Convergence Rates for Tikhonov Regularization from Different Kinds of Smoothness Conditions

Albrecht Böttcher[1], **Bernd Hofmann**[2],
Ulrich Tautenhahn[3] and **Masahiro Yamamoto**[4]

June 2005

Abstract. The paper is concerned with ill-posed operator equations $Ax = y$ where $A : X \to Y$ is an injective bounded linear operator with non-closed range $\mathcal{R}(A)$ and X and Y are Hilbert spaces. The solution $x = x^\dagger$ is assumed to be in the range $\mathcal{R}(G)$ of some selfadjoint strictly positive bounded linear operator $G : X \to X$. Under several assumptions on G, such as $G = \varphi(A^*A)$ or more generally $\mathcal{R}(G) \subset \mathcal{R}(\varphi(A^*A))$, inequalities of the form $\varrho^2(G) \le A^*A$, or range inclusions $\mathcal{R}(\varrho(G)) \subset \mathcal{R}(|A|)$, convergence rates for the regularization error $\|x_\alpha - x^\dagger\|$ of Tikhonov regularization are established. We also show that part of our assumptions automatically imply so-called source conditions. The paper contains a series of new results but also intends to uncover cross-connections between the different kinds of smoothness conditions that have been discussed in the literature on convergence rates for Tikhonov regularization.

Key words. linear ill-posed problems, Tikhonov regularization, convergence rates, smoothness conditions, index functions, operator monotone functions, range inclusions

MSC 2000 subject classifications: 47A52, 65J20, 65R30

1 Introduction

Let $A : X \to Y$ be a bounded linear operator mapping between infinite-dimensional Hilbert spaces X and Y with norms $\|\cdot\|$ and inner products $(\,\cdot\,,\cdot\,)$. Throughout this paper we assume that A is injective and that the range $\mathcal{R}(A)$ is not closed. The equation

$$Ax = y \tag{1.1}$$

has a unique solution $x = x^\dagger \in X$ for every $y \in \mathcal{R}(A)$ but is ill-posed. Consequently, a regularization approach is required to find stable approximate solutions. In this Hilbert

[1]Department of Mathematics, Chemnitz University of Technology, 09107 Chemnitz, Germany, e-mail: aboettch@mathematik.tu-chemnitz.de

[2]Department of Mathematics, Chemnitz University of Technology, 09107 Chemnitz, Germany, e-mail: hofmannb@mathematik.tu-chemnitz.de

[3]Department of Mathematics, University of Applied Sciences Zittau/Görlitz, P.O. Box 1454, 02754 Zittau, Germany, e-mail: u.tautenhahn@hs-zigr.de

[4]Department of Mathematical Sciences, The University of Tokyo, 3-8-1 Komaba Meguro, Tokyo 153, Japan, e-mail: myama@ms.u-tokyo.ac.jp

space setting Tikhonov regularization consists of choosing a parameter $\alpha > 0$ and solving the extremal problem

$$\|Ax - y\|^2 + \alpha\|x\|^2 \to \min \,. \tag{1.2}$$

The last problem has a unique solution $x = x_\alpha \in X$. This solution is given by

$$x_\alpha = (A^*A + \alpha I)^{-1} A^* y \tag{1.3}$$

and is called regularized solution. It is well known that $\|x_\alpha - x^\dagger\| \to 0$ as $\alpha \to 0$. Notice that, however, the convergence $x_\alpha \to x^\dagger$ may be spectacularly slow (see, for example, [6, Proposition 3.11] and [29]). For treatises of several aspects of Tikhonov regularization we refer to the seminal paper [33] and, for example, to the books [2], [3, chapter 6], [6, Chapter 5], [7], [9], [16], [17], [18], [23], [34], [36].

This paper is mainly concerned with estimates of the form

$$\|x_\alpha - x^\dagger\| \;\le\; K\,\varphi(\alpha) \;\text{ for }\; 0 < \alpha \le \overline{\alpha} \tag{1.4}$$

with a constant $K < \infty$ and a certain continuous and strictly increasing function φ with $\varphi(0) = 0$. Such estimates (1.4) are obtained under different kinds of smoothness conditions concerning the solution $x^\dagger$, the operator A, and their cross-relations. They express convergence rates of regularized solutions in the case of noiseless data y.

To obtain the desired estimates (1.4) we employ certain a priori assumptions on $x^\dagger$. To be specific, we assume throughout this paper the solution smoothness

$$x^\dagger \;=\; G\,v \;\text{ with }\; v \in X \;\text{ and }\; \|v\| \le R\,, \tag{1.5}$$

where $G : X \to X$ is a bounded linear and selfadjoint operator which is strictly positive in the sense that $(G\,x, x) > 0$ for all $x \neq 0$. The operator G is required to satisfy one of the following assumptions. In accordance with [8] and [21] we call a function $f : [0, \overline{t}\,] \to \mathbf{R}$ an index function if it is continuous and strictly increasing with $f(0) = 0$.

Assumption A1. We have $G = \varphi(A^*A)$ with some index function φ on $[0, \|A\|^2]$.

Assumption A2. We have $\mathcal{R}(G) \subset \mathcal{R}(\varphi(A^*A))$ with some index function φ on $[0, \|A\|^2]$.

Assumption A3. We have $\|\varrho(G)x\| \le \|Ax\|$ for all $x \in X$ with some index function ϱ on $[0, \|G\|]$.

Assumption A4. We have $\|\varrho(G)x\| \le C\,\|Ax\|$ for all $x \in X$ with some index function ϱ on $[0, \|G\|]$ and some constant $C < \infty$.

Assumption A5. We have $\mathcal{R}(\varrho(G)) \subset \mathcal{R}(|A|)$ for some index function ϱ on $[0, \|G\|]$, where $|A| := (A^*A)^{1/2}$.

We note that $\mathcal{R}(|A|) = \mathcal{R}(A^*)$ (which follows, for example, from the polar decomposition $A = U\,|A|$), so that the range inclusion in assumption A5 may be replaced with the inclusion $\mathcal{R}(\varrho(G)) \subset \mathcal{R}(A^*)$.

