}$$

for $p > 1$, we may conclude that

$$D_p(y, \bar{y}) := \tfrac{1}{p}\|y\|^p - \tfrac{1}{p}\|\bar{y}\|^p - \|\bar{y}\|^{p-2}\langle J_2(\bar{y}), y - \bar{y}\rangle \,.$$

If $\bar{y} \neq 0$, then the limit $p \to 1$ exists and we may define:

$$D_1(y, \bar{y}) := \|y\| - \|\bar{y}\| - \|\bar{y}\|^{-1}\langle J_2(\bar{y}), y - \bar{y}\rangle \,, \quad \bar{y} \neq 0 \,. \tag{9}$$

We are now in the position to prove the following estimate:

Lemma 1 *Let Assumption (A2) hold. Then for some $c_s > 0$ the estimate*

$$D_p(y, \bar{y}) \leq c_s \|\bar{y}\|^{p-s} \|y - \bar{y}\|^s$$

holds if $\bar{y} \neq 0$ and if $1 \leq p \leq s$.

Proof. Let $p > 1$. Using [20, Theorem 2] we obtain the estimate

$$D_p(y, \bar{y}) \leq c_p \int_0^1 t^{-1} \max\{\|\bar{y} + t(y - \bar{y})\|, \|\bar{y}\|\}^p \cdot \rho_Y(t\|y - \bar{y}\| \max\{\|\bar{y} + t(y - \bar{y})\|, \|\bar{y}\|\}^{-1})\, dt$$

for some $c_p > 0$. An inspection of the proof of [20, Theorem 2] shows that c_p may be bounded independently of p for all $1 < p < s$. Thus, the assertion immediately follows also for $p = 1$ using (A2), definition (9), and noting that $\max\{\|\bar{y} + t(y - \bar{y})\|, \|\bar{y}\|\} \geq \|\bar{y}\|$. ∎

3. Convergence rates

Using the notation of [14, Theorem 3] we first present the improved result. In the proof we only show the steps that differ from the one in [14]:

Theorem 2 *Let the assumptions (A1) – (A10) hold. Furthermore, let $p > 1$ and assume that $x^\dagger$ satisfies condition (7). Moreover, assume that $z_\alpha := x^\dagger - \alpha^{\frac{1}{p-1}} v \in \mathcal{D}$ for all $0 < \alpha \leq \bar{\alpha}$ with v as in (7), that $c_F\|w\| < 1$ with $w := J_p(F'(x^\dagger)v)$, and that $F'(x^\dagger)v \neq 0$.*

Then the Tikhonov regularized solutions converge with the rate

$$D_R(x_\alpha^\delta, x^\dagger) = \mathcal{O}(\delta^{\frac{rs}{r+s-1}}) \qquad as \quad \delta \to 0 \tag{10}$$

if the regularization parameter is chosen as

$$\alpha \sim \delta^{(p-1)\frac{s}{r+s-1}} \,. \tag{11}$$

Proof. Obviously, the result immediately follows with [14, Theorem 3] for the case $p \geq s$. Let us now assume that $1 < p < s$ and that $\alpha = \alpha(\delta)$ satisfies (4). From the proof of [14, Theorem 3] we can exploit the results

$$D_R(x_\alpha^\delta, x^\dagger) = \mathcal{O}\Big(\alpha^{-1} D_p(F(z_\alpha) - y^\delta, b_\alpha) + D_R(z_\alpha, x^\dagger)\Big) \quad \text{as} \quad \delta \to 0 \,,$$

$$D_R(z_\alpha, x^\dagger) = \mathcal{O}(\alpha^{\frac{r}{p-1}}) \quad \text{as} \quad \alpha \to 0$$

and

$$\|F(z_\alpha) - y^\delta - b_\alpha\| \leq \delta + c_F D_R(z_\alpha, x^\dagger)$$

with $b_\alpha := -\alpha^{\frac{1}{p-1}} F'(x^\dagger)v$ and $\delta > 0$ sufficiently small. Now Lemma 1 implies that

$$D_p(F(z_\alpha) - y^\delta, b_\alpha) = \mathcal{O}\Big(\|b_\alpha\|^{p-s} (\delta^s + D_R(z_\alpha, x^\dagger)^s) \Big) \quad \text{as} \quad \delta \to 0\,.$$

Combining all estimates and noting that $rs - s + 1 > r$ yields

$$\begin{aligned} D_R(x_\alpha^\delta, x^\dagger) &= \mathcal{O}(\alpha^{-\frac{s-1}{p-1}} \delta^s + \alpha^{\frac{rs-s+1}{p-1}} + \alpha^{\frac{r}{p-1}}) \\ &= \mathcal{O}(\alpha^{-\frac{s-1}{p-1}} \delta^s + \alpha^{\frac{r}{p-1}}) \quad \text{as} \quad \delta \to 0\,. \end{aligned}$$

Now the a-priori parameter choice (11) implies the assertion. ∎

Of course all other convergence rates results in [14] may be improved in the same way, i.e, $\min\{p, s\}$ may always be replaced by s.

We will now turn to the case $p = 1$. Before doing so, we want to mention that, due to (8), an element $x^\dagger$ satisfying source condition (7) for some $p > 1$ will also satisfy the source condition with $p = 2$, however, with a scaled element v and vice versa, i.e., if $x^\dagger$ satisfies condition

$$dR(x^\dagger) = F'(x^\dagger)^\# J_2(F'(x^\dagger)\bar{v}) \quad \text{for some} \quad \bar{v} \in X\,, \tag{12}$$

then it also satisfies condition (7) for any $p > 1$ with $v = \|F'(x^\dagger)\bar{v}\|^{\frac{2-p}{p-1}} \bar{v}$. Note that for $p \to 1$ we have the limiting relations $\|v\| \to 0$ in case $\|F'(x^\dagger)\bar{v}\| < 1$ and $\|v\| \to \infty$ in case $\|F'(x^\dagger)\bar{v}\| > 1$.

To be able to prove enhanced rates for the case $p = 1$, condition (A10) has to hold with a constant $\rho_F > 0$ sufficiently large, namely

$$\|F'(x^\dagger)\bar{v}\|^{-1} R(x^\dagger) < \rho_F \tag{13}$$

with $\bar{v}$ as in (12).

Theorem 3 *Let the assumptions (A1) – (A10) and* (13) *hold. Furthermore, let* $p = 1$ *and assume that* $x^\dagger$ *satisfies condition* (12)*. Moreover, assume that* $z_\kappa := x^\dagger - \kappa\bar{v} \in \mathcal{D}$ *for all* $0 < \kappa \le \bar{\kappa}$ *with* $\bar{v}$ *as in* (12)*, that* $c_F\|w\| < 1$ *with* $w := J_2(F'(x^\dagger)\bar{v})$*, and that* $F'(x^\dagger)\bar{v} \neq 0$.

Then the Tikhonov regularized solutions converge with the rate (10) *if the regularization parameter is chosen as* $\alpha = \|F'(x^\dagger)\bar{v}\|^{-1}$.

Proof. We set $b_\kappa := -\kappa F'(x^\dagger)\bar{v}$. Then, due to (5), it holds that

$$J_2(b_\kappa) = -\kappa J_2(F'(x^\dagger)\bar{v}) = -\kappa\, w\,. \tag{14}$$

Since x_α^δ is a minimizer of the Tikhonov functional (2), it holds that

$$\|F(x_\alpha^\delta) - y^\delta\| + \alpha R(x_\alpha^\delta) \le \|F(z_\kappa) - y^\delta\| + \alpha R(z_\kappa)$$