We refer to the above assumptions as smoothness assumptions. Clearly, the equality $x^\dagger = Gv$ is a general smoothness condition on $x^\dagger$. Assumptions A1 to A5 characterize the relative severity of ill-posedness of equation (1.1) with respect to the operator G. Assumptions A3 to A5 state that, in a sense, $\varrho(G)$ is not less smoothing than A. Using that an operator B has closed range if and only if zero is an isolated point of the spectrum of B^*B, it is not difficult to prove that under any of the assumptions A1 to A5 the operators $\varphi(A^*A)$, $\varrho(G)$, and G have non-closed ranges (recall that we always require that $\mathcal{R}(A)$ is non-closed). If (following the terminology of [26]) the linear equation (1.1) is ill-posed of type II (A compact), then G is necessarily compact whenever one of the assumptions A1 to A5 holds. On the other hand, the operator G in assumptions A2 to A5 can be compact or not if (1.1) is ill-posed of type I (A non-compact with non-closed range $\mathcal{R}(A)$). For more details around this field we also refer, for example, to discussions in [11], [12] and [13].

The paper is organized as follows. Section 2 is devoted to cross-connections between the assumptions listed above. Roughly speaking, we show that

$$\begin{array}{ccccccc} A1 & \Rightarrow & A3 & \overset{\Leftarrow *}{\Rightarrow} & A4 & \Leftrightarrow & A5 \\ \Downarrow & & \Downarrow * & & \Downarrow * & & \Downarrow * \\ A2 & & A2 & & A2 & & A2 \end{array}$$

where an arrow without asterisk means that the implication is true with the same index function φ or ϱ, respectively. On the other hand, an arrow with an asterisk indicates that the implication is true under an additional condition or for a different index function. The philosophy of Sections 3 to 7 is that each of the assumptions A1 to A5 can be accompanied by an extra assumption to yield convergence rates for the regularization error $\|x_\alpha - x^\dagger\|$. In Section 3 we consider assumptions A1 and A2 together with some mild additional hypotheses on the function φ, in Section 4 we combine assumption A3 with operator monotonicity, and in Section 5 we take assumption A3 in conjunction with a concavity condition. Assumptions A4 and A5 turn out to be equivalent, and their implications are the subject of Sections 6 and 7. In Section 6, assumption A5 is paired with the requirement that G be a compact operator, and in Section 7 we return to operator monotonicity or concavity, but this time not in connection with assumption A3 but with assumptions A4 and A5. In Section 8 we show that some saturation phenomenon occurring in the general case disappears if the operators G and A^*A commute. Section 9 summarizes the consequences of the error estimates obtained in Sections 3 to 8 for the error of regularized solutions in the case of noisy data $y^\delta \in Y$ with $\|y - y^\delta\| \le \delta$. In this context, we focus on a priori parameter choices for the regularization parameter α leading to order optimal convergence rates. Section 10 illustrates the abstract theory by a couple of examples related to inverse problems.

2 Connections between the smoothness conditions

Assumption A1 obviously implies assumption A2 with the same index function φ, and assumption A3 trivially implies assumption A4 with the same index function ϱ and $C = 1$. If assumption A4 holds with an index function ϱ and a constant $C > 0$, then assumption A3 is evidently satisfied with the index function ϱ/C. It is also clear that assumption A1 yields assumption A4 with $\varrho = \varphi^{-1}$ and $C = \|A^*\|$. More relationships between assumptions A1 to A5 will be deduced from the following result. This result is not terribly new, but we have not found it explicitly in the literature. Our proof of part (b) mimics arguments of [35, Section 8.6].

Proposition 2.1 *Let S and T be selfadjoint bounded linear operators on X and suppose T is injective.*

(a) If $\mathcal{R}(S) \subset \mathcal{R}(T)$, then $T^{-1}S$ is a well-defined bounded linear operator on X and $\|Sx\| \le C\|Tx\|$ for all $x \in X$ with $C = \|T^{-1}S\|$.

(b) If there is a constant $C < \infty$ such that $\|Sx\| \le C\|Tx\|$ for all $x \in X$, then $\mathcal{R}(S) \subset \mathcal{R}(T)$ and $\|T^{-1}S\| \le C$.

Proof. We first of all remark that the linear operator $T^{-1} : \mathcal{R}(T) \to X$ is well-defined and that $\mathcal{R}(T)$ is dense in X because the null space $\mathcal{N}(T)$ is $\{0\}$ and $\overline{\mathcal{R}(T)} = \mathcal{N}(T^*)^\perp = \mathcal{N}(T)^\perp$.

(a) The range inclusion implies that $D := T^{-1}S : X \to X$ is a well-defined linear operator. To prove that this operator is bounded, we use the closed graph theorem. Thus, suppose $x_n \to x$ and $Dx_n \to y$. We must show that $Dx = y$. But since $Sx_n \to Sx$ and $Sx_n = TT^{-1}Sx_n = TDx_n \to Ty$, it follows that $Sx = Ty$, which gives the desired equality $T^{-1}Sx = y$. Since $S = TD$, we have $S = S^* = D^*T^* = D^*T$ and hence $\|Sx\| \le \|D^*\| \, \|Tx\|$ for all $x \in X$. As $\|D^*\| = \|D\|$, we arrive at the assertion.

(b) Let $u = Sw \in \mathcal{R}(S)$. We consider the functional $F : x \mapsto (T^{-1}x, u)$ on $\mathcal{R}(T)$. For $x \in \mathcal{R}(T)$,

$$\begin{aligned} |F(x)| &= |(T^{-1}x, u)| = |(T^{-1}x, Sw)| = |(ST^{-1}x, w)| \\ &\le \|ST^{-1}\| \, \|w\| \le C\|TT^{-1}x\| \, \|w\| = C\|w\| \, \|x\| \, . \end{aligned}$$

Thus, F is bounded on $\mathcal{R}(T)$. As $\mathcal{R}(T)$ is dense in X, the functional F can be continued to a bounded linear functional $\widetilde{F}$ on all of X. By the Riesz representation theorem, there is a $y^* \in X$ such that $\widetilde{F}(x) = (x, y^*)$ for all $x \in X$. For $x \in \mathcal{R}(T)$ we therefore have $(T^{-1}x, u) = (x, y^*)$. Hence for every $z \in X$,

$$(z, u) = (T^{-1}Tz, u) = (Tz, y^*) = (z, Ty^*) \, ,$$

which implies that $u = Ty^*$. This shows that u is in $\mathcal{R}(T)$ and completes the proof of the inclusion $\mathcal{R}(S) \subset \mathcal{R}(T)$.