Using (12), (14), $\kappa\|b_\kappa\|^{-1} = \|F'(x^\dagger)\bar{v}\|^{-1} = \alpha$, and setting

$$\bar{r}_\kappa := F(z_\kappa) - F(x^\dagger) - F'(x^\dagger)(z_\kappa - x^\dagger)\,, \tag{15}$$

$$r_\alpha^\delta := F(x_\alpha^\delta) - F(x^\dagger) - F'(x^\dagger)(x_\alpha^\delta - x^\dagger)\,, \tag{16}$$

this estimate is equivalent to

$$\begin{aligned} &D_1(F(x_\alpha^\delta) - y^\delta, b_\kappa) + \alpha D_R(x_\alpha^\delta, x^\dagger) \\ &\quad\le D_1(F(z_\kappa) - y^\delta, b_\kappa) + \alpha D_R(z_\kappa, x^\dagger) \\ &\qquad + \|b_\kappa\|^{-1} \langle J_2(b_\kappa), F(z_\kappa) - F(x_\alpha^\delta) \rangle + \alpha \langle dR(x^\dagger), z_\kappa - x_\alpha^\delta \rangle \\ &\quad= D_1(F(z_\kappa) - y^\delta, b_\kappa) + \alpha D_R(z_\kappa, x^\dagger) + \alpha \langle w, r_\alpha^\delta - \bar{r}_\kappa \rangle\,, \end{aligned}$$

where D_1 is defined as in (9). Due to (13) and $\|F(x_\alpha^\delta) - y^\delta\| + \alpha R(x_\alpha^\delta) \le \delta + \alpha R(x^\dagger)$, it follows that $R(x_\alpha^\delta) \le R(x^\dagger) + \rho_F$ and $\|F(x_\alpha^\delta) - F(x^\dagger)\| \le \rho_F$ for $\delta > 0$ sufficiently small. Thus, (A10) is applicable and we obtain that

$$\begin{aligned} &D_1(F(x_\alpha^\delta) - y^\delta, b_\kappa) + \alpha\,(1 - c_F\,\|w\|)\,D_R(x_\alpha^\delta, x^\dagger) \\ &\quad\le D_1(F(z_\kappa) - y^\delta, b_\kappa) + \alpha\,(1 + c_F\,\|w\|)\,D_R(z_\kappa, x^\dagger) \end{aligned}$$

for $\delta, \kappa > 0$ sufficiently small. Assuming that

$$\kappa = \kappa(\delta) \to 0 \qquad \text{as} \quad \delta \to 0\,,$$

we, therefore, get the estimate

$$D_R(x_\alpha^\delta, x^\dagger) = \mathcal{O}\Big(D_1(F(z_\kappa) - y^\delta, b_\kappa) + D_R(z_\kappa, x^\dagger)\Big) \quad \text{as} \quad \delta \to 0\,. \tag{17}$$

Using the special setting of z_κ and b_κ and noting that (A10) and (15) yield that

$$\|F(z_\kappa) - y^\delta - b_\kappa\| = \|y - y^\delta + \bar{r}_\kappa\| \le \delta + c_F D_R(z_\kappa, x^\dagger)$$

for $\delta > 0$ sufficiently small, we obtain together with (A9) and Lemma 1 (case $p = 1$) that

$$\begin{aligned} D_R(x_\alpha^\delta, x^\dagger) &= \mathcal{O}\Big(\,\|b_\kappa\|^{\,1-s}\,\|F(z_\kappa) - y^\delta - b_\kappa\|^{\,s} + D_R(z_\kappa, x^\dagger)\Big) \\ &= \mathcal{O}\Big(\kappa^{1-s}(\delta^s + D_R(z_\kappa, x^\dagger)^s) + D_R(z_\kappa, x^\dagger)\Big) \\ &= \mathcal{O}\Big(\kappa^{1-s}\delta^s + \kappa^r\Big) \qquad \text{as} \quad \delta \to 0\,. \end{aligned}$$

Choosing $\kappa = \delta^{\frac{s}{r+s-1}}$ yields the desired result. ∎

An inspection of the proof above shows that the result of Theorem 3 remains valid if the constant parameter choice $\alpha = \alpha_0 := \|F'(x^\dagger)\bar{v}\|^{-1}$ is replaced by an a-priori choice $\alpha(\delta)$ with

$$|\alpha(\delta) - \alpha_0| \le \mathcal{O}(\delta^{\frac{rs}{r+s-1}})\,. \tag{18}$$

Nevertheless, $\alpha(\delta)$ has to converge towards a number $\alpha_0 > 0$ that depends on $x^\dagger$ and is, therefore, not known. However, as Proposition 4 below shows, for $p = 1$ enhanced rates can, in general, only be obtained if (18) holds.

4. Discussion

4.1. Interpretation of limiting cases

From Theorems 2 and 3 we find for all $p \ge 1$ that in case of appropriate parameter choices and under the assumption that condition (A9) holds with $r = 2$ the rate

$$D_R(x_\alpha^\delta, x^\dagger) = \mathcal{O}(\delta^{\frac{2s}{s+1}}) \qquad \text{as} \quad \delta \to 0 \tag{19}$$

can be established. Its rate exponent grows with the smoothness $s \in (0, 2]$ of the space Y from 0 to 4/3. The maximal exponent 4/3 characterizes the limit situation $s = 2$ of a Hilbert space Y.

If both X and Y are Hilbert spaces and if $R(x) = \|x - x^*\|^2$, then $D_R(x, x^\dagger) = \|x - x^\dagger\|^2$, i.e., (A9) is then satisfied with $r = 2$. In this case the theorems above imply for all $p \ge 1$ that

$$\|x_\alpha^\delta - x^\dagger\| = \mathcal{O}(\delta^{\frac{2}{3}}) \qquad \text{as} \quad \delta \to 0\,.$$

A well-known saturation result (see [13]) shows that for bounded linear operators F this rate cannot be improved.

We emphasize that all rate results discussed here including the limiting rate with exponent 4/3 in the Bregman case and 2/3 in the Hilbert space case with $R(x) = \|x - x^*\|^2$ do not depend on $p \in [1, \infty)$. However, it is an open problem whether such limiting rates can also be proven if $0 < p < 1$. Convergence rate assertions for that p-interval for the low rate world were made in [6] using variational inequalities.

4.2. Discussion of different parameter choices

For $p > 1$ in the low rate world of Banach space theory established by variational inequalities (see [9]) the occurring convergence rates $D_R(x_\alpha^\delta, x^\dagger) = \mathcal{O}(\delta^\kappa)$ as $\delta \to 0$ can, in general, be obtained by using the regularization parameter as $\alpha \sim \delta^{p-\kappa}$, where $0 < \kappa \le 1$ expresses the solution smoothness of $x^\dagger$ and the structure of nonlinearity of F in a compressed form. This parameter choice always satisfies the general convergence condition (4). In particular for the limiting case $\kappa = 1$ analyzed comprehensively in [16] we have $\alpha \sim \delta^{p-1}$. The enhanced rates of Theorem 2, however, require $\alpha \sim \delta^{(p-1)\frac{s}{r+s-1}}$ with smaller exponent, i.e., the decay of $\alpha(\delta) \to 0$ as $\delta \to 0$ is slower, but condition (4) is also satisfied.

For $p = 1$ we have *exact penalization* whenever $\alpha > 0$ is chosen sufficiently small. The consequences were discussed in [2] yielding also the convergence rate $D_R(x_\alpha^\delta, x^\dagger) = \mathcal{O}(\delta)$ as $\delta \to 0$, but there the regularization parameter has to be fixed as a sufficiently small value $\alpha(\delta) := \alpha_0 > 0$. Then condition (4) fails to be satisfied, but at least $\delta/\alpha(\delta) \to 0$ as $\delta \to 0$. On the other hand, for getting the rate $D_R(x_\alpha^\delta, x^\dagger) = \mathcal{O}(\delta)$ the value α_0 can be arbitrarily small, but this may lead to exploding rate constants.