Since $\|ST^{-1}y\| \le C\|TT^{-1}y\| = C\|y\|$ for every $y \in \mathcal{R}(T)$ and $\mathcal{R}(T)$ is dense in X, the operator $ST^{-1} : \mathcal{R}(T) \to X$ can be continued to a bounded linear operator $B : X \to X$ satisfying $\|B\| \le C$. If $x \in X$ and $y \in \mathcal{R}(T)$, then

$$(x, ST^{-1}y) = (Sx, T^{-1}y) = (TT^{-1}Sx, T^{-1}y) = (T^{-1}Sx, TT^{-1}y) = (T^{-1}Sx, y).$$

This equality can be extended to the equality $(x, By) = (T^{-1}Sx, y)$ for all $x \in X$ and all $y \in Y$. It follows that $(T^{-1}S)^* = B$, whence $\|T^{-1}S\| = \|B\| \le C$. $\square$

Corollary 2.2 *(a) Assumption A4 implies assumption A5.*

(b) Assumption A5 implies assumption A4 with $C = \| \, |A|^{-1} \varrho(G) \|$.

Proof. (a) We apply Proposition 2.1 to $S = \varrho(G)$ and $T = |A|$ just noticing that $\| \, |A|x\| = \|Ax\|$ for all $x \in X$. $\square$

As usual, we write $S \le T$ for two selfadjoint bounded linear operators S and T if $(Sx, x) \le (Tx, x)$ for all $x \in X$. A function $f : [0, \bar{t}\,] \to \mathbf{R}$ is said to be operator monotone on $[0, \bar{t}\,]$ if $f(S) \le f(T)$ whenever S and T are two selfadjoint bounded linear operators with $S \le T$ whose spectra are contained in $[0, \bar{t}\,]$. We remark that the use of operator monotone functions in connection with the analysis of ill-posed problems is due to [20] (see also some more results in [22]).

We denote by ψ^{-1} the inverse of a bijective function ψ and by ψ^2 the square (and not the second iterate) of ψ, that is, $\psi^2(t) = [\psi(t)]^2$.

In connection with the following corollary we note that assumption A1 implies the equality $\|G\| = \varphi(\|A\|^2)$ and that in the case of assumption A3 we have the inequality $\varrho(\|G\|) \le \|A\|$ and thus $\varrho^2(\|G\|) \le \|A\|^2$.

Corollary 2.3 *(a) If assumption A1 holds, then assumption A3 is true in the equality form $\|\varrho(G)x\| = \|Ax\|$ for all $x \in X$ with ϱ defined by $\varrho(s) = \sqrt{\varphi^{-1}(s)}$ for $s \in [0, \varphi(\|A\|^2)]$.*

(b) If assumption A3 holds and the function φ defined by $\varphi(t) = \varrho^{-1}(\sqrt{t})$ for $t \in [0, \varrho^2(\|G\|)]$ has the property that φ^2 can be continued to an operator monotone index function on $[0, \|A\|^2]$, then assumption A2 is valid with this φ. Furthermore, in that case we have $x^\dagger = \varphi(A^*A)\,u$ with $u \in X$ and $\|u\| \le R$.

(c) If assumption A3 holds and the function ϱ^{-1} defined on $[0, \varrho(\|G\|)]$ can be continued to an operator monotone index function on $[0, \|A\|]$, then assumption A2 is valid with $\varphi(t) = \sqrt{\varrho^{-1}(\sqrt{t})}$ for $t \in [0, \|A\|^2]$.

Proof. (a) We have $\varrho^2(\varphi(t)) = t$ for $t \in [0, \|A\|^2]$ and hence $\varrho^2(G) = \varrho^2(\varphi(A^*A)) = A^*A$, which implies that $\|\varrho(G)x\| = \|Ax\|$ for all $x \in X$.

(b) Since $\|\varrho(G)x\| \le \|Ax\|$ for all $x \in X$, we may conclude that $\varrho^2(G) \le A^*A$. The function φ^2 with $\varphi(t) = \varrho^{-1}(\sqrt{t})$ is operator monotone on $[0, \|A\|^2]$ and the spectra of $\varrho^2(G)$ and A^*A are obviously contained in $[0, \|A\|^2]$. Consequently, $\varphi^2(\varrho^2(G)) \le \varphi^2(A^*A)$. Because $\varphi^2(\varrho^2(s)) = s^2$ for $s \in [0, \|G\|]$, it follows that $G^2 \le \varphi^2(A^*A)$, which is equivalent to the inequality $\|Gx\| \le \|\varphi(A^*A)x\|$ for all $x \in X$. Proposition 2.1(b) now implies that $\mathcal{R}(G) \subset \mathcal{R}(\varphi(A^*A))$ and $\|\varphi(A^*A)^{-1}G\| \le 1$. Thus, $x^\dagger = \varphi(A^*A)u$ with $u := \varphi(A^*A)^{-1}Gv$ satisfying $\|u\| \le \|\varphi(A^*A)^{-1}G\| \, \|v\| \le R$.

(c) Again we have $\varrho^2(G) \le A^*A$. The function $\sqrt{t}$ is known to be operator monotone on $[0, \infty)$ (see, e.g., [4, Theorem V.1.8]). This implies that $\varrho(G) \le |A|$, and the operator monotonicity of ϱ^{-1} yields $G \le \varrho^{-1}((A^*A)^{1/2})$. Thus, from Proposition 2.1(b) we obtain $\mathcal{R}(G^{1/2}) \subset \mathcal{R}((\varrho^{-1}((A^*A)^{1/2}))^{1/2})$, and since $\|Gx\| \le \|G^{1/2}\| \, \|G^{1/2}x\|$ for all $x \in X$, we arrive at the inclusions

$$\mathcal{R}(G) \subset \mathcal{R}(G^{1/2}) \subset \mathcal{R}((\varrho^{-1}((A^*A)^{1/2}))^{1/2}).$$

So we have assumption A2 with $\varphi(t) = \sqrt{\varrho^{-1}(\sqrt{t})}$. $\square$

3 Source conditions

A condition

$$x^\dagger = \varphi(A^*A)\, u \quad \text{with} \quad u \in X \tag{3.1}$$

and with an index function φ defined on $[0, \|A\|^2]$ is called a source condition. Throughout this paper, we assume that (1.5) characterizes the solution smoothness. Then a source condition (3.1) can hold purely accidental with some accidental φ und some accidental u. Note that for fixed $x^\dagger$ condition (3.1) can hold for very different pairs (φ, u). Here we focus on source conditions (3.1) which are implied systematically by one of the assumptions A1 to A5. The following simple proposition reveals that under assumptions A1 and A2 a source condition (3.1) is automatically satisfied.