In Theorem 3 we showed that the optimal rate for $p = 1$ comes with the prescribed $\alpha_0 := \|F'(x^\dagger)\bar{v}\|^{-1}$. Here, at least in the worst case, there is no α_1 independent of δ with $0 < \alpha_1 < \alpha_0$ such that this maximum rate is also obtained for $\alpha(\delta) := \alpha_1$ as the following considerations show:

Let X and Y be Hilbert spaces and assume that $R(x) := \frac{1}{2}\|x\|^2$ and that $F(x) := Ax$, where $A : X \to Y$ is a bounded linear operator with unbounded pseudoinverse $A^\dagger$, i.e., $\mathcal{R}(A)$ is not closed. The regularized solution x_α^δ is then the minimizer of the functional

$$\|Ax - y^\delta\| + \alpha\tfrac{1}{2}\|x\|^2, \qquad \alpha > 0. \tag{20}$$

Let us assume that $Ax_\alpha^\delta - y^\delta \neq 0$. Then the mapping $x \mapsto \|Ax - y^\delta\|$ is differentiable in $x = x_\alpha^\delta$ and the optimality conditions for (20) yield

$$\frac{1}{\|Ax_\alpha^\delta - y^\delta\|} A^*(Ax_\alpha^\delta - y^\delta) + \alpha x_\alpha^\delta = 0$$

or equivalently

$$A^*(Ax_\alpha^\delta - y^\delta) + \alpha\|Ax_\alpha^\delta - y^\delta\| x_\alpha^\delta = 0\,.$$

Setting $\beta := \alpha\|Ax_\alpha^\delta - y^\delta\|$ and $\bar{x}_\beta^\delta := x_\alpha^\delta$, this shows that

$$(A^*A + \beta I)\,\bar{x}_\beta^\delta = A^*y^\delta\,.$$

Thus, x_α^δ is equal to a standard Tikhonov regularized solution $\bar{x}_\beta^\delta$, where $\beta > 0$ satisfies the condition

$$\alpha^{-1} = \beta^{-1}\|A\bar{x}_\beta^\delta - y^\delta\| = \|(AA^* + \beta I)^{-1}y^\delta\| =: f_\beta(y^\delta)\,. \tag{21}$$

Noting that $\lim_{\beta\to\infty} f_\beta(y^\delta) = 0$ and that

$$\lim_{\beta\to 0^+} f_\beta(y^\delta) = \begin{cases} \|(AA^*)^\dagger y^\delta\| & \text{if} \quad Qy^\delta \in \mathcal{R}(AA^*)\,, \\ \infty & \text{else}\,, \end{cases}$$

we may conclude together with the monotonicity of $f_\beta(y^\delta)$ that

$$x_\alpha^\delta = \begin{cases} A^\dagger y^\delta \;\text{ if }\; Qy^\delta \in \mathcal{R}(AA^*) \;\text{ and }\; \alpha^{-1} \geq \|(AA^*)^\dagger y^\delta\|\,, \\ (A^*A + \beta I)^{-1} A^* y^\delta \;\text{ with }\; \beta \;\text{ solving (21) else }\,. \end{cases} \tag{22}$$

Here, Q denotes the orthogonal projector onto $\overline{\mathcal{R}(A)}$.

We are now in the position to prove that the convergence rate in Theorem 3 can only be obtained if $\alpha(\delta) \to \|A\bar{v}\|^{-1}$.

Proposition 4 *Let X, Y, $R(x)$, and A be as above and let x_α^δ be the minimizer of* (20). *Moreover, we assume that a sequence $\{\lambda_k\}$ exists in the spectrum of AA^* satisfying*

$$\lambda_k \searrow 0 \qquad \textit{and} \qquad \frac{\lambda_k}{\lambda_{k+1}} \leq C \quad \textit{for all} \quad k \in \mathbb{N} \tag{23}$$

for some $C \geq 1$.

If $\alpha = \alpha(\delta)$ is an a-priori parameter choice rule such that

$$\sup\{\,\|x_\alpha^\delta - x^\dagger\| \,:\, \|y^\delta - y\| \leq \delta\} = o(\delta^{\frac{1}{2}}) \tag{24}$$

and if $x^\dagger = A^*w$ with $w \in \overline{\mathcal{R}(A)}\backslash\{0\}$. Then

$$\lim_{\delta\to 0} \alpha(\delta) = \|w\|^{-1}. \tag{25}$$

Proof. Since we consider worst case estimates, we may choose the data y^δ as we like. Similar as in [13], we suggest the choice

$$y^\delta := AA^*w + \varepsilon AA^*G_k z\,, \tag{26}$$

where $\varepsilon > 0$ is a fixed number to be chosen later,

$$z := \begin{cases} w\|G_k w\|^{-1} & \text{if } \; G_k w \neq 0\,, \\ \text{arbitrary with } \|G_k z\| = 1 & \text{otherwise}\,, \end{cases} \tag{27}$$

and

$$G_k := F_{\frac{3}{2}\lambda_{k+1}} - F_{\frac{1}{2}\lambda_{k+1}}\,. \tag{28}$$

Here, $\{F_\lambda\}$ denotes a spectral family of AA^* and k is chosen such that

$$\varepsilon\frac{3}{2}\lambda_{k+1} \leq \delta \leq \varepsilon\frac{3}{2}\lambda_k\,, \tag{29}$$

which is always possible for $\delta > 0$ sufficiently small. (26) – (29) immediately imply that

$$\|y - y^\delta\| = \varepsilon\|AA^*G_k z\| \leq \varepsilon\frac{3}{2}\lambda_{k+1} \leq \delta\,,$$

and hence the data are feasible, and that

$$\begin{aligned} \|(AA^* + \beta I)^{-1}y^\delta\|^2 &= \|(AA^* + \beta I)^{-1}AA^*w\|^2 \\ &\quad + \|(AA^* + \beta I)^{-1}AA^*G_k z\|^2(\varepsilon^2 + 2\varepsilon\|G_k w\|) \qquad (30) \\ &\quad \to \|(AA^*)^\dagger y^\delta\|^2 = \|w\|^2 + \varepsilon^2 + 2\varepsilon\|G_k w\| \quad \text{as} \quad \beta \to 0\,, \qquad (31) \end{aligned}$$

First we assume that

$$\limsup_{\delta\to 0} \alpha(\delta) > \|w\|^{-1}. \tag{32}$$

Then there is a sequence $\delta_n > 0$ such that $\lambda_1 \geq \delta_n \to 0$ and $\alpha_n := \alpha(\delta_n) \to \bar{\alpha} > \|w\|^{-1}$ as $n \to \infty$. We choose the data y^{δ_n} as in (26) – (29) with $\varepsilon = 1$. Due to (22) and (31), there is a $\beta_n = \beta(\delta_n, y^{\delta_n})$ such that $x_{\alpha_n}^{\delta_n} = \bar{x}_{\beta_n}^{\delta_n} = (A^*A + \beta_n I)^{-1} A^* y^{\delta_n}$ for n sufficiently large. Since $x_\alpha^\delta \to x^\dagger$, we also have that $\bar{x}_{\beta_n}^{\delta_n} \to x^\dagger$. Together with $x^\dagger \neq 0$ this implies that $\beta_n \to 0$ as $n \to \infty$. However, then (21) and (30) imply that $\bar{\alpha} \leq \|w\|^{-1}$ which is a contradiction. Hence, assumption (32) is wrong.