Proposition 3.1 *Let one of the assumptions A1 or A2 hold. Then we have a source condition (3.1) as a consequence of the smoothness condition (1.5).*

Proof. If the operators G and A^*A are related by a function φ as in Assumption A1, then in view of (1.5) the source condition (3.1) holds trivially with $u = v$. Under assumption A2, it follows from (1.5) that $x^\dagger \in \mathcal{R}(\varphi(A^*A))$, which is (3.1). $\square$

Now we formulate consequences of source conditions (3.1) for the error of Tikhonov regularization. The following result is from [21].

Proposition 3.2 *Suppose (3.1) holds and put $\bar{t} = \|A\|^2$. If there is a constant $k < \infty$ such that*

$$\sup_{t \in [0,\bar{t}]} \frac{\alpha}{t+\alpha}\, \varphi(t) \le k\varphi(\alpha) \ \textit{ for all } \ \alpha \in (0, \bar{t}\,]\,, \tag{3.2}$$

then

$$\|x_\alpha - x^\dagger\| \le kR\varphi(\alpha) \ \textit{ for all } \ \alpha \in (0, \bar{t}\,]\,. \tag{3.3}$$

Proof. From (1.3) it is readily seen that $x^\dagger - x_\alpha = \alpha(A^*A + \alpha I)^{-1} x^\dagger$. Since $x^\dagger = \varphi(A^*A)u$ with $\|u\| \le R$, it follows that

$$\begin{aligned} &\|x_\alpha - x^\dagger\| = \|\alpha(A^*A + \alpha I)^{-1}\varphi(A^*A)u\| \\ &\le \|\alpha(A^*A + \alpha I)^{-1}\varphi(A^*A)\|\,\|u\| \le \sup_{t\in[0,\bar{t}]} \frac{\alpha}{t+\alpha}\,\varphi(t)\,R\,, \end{aligned}$$

and hence (3.2) yields (3.3). $\square$

The following result provides us with sufficient conditions for the validity of (3.2).

Proposition 3.3 *Let $\varphi : [0, \bar{t}\,] \to \mathbf{R}$ be an index function with $\bar{t} = \|A\|^2$. If (a) $t/\varphi(t)$ is monotonically increasing on $[0, \bar{t}\,]$ or (b) $\varphi(t)$ is concave on $[0, \bar{t}\,]$ then (3.2) holds with $k = 1$. If (c) $\varphi(t)$ is operator monotone on $[0, \bar{t}\,]$, then (3.2) is valid with some $k \ge 1$. If there exists a $\hat{t} \in (0, \bar{t}\,)$ such that (d) $t/\varphi(t)$ is monotonically increasing on $[0, \hat{t}\,]$ or (e) $\varphi(t)$ is concave on $[0, \hat{t}\,]$, then (3.2) is true with $k = \varphi(\bar{t})/\varphi(\hat{t})$.*

Proof. Part (a) may be found in [21]. To prove (b), let φ be concave on $[0,\bar{t}]$. Since $\varphi(0)=0$, for arbitrary $0<s<t\le\bar{t}$ we have $\varphi(t)/t\le\varphi(s)/s$, or equivalently, $s/\varphi(s)\le t/\varphi(t)$. Hence, $t/\varphi(t)$ is monotonically increasing on $(0,\bar{t}\,]$ and (b) follows from (a). Part (c) is established in [22]. It is based on the fact that an operator monotone index function on $[0,\bar{t}]$ can be decomposed into a sum $\varphi=\varphi_0+\varphi_1$ of a concave index function φ_0 and a Lipschitz continuous index function φ_1 with Lipschitz constant L_1 (see [20]). The constant $k\ge 1$ in (c) then has the form $k=1+L_1\,\bar{t}/\varphi_0(\bar{t})$ ([22, Lemma 3 and Proposition 2]) and attains in accordance with (b) the value $k=1$ if we have $L_1=0$, i.e., if the operator monotone function φ is concave. To prove (d), we distinguish three cases. In the first case $\alpha\ge t$ we use the monotonicity of φ to obtain that

$$\frac{\alpha\,\varphi(t)}{t+\alpha}\le\varphi(t)\le\varphi(\alpha)\,.$$

In the second case $\alpha\le t$ and $t\le\hat{t}$ we take into account that the function $t/\varphi(t)$ is monotonically increasing to get

$$\frac{\alpha\,\varphi(t)}{t+\alpha}\le\frac{\alpha\,\varphi(t)}{t}\le\varphi(\alpha)\,.$$

In the third case $\alpha\le t$ and $t\in[\hat{t},\bar{t}]$ we employ the monotonicity of both $\varphi(t)$ and $t/\varphi(t)$ to conclude that

$$\frac{\alpha\,\varphi(t)}{t+\alpha}\le\frac{\alpha\,\varphi(t)}{t}\le\frac{\alpha\,\varphi(t)}{\hat{t}}\le\frac{\alpha\,\varphi(\bar{t})}{\hat{t}}=k\,\frac{\alpha\,\varphi(\hat{t})}{\hat{t}}\le k\,\varphi(\alpha)\,.$$

This completes the proof of part (d). The proof of (e) is similar to the proof of (d). $\Box$

From Propositions 3.1, 3.2, and 3.3 we immediately obtain the following.

Corollary 3.4 *Let one of the assumptions A1 or A2 hold with an index function φ that satisfies at least one of the conditions (a) to (e) of Proposition 3.3. Then we have the error estimate (3.3).*

We now introduce two families of index functions φ in (3.1) or in assumptions A1 and A2, respectively, that are frequently used in the literature. These families will also be discussed in subsequent sections.

Example 3.5 (Hölder rates) Let $0<\mu\le 1$. Then $\varphi(t)=t^\mu$ is a concave index function on $[0,\infty)$ and by Proposition 3.3(b) it satisfies the hypothesis (3.2) of Proposition 3.2. Thus, under (3.1) we get from (3.3) the estimate $\|x_\alpha-x^\dagger\|=O(\alpha^\mu)$ as $\alpha\to 0$, i.e., we have Hölder convergence rates in that case. The best possible rate of the error is $O(\alpha)$ for $\mu=1$.