Now we assume that

$$\liminf_{\delta\to 0} \alpha(\delta) < \|w\|^{-1}. \tag{33}$$

Then there is a sequence $\delta_n > 0$ such that $\lambda_1 \geq \delta_n \to 0$ and $\alpha_n := \alpha(\delta_n) \to \bar{\alpha} < \|w\|^{-1}$ as $n \to \infty$. We choose the data y^{δ_n} as in (26) – (29) with $\varepsilon > 0$ such that

$$\varepsilon^2 + 2\varepsilon\|w\| \leq \frac{1}{4}(\bar{\alpha}^{-1} + \|w\|)^2 - \|w\|^2 .$$

Then (22) and (31) imply that $x_{\alpha_n}^{\delta_n} = A^\dagger y^{\delta_n}$ for n sufficiently large. However, then

$$\begin{aligned} \|x_{\alpha_n}^{\delta_n} - x^\dagger\| &= \varepsilon\|A^* G_k z\| \geq \varepsilon\Big(\frac{1}{2}\lambda_{k+1}\Big)^{\frac{1}{2}} = \Big(\varepsilon\frac{3}{2}\lambda_k\frac{\varepsilon}{3}\frac{\lambda_{k+1}}{\lambda_k}\Big)^{\frac{1}{2}} \\ &\geq \Big(\frac{\varepsilon\delta}{3C}\Big)^{\frac{1}{2}} , \end{aligned}$$

which is a contradiction to (24). Hence, assumption (33) is wrong. This finally proves that assertion (25) holds. ∎

Since it follows from converse results in [13] that the convergence rate $\mathcal{O}(\delta^{\frac{2}{3}})$ implies that $x^\dagger = A^*A\bar{v}$ for some $\bar{v} \in X$, Proposition 4 shows that the enhanced rates in Theorem 3 can, in general, not be achieved if $\alpha(\delta) \not\to \|F'(x^\dagger)\bar{v}\|^{-1}$.

The required α-choice of Theorem 3 and its justification by Proposition 4 indicate that the so-called Bakushinskii veto (cf. [1], see also [4, Theorem 3.3]) must not be overestimated. This veto says that the Moore-Penrose pseudoinverse $A^\dagger$ of the bounded linear operator A is bounded whenever the worst case error of an arbitrary regularization method converges to zero for all $y \in \mathcal{D}(A^\dagger)$ and $\alpha = \alpha(y^\delta)$ is chosen independently of δ. Although the parameter α chosen in Theorem 3 does not depend on δ this is not a contradiction to this veto. Namely, in our case α depends on y. In this case the proof of the Bakushinskii veto is not applicable.

4.3. Extension to convex functions and conclusions on the choice of p

In [19, Chapter 2] the generalized version

$$f(\|F(x) - y^\delta\|) + \alpha R(x)$$

of a Tikhonov functional with monotone functions f is considered. We will discuss this problem in a Hilbert space setting for special functions f:

We assume that X and Y are Hilbert spaces and that $A : X \to Y$ is a bounded linear operator. Moreover, we assume that $g : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ is a convex, strictly monotonically increasing function that is continuously differentiable on $\mathbb{R}^+$. We look for regularized solutions x_α^δ minimizing

$$g(\|Ax - y^\delta\|^2) + \alpha\tfrac{1}{2}\|x\|^2 , \qquad \alpha > 0 . \tag{34}$$

Obviously, $f(t) = g(t^2)$ in our considerations. As above it now follows that x_α^δ solves the equation

$$2\,g'(\|Ax_\alpha^\delta - y^\delta\|^2)A^*(Ax_\alpha^\delta - y^\delta) + \alpha x_\alpha^\delta = 0$$

if $Ax_\alpha^\delta \neq y^\delta$. Thus, x_α^δ is then a standard Tikhonov regularized solution

$$x_\alpha^\delta = \bar{x}_\beta^\delta = (A^*A + \beta I)^{-1}A^*y^\delta \quad \text{with} \quad \alpha = \beta\,2\,g'(\|A\bar{x}_\beta^\delta - y^\delta\|^2)\,. \tag{35}$$