So-called converse results allow us to conclude that if an estimate $\|x_\alpha-x^\dagger\|=O(\psi(\alpha))$ with some index function ψ holds, then (3.1) is valid with some index function φ. For example, in [28] we find a converse result for Hölder rates. It says that if $\|x_\alpha-x^\dagger\|=O(\alpha)$ as $\alpha\to 0$, then $x^\dagger=A^*A\,u_1$ with some $u_1\in X$, and if $\|x_\alpha-x^\dagger\|=O(\alpha^\mu)$ as $\alpha\to 0$ with some $0<\mu<1$, then for each $0<\nu<\mu$ there is a $u_\nu\in X$ such that $x^\dagger=(A^*A)^\nu\,u_\nu$.

Using for nonnegative arguments t the relation $\varrho(t) = \sqrt{\varphi^{-1}(t)}$ already exploited in Corollary 2.3(a) we obtain for power functions $\varphi(t) = t^\mu$ with exponents $0 < \mu \le 1/2$ the power functions $\varrho(t) = t^{\frac{1}{2\mu}}$ with exponents in the interval $[1,\infty)$. Then the well-known Heinz-Kato inequality (see, e.g., the corollary of Theorem 2.3.3 in [30, p. 45] and [6, Proposition 8.21]) yields the implication

$$\mathcal{R}(G^{\frac{1}{2\mu}}) \subset \mathcal{R}(|A|) \implies \mathcal{R}(G) \subset \mathcal{R}(|A|^{2\mu}) = \mathcal{R}((A^*A)^\mu)\,, \tag{3.4}$$

and hence for that specific ϱ our assumption A5 implies assumption A2. Consequently, error estimates (3.3) can be obtained along the lines of Corollary 3.4 provided that a range inclusion $\mathcal{R}(G^{\frac{1}{2\mu}}) \subset \mathcal{R}(|A|)$ is valid for $0 < \mu \le 1/2$. The examples 1 and 2 in [14] fit here with $\varrho_1(t) = \varrho(t)$ and $\varrho_2(t) = t$ and illustrate the Hölder rate situation: assumption A5 and hence assumption A2 are fulfilled in those examples, but assumption A1 is not valid. Note that the implication (3.4) fails to hold for $\mu > 1/2$ and general operators G and A. However, in Section 8 we will point out that the implication is true for all $\mu > 0$ if G and A^*A commute. □

Example 3.6 (logarithmic rates) Let $p > 0$ and $K > 0$. Then the function given by $\varphi(t) = K(-\ln t)^{-p}$ for $0 < t \le \bar{t} < 1$ and $\varphi(0) = 0$ is an index function on $[0, \bar{t}\,]$. This function is concave for $0 \le t \le \hat{t} = e^{-p-1}$ and hence Propositions 3.3(e) and 3.2 show that if (3.1) is satisfied with this index function, then we have the logarithmic convergence rates $\|x_\alpha - x^\dagger\| = O\left((-\ln \alpha)^{-p}\right)$ as $\alpha \to 0$.

For such logarithmic rates, converse results were established in [15]: if $\|x_\alpha - x^\dagger\| = O\left((-\ln \alpha)^{-p}\right)$ as $\alpha \to 0$, then for each $0 < q < p$ there exists a $u_q \in X$ such that $x^\dagger = (-\ln(A^*A))^{-q}\, u_q$. □

4 Operator monotonicity

In contrast to assumption A2 and its special case A1, assumptions A3, A4, A5 do not provide us directly with a source condition (3.1) which implies an estimate (3.3) for the regularization error under weak requirements on φ. Nevertheless, such estimates can be derived indirectly from assumptions A3 to A5 by imposing stronger conditions on the index function ϱ. We start with assumption A3 and our objective is to deduce a source condition (3.1) and the hypothesis of Proposition 3.2 from it. The price we have to pay for this is a quite strong requirement concerning the index function ϱ.

Proposition 4.1 *(a) Let assumption A3 hold, define φ on $[0, \varrho^2(\|G\|)]$ by $\varphi(t) = \varrho^{-1}(\sqrt{t})$, and suppose φ^2 can be continued to an operator monotone index function on $[0, \|A\|^2]$. Then*

$$\|x_\alpha - x^\dagger\| \le kR\varphi(\alpha) \;\; \textit{for all} \;\; \alpha \in [0, \|A\|^2]\,, \tag{4.1}$$

where $k \ge 1$ is the constant from Proposition 3.3(c).

(b) Let assumption A3 hold and suppose ϱ^{-1} can be continued to an operator monotone index function on $[0, \|A\|]$. Then we have inequality (4.1) with some constant $k \ge 1$ and with $\varphi(t) = \sqrt{\varrho^{-1}(\sqrt{t})}$ for $t \in [0, \|A\|^2]$.

Proof. (a) By virtue of Corollary 2.3(b), assumption A2 is satisfied with $\varphi(t) = \varrho^{-1}(\sqrt{t})$ continued to $[0, \|A\|^2]$ whenever assumption A3 holds with index function ϱ. Since φ^2 is assumed to be operator monotone and the function $\sqrt{t}$ is known to be operator monotone on $[0, \infty)$, we have operator monotonicity of φ. From Proposition 3.3(c) we now infer that (3.2) is true with some $k \geq 1$. Proposition 3.2 therefore yields (4.1).

(b) As a consequence of Corollary 2.3(c), assumption A2 is satisfied with the function $\varphi(t) = \sqrt{\varrho^{-1}(\sqrt{t})}$ continued to $[0, \|A\|^2]$ whenever assumption A3 holds with the index function ϱ. Taking into account Proposition 3.3(c) we can therefore argue as in the proof of part (a). The operator monotonicity of φ now follows from the fact that the composition of operator monotone functions remains operator monotone. $\square$

Note that the requirement concerning operator monotonicity in Proposition 4.1(a) is stronger than that in Proposition 4.1(b). However, the convergence rate (4.1) gained from Proposition 4.1(b) is only the square root of the rate obtained from Proposition 4.1(a).

Example 4.2 We come back to the Hölder rates of Example 3.5 and consider the function $\varrho(t) = t^{\frac{1}{2\mu}}$ with $0 < \mu \leq 1$ in assumption A3. In this case we deduce from Proposition 4.1(a) that assumption A2 is true with $\varphi(t) = \varrho^{-1}(\sqrt{t}) = t^\mu$ whenever $0 < \mu \leq 1/2$, since φ^2 is operator monotone on $[0, \infty)$ for $0 < \mu \leq 1/2$ ([4, Theorem V.1.9]). This yields the convergence rates (4.1) in the form $\|x_\alpha - x^\dagger\| = O(\alpha^\mu)$ as $\alpha \to 0$. The best possible rate to be expected is $O(\sqrt{\alpha})$ in the case $\mu = 1/2$. Higher order Hölder rates $O(\alpha^\mu)$ with $1/2 < \mu \leq 1$ as in Example 3.5 cannot be derived from Proposition 4.1(a) if the operator monotone function φ^2 appears as a companion of assumption A3 (see also Remark 7.2 and Section 8). Note that Proposition 4.1(b) can also be used for $1/2 < \mu \leq 1$, but that then the resulting convergence rate is only

$$\|x_\alpha - x^\dagger\| = O\left(\sqrt{\varrho^{-1}(\sqrt{\alpha})}\right) = O\left(\alpha^{\mu/2}\right),$$

with the best possible case $O(\sqrt{\alpha})$ for $\mu = 1$. $\square$

Example 4.3 Now we return to the logarithmic rates of Example 3.6 and consider the function $f_p(t) = (-\ln t)^{-p}$ with some exponent $p > 0$ and $f_p(0) = 0$. This function is operator monotone on $[0, \bar{t}\,]$ for $0 < \bar{t} < 1$ whenever $0 < p \leq 1$ (see [20, Examples 2 and 3]). Let assumption A3 hold with $\varrho(t) = \exp\left(-K/t^{1/p}\right)$. Then

$$\varrho^{-1}(t) = K^p f_p(t) \implies \varrho^{-1}(\sqrt{t}) = (2K)^p f_p(t) \implies (\varrho^{-1}(\sqrt{t}))^2 = (2K)^{2p} f_{2p}(t)$$

and hence $\sqrt{\varrho^{-1}(\sqrt{t})} = (2K)^{\frac{p}{2}} f_{\frac{p}{2}}(t)$. For $\|A\| < 1$, Proposition 4.1(a) yields the convergence rates

$$\|x_\alpha - x^\dagger\| = O(f_p(\alpha)) \;\text{ as }\; \alpha \to 0 \tag{4.2}$$

whenever $0 < p \leq 1/2$, but Proposition 4.1(a) is not applicable for $p > 1/2$. Namely, as noted in [20, p. 630], for $p > 1/2$ the function $(\varrho^{-1}(\sqrt{t}))^2 = (2T)^{2p} f_{2p}(t)$ fails to be operator monotone on any interval $[0, \varepsilon]$ with $\varepsilon > 0$. On the other hand, Proposition 4.1(b) also applies for $1/2 < p \leq 1$ and gives the convergence rates $\|x_\alpha - x^\dagger\| = O(f_{\frac{p}{2}}(\alpha))$. The best rate is $O((-\ln \alpha)^{-1/2})$ in both cases. $\square$

5 Concavity

Operator monotonicity is an extremely involved property and difficult to check in concrete cases. Contrary to this, concavity is easy to comprehend and to verify, and in this section we will establish a result on convergence rates based on that property for $\varphi^2(t) = (\varrho^{-1}(\sqrt{t}))^2$ in addition to assumption A3. As already mentioned, any operator monotone index function on an interval $[0,\bar{t}\,]$ is a sum of a concave index function and a Lipschitz continuous index function, but neither such a sum nor a concave index function alone are necessarily operator monotone. In particular, for $p > 1/2$ the function $f_p^2(t) = f_{2p}(t)$ and its multiples ocurring in Example 4.3 fail to be operator monotone on any interval $[0,\varepsilon]$, although this function is concave on the interval $[0, e^{-2p-1}]$.

Under assumption A3 with $\varrho(t) = \exp\left(-1/(2\,t^{1/p})\right)$ and with the additional restriction $\varrho^2(\|G\|) \le e^{-2p-1}$, which is equivalent to $\|G\| \le (2p+1)^{-p}$, the following theorem allows us to derive logarithmic convergence rates (4.2) for all $p > 0$. Based on Proposition 4.1(a) such rates can only be derived for $0 < p \le 1/2$.

Theorem 5.1 *Let assumption A3 hold and define the index function φ by $\varphi(t) = \varrho^{-1}(\sqrt{t})$ for $t \in [0, \varrho^2(\|G\|)]$. If φ^2 is concave on $[0, \varrho^2(\|G\|)]$, then there exists a number $\overline{\alpha} > 0$ such that*

$$\|x_\alpha - x^\dagger\| \le R\,\varphi(\alpha) \quad \textit{for} \quad 0 < \alpha \le \overline{\alpha}\,. \tag{5.1}$$

Proof. Let us introduce the abbreviation $z_\alpha = x^\dagger - x_\alpha$. We already observed in Section 3 that (1.3) implies the identity $z_\alpha = x^\dagger - x_\alpha = B_\alpha x^\dagger$ with $B_\alpha = \alpha(A^*A + \alpha I)^{-1}$. Taking into account that $\|B_\alpha\| \le 1$, we obtain from (1.5) the estimate

$$\|z_\alpha\|^2 \le \|B_\alpha^{1/2}\, x^\dagger\|^2 = (z_\alpha, x^\dagger) = (Gz_\alpha, v) \le R\,\|Gz_\alpha\|\,. \tag{5.2}$$

Since $\|(A^*A)^{1/2} B_\alpha^{1/2}\| \le \sqrt{\alpha}$, we therefore get

$$\|Az_\alpha\|^2 = \|(A^*A)^{1/2} B_\alpha x^\dagger\|^2 \le \alpha \|B_\alpha^{1/2} x^\dagger\|^2 \le \alpha R\,\|Gz_\alpha\|\,. \tag{5.3}$$