This means, whenever we have a parameter rule $\beta = \beta(\delta, y^\delta) > 0$ yielding a convergence rate for $\|\bar{x}_\beta^\delta - x^\dagger\|$, we obtain the same rate for $\|x_\alpha^\delta - y^\delta\|$ if we choose α as in (35) and if $A\bar{x}_\beta^\delta - y^\delta \neq 0$. The last assumption is always satisfied if $y^\delta \notin \mathcal{R}(A)^\perp$.

Since, for Hilbert spaces X and Y the computation of standard regularized solutions with residual norm square is much easier than the calculation of minimizers of (34) or of (2) with $p \neq 2$, it is questionable why one should prefer such generalizations in the Hilbert space setting. However, note that the situation may be different for Banach spaces. If we consider Lebesgue spaces $Y = L^p$ with $p > 1$, $p \neq 2$, then the choice of that exponent p in (2) simplifies the structure of the functional and helps to reduce the amount of computations for finding regularized solutions.

All considerations above are no longer true if g is not convex. In that case it is not obvious how the convergence analysis for standard Tikhonov regularization could be used to obtain results for minimizers of (34). Non-convex residual terms also occur in (2) whenever $0 < p < 1$, see [6].

Acknowledgement

The work has been 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. B. Hofmann and U. Tautenhahn as long-term guests of that scientific event thank RICAM for kind financial support. Research of B. Hofmann was partly supported by Deutsche Forschungsgemeinschaft (DFG) under Grant HO1454/7-2.

References

[1] A. B. Bakushinskii, Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion, *USSR Comp. Math. Math. Phys.*, 24:181-182, 1984.

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

[3] I. Cioranescu, *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*, Vol. 62 of Mathematics and its Applications, Kluwer, Dordrecht, 1990.

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

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

[6] J. Geissler and B. Hofmann, A new approach to source conditions in regularization with general residual term, 2009, published electronically in arXiv:0906.3438v1.

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

[8] T. Hein, On Tihkonov regularization in Banach spaces - optimal convergence rates, *Appl. Anal.*, 88, 2009, DOI: 10.1080/00036810802555474, to appear.

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

[10] 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*, 23:987–1010, 2007.

[11] J. Köhler and U. Tautenhahn, Error bounds for regularized solutions of nonlinear ill-posed problems, *J. Inv. Ill-Posed Problems*, 3:47–74, 1995.

[12] A. Neubauer, Tikhonov regularisation for nonlinear ill-posed problems: optimal convergence rates and finite-dimensional approximation, *Inverse Problems*, 5:541–557, 1989.

[13] A. Neubauer, On converse and saturation results for Tikhonov regularization of linear ill-posed problems, *SIAM J. Numer. Anal.*, 34:517–527, 1997.

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

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

[16] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational Methods in Imaging*. Vol. 167 of Applied Mathematical Sciences. Springer, New York, 2008.

[17] U. Tautenhahn, Tikhonov regularization for identification problems in differential equations, in: J. Gottlieb etal, *Parameter Identification and Inverse Problems in Hydrology, Geology and Ecology*, Kluwer, Dordrecht, 261–270, 1996.

[18] U. Tautenhahn and Q. Jin, Tikhonov regularization and a posteriori rules for solving nonlinear ill-posed problems, *Inverse Problems*, 19:1–21, 2003.

[19] A.N. Tikhonov, A. N., A.S. Leonov, and A.G. Yagola, *Nonlinear Ill-Posed Problems*, Vol. 1., Series Applied Mathematics and Mathematical Computation, Vol. 14. Translated from the Russian. Chapman & Hall, London, 1998.

[20] Z.-B. Xu and G. F. Roach, Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces, *J. Math. Anal. Appl.*, 157:189–210, 1991.