Our next objective is to derive a third estimate that relates the three quantities $\|z_\alpha\|$, $\|Az_\alpha\|$ and $\|Gz_\alpha\|$. This estimate will be derived by appropriate interpolation. We define the index function ξ on the interval $[0, \varrho^2(\|G\|)]$ by $\xi(t) = (\varphi^2)^{-1}(t) = \varrho^2(\sqrt{t})$. This function is convex because φ^2 is concave. We may therefore employ Jensen's inequality and have

$$\begin{aligned}
\xi\left(\frac{\|Gz_\alpha\|^2}{\|z_\alpha\|^2}\right) &= \xi\left(\frac{\int \lambda^2\, \mathrm{d}\|E_\lambda z_\alpha\|^2}{\int \mathrm{d}\|E_\lambda z_\alpha\|^2}\right) \\
&\le \frac{\int \xi(\lambda^2)\, \mathrm{d}\|E_\lambda z_\alpha\|^2}{\int \mathrm{d}\|E_\lambda z_\alpha\|^2} \qquad \text{(by Jensen's inequality)} \\
&= \frac{\int \varrho^2(\lambda)\, \mathrm{d}\|E_\lambda z_\alpha\|^2}{\|z_\alpha\|^2} = \frac{\|\varrho(G)\, z_\alpha\|^2}{\|z_\alpha\|^2} \\
&\le \frac{\|Az_\alpha\|^2}{\|z_\alpha\|^2} \qquad \text{(by assumption A3)}\,.
\end{aligned} \tag{5.4}$$

Now we derive the desired estimate (5.1) by combining the estimates (5.2), (5.3) and (5.4). We introduce the function $f(t) = \xi\left(t^2\right)/t^2$. Since ξ is a convex index function we conclude that f is monotonically increasing. Hence by (5.2), which may be rewritten as $\|Gz_\alpha\|^{1/2}/R^{1/2} \le \|Gz_\alpha\|/\|z_\alpha\|$, the monotonicity of f, and (5.4),

$$f\left(\frac{\|Gz_\alpha\|^{1/2}}{R^{1/2}}\right) \le f\left(\frac{\|Gz_\alpha\|}{\|z_\alpha\|}\right) = \frac{\|z_\alpha\|^2}{\|Gz_\alpha\|^2}\cdot\xi\left(\frac{\|Gz_\alpha\|^2}{\|z_\alpha\|^2}\right) \le \frac{\|Az_\alpha\|^2}{\|Gz_\alpha\|^2}\,.$$

Multiplying by $\|Gz_\alpha\|/R$ and exploiting (5.3) we get

$$\xi\left(\frac{\|Gz_\alpha\|}{R}\right) \le \frac{\|Az_\alpha\|^2}{R\|Gz_\alpha\|} \le \alpha \quad \text{for all sufficiently small } \alpha\,.$$

Since ξ is the inverse function of φ^2, this estimate is equivalent to the inequality

$$\|Gz_\alpha\| \le R\,\varphi^2(\alpha)\,.$$

From the last inequality and (5.2) we obtain (5.1). $\square$

The above proof is based on spectral calculus, which is frequently used in regularization theory (see, e.g., [6], [18], [19], [28], [32]). Hence, Theorem 5.1 is applicable to compact and non-compact operators G in like manner. On the other hand, the approach of the following section is focused on compact G only, because it makes explicit use of the discrete spectrum of this operator and the corresponding eigenvalue expansion.

6 Compactness

Assumption A5 seems to be the one that may be verified in the easiest way. This can be seen by considering the examples in [14] concerning inverse problems in partial differential equations. We are able to derive convergence rates from assumption A5 provided the operator G is compact.

Theorem 6.1 *Let assumption A5 hold and suppose G is compact. In addition, assume that the function $t/\varrho(t)$ is strictly monotonically decreasing in some interval $(0,\varepsilon]$ $(\varepsilon > 0)$ and $t/\varrho(t) \to \infty$ as $t \to 0$. Then, with some constant $\kappa < \infty$,*

$$\|x_\alpha - x^\dagger\| \le \kappa\,\varphi(\alpha) \;\; \textit{for} \;\; 0 < \alpha \le \overline{\alpha}\,, \tag{6.1}$$

where $\varphi(t) = \varrho^{-1}(\sqrt{t})$ for $t \in [0, \|G\|]$.

Proof. The hypotheses of the theorem ensure that the standing assumptions of [14] are satisfied with $\varrho_1(t) = \varrho(t)$ and $\varrho_2(t) = t$. Formula (3.7) of [14] says that there is a constant $K < \infty$ such that

$$\|x_\alpha - x^\dagger\| \le K\left(f^{-1}(M) + \sqrt{\alpha}\,M\right) \;\; \text{for all sufficiently large } \; M\,, \tag{6.2}$$

where $f(t) = t/\varrho(t)$. Put $M(\alpha) = \varrho^{-1}(\sqrt{\alpha})/\sqrt{\alpha} = \varphi(\alpha)/\sqrt{\alpha}$. Since $t/\varrho(t) \to \infty$ as $t \to 0$, we see that $M(\alpha) \to \infty$ as $\alpha \to 0$. We have $\varrho(\sqrt{\alpha}\,M(\alpha)) = \varrho(\varrho^{-1}(\sqrt{\alpha})) = \sqrt{\alpha}$ and hence

$$M(\alpha) = \frac{\sqrt{\alpha}\,M(\alpha)}{\varrho(\sqrt{\alpha}\,M(\alpha))} = f(\sqrt{\alpha}\,M(\alpha))\,,$$

which gives $f^{-1}(M(\alpha)) = \sqrt{\alpha}\, M(\alpha)$. Consequently, we have for sufficiently small $\alpha > 0$

$$K\left(f^{-1}(M(\alpha)) + \sqrt{\alpha}\, M(\alpha)\right) = 2K\sqrt{\alpha}\, M(\alpha) = 2K\varphi(\alpha)\,.$$

Setting $\kappa = 2\,K$ this yields (6.1). $\square$

Note that the rates $\varphi(\alpha)$ delivered by Theorem 6.1 are by construction always slower than $O(\sqrt{\alpha})$.

Obviously, the above is nothing but an adaption of the main result of [14] to the present context. We remark that the last requirement formulated in condition (2.3) of [14], in our notation $\max_{\varepsilon \le t \le \|G\|} t/\varrho(t) \le C_1 \varepsilon/\varrho(\varepsilon)$ for some constant $1 \le C_1 < \infty$, is automatically satisfied due to the continuity and positivity of the index function $\varrho(t)$ for $0 < t \le \|G\|$. Moreover, note that an essential ingredient of the proof in [14] is the error estimate

$$\|x_\alpha - x^\dagger\| \le d(M) + \sqrt{\alpha}\, M \quad \text{for all sufficiently large } M \tag{6.3}$$

with the distance function

$$d(M) \;=\; \inf\{\|x^\dagger - A^*\, v\| \,:\, v \in Y,\ \|v\| \le M\}\,. \tag{6.4}$$

In order to deduce (6.2) from (6.3), the hypotheses of Theorem 6.1 are essential. They are sufficient to estimate $d(M) \le K\, f^{-1}(M)$ with some $K < \infty$ for sufficiently large M. On the other hand, estimate (6.3) follows, for example, from Theorem 6.8 of [3]. For some more studies of this subject see also [10].

Even though the compactness of G is a severe constraint, the requirements of Theorem 6.1 concerning φ are essentially weaker than the corresponding conditions in Proposition 4.1(a) and Theorem 5.1. So, under assumption A5 with $\varrho(t) = \exp\left(-1/(2\,t^{1/p})\right)$ the logarithmic convergence rates (4.2) can be established for all $p > 0$ whenever $\|A\| < 1$, since $t/\varrho(t)$ is decreasing for sufficiently small $t > 0$ and goes to infinity as $t \to 0$. With slightly modified constants, such a situation occurs in the cases $p = 1$ and $p = 2$ in some inverse problems for the heat equation (see [14, Examples 3 and 4]).

7 Scaling

We finally embark on assumption A5 without the additional requirement that G be compact, which was essential in Section 6. From Corollary 2.2 we know that assumption A5 is equivalent to assumption A4, the only problem being that sole knowledge of the range inclusion A5 does not tell us anything about the constant C in assumption A4.

Theorem 7.1 *Let assumption A5 be satisfied or, equivalently, let us assume that assumption A4 is given. Define the index function φ by $\varphi(t) = \varrho^{-1}(\sqrt{t})$ for $t \in [0, \varrho^2(\|G\|)]$ and suppose that either (a) φ^2 is concave on $[0, \varrho^2(\|G\|)]$ or (b) φ^2 can be continued to an index function on $[0, C^2\|A\|^2]$ that is operator monotone, where C is the constant from assumption A4. Then there exists an $\overline{\alpha} > 0$ such that*

$$\|x_\alpha - x^\dagger\| \le K\,\varphi(\alpha) \quad \textit{for all } 0 < \alpha \le \overline{\alpha}\,, \tag{7.1}$$

where $K = R\,\max(C,1)$ in the case (a) and $K = k^2 R\,\max(C^2,1)$ with the constant $k \ge 1$ of Proposition 3.3(c) in the case (b).

Proof. Assumption A4 says that $\|\varrho(G)x\| \le C\|Ax\|$ for all $x \in X$. We use a scaling idea and replace the equation $Ax = y$ by the equivalent equation $CAx = Cy$. Let us denote by $\widetilde{x}_\beta$ the solution of the extremal problem $\|CAx - Cy\|^2 + \beta\|x\|^2 \to \min$. Then for the equation $CAx = Cy$, assumption A3 is valid. Hence, Theorem 5.1 in the case (a) and Proposition 4.1(a) in the case (b) apply and provide us with inequalities of the form

$$\|\widetilde{x}_\beta - x^\dagger\| \le k\, R\, \varphi(\beta) \tag{7.2}$$

for all sufficiently small β. From (1.3) we infer that

$$\widetilde{x}_\beta = (C^2 A^* A + \beta I)^{-1} C A^* C y$$

and that the solution of the problem $\|Ax - y\|^2 + \alpha\|x\|^2 \to \min$ equals

$$x_\alpha = (A^* A + \alpha I)^{-1} A^* y\,.$$

Consequently, $\widetilde{x}_{C^2\alpha} = x_\alpha$. From (7.2) we therefore obtain

$$\|x_\alpha - x^\dagger\| \le k\, R\, \varphi(C^2\alpha) \tag{7.3}$$

for all sufficiently small α. Notice that $k = 1$ in the case (a). If $C \le 1$, then (7.1) follows from the monotonicity of φ and the observation that $1 \le k \le k^2$ in the case (b). Now let $C > 1$. If φ^2 is concave, we have $\varphi^2(C^2\alpha) \le C^2\varphi^2(\alpha)$ (recall the proof of Proposition 3.3(b)). It follows that $\varphi(C^2\alpha) \le C\varphi(\alpha)$ and hence (7.3) yields (7.1). Now suppose φ^2 is operator monotone on $[0, C^2\|A\|^2]$. Then φ itself is operator monotone on the same interval (see the proof of Proposition 4.1)(a). In that case we use the inequality $\varphi(t)/t \le k\,\varphi(s)/s$ for $0 < s < t \le C^2\|A\|^2$ of Lemma 3 in [22] (recall the proof of Proposition 3.3(c)). This gives the estimate $\varphi(C^2\alpha) \le k\, C^2\varphi(\alpha)$ for all sufficiently small $\alpha > 0$, and hence (7.3) implies the desired estimate (7.1). $\square$

We remark that a convergence rate of type (7.1) can also be derived from assumption A5 if ϱ^{-1} is operator monotone (see Corollary 2.3(c) and Proposition 4.1(b)). Then, however, we only get the reduced rate $O(\sqrt{\varrho^{-1}(\sqrt{\alpha})})$.

Remark 7.2 If we focus on the Hölder rates considered in Examples 3.3 and 4.2, then it becomes clear that all of Sections 4 to 7 based on assumptions A3 to A5 yield convergence rates up to $O(\sqrt{\alpha})$. It is only source conditions of higher order in assumption A1 or A2 that can provide us with convergence rates up to $O(\alpha)$ along the lines of Section 3. This phenomenon reflects the problem that we have found no universal way to go from one of assumptions A3 to A5 or from an estimate of the form (1.4) to assumption A2 with $\varphi(t) = t^\mu$ for $1/2 < \mu \le 1$. Although the converse result mentioned in Example 3.3 could provide us with such source conditions for $\mu > 1/2$, the estimates (1.4) obtained in Sections 4 to 7 lead to rate functions φ that are never of higher order than $O(\sqrt{\alpha})$. Corollaries 2.3(b) and 2.3(c), which imply assumption A2 directly from assumption A3, lead to $O(\sqrt{\alpha})$ as highest order of convergence. In this sense, the smoothness conditions of assumptions A3 to A5 show a saturation behavior for the method of Tikhonov regularization. This seems to correspond with the exponent limitation of the Heinz-Kato inequality

$$\begin{aligned} &\|G^{\frac{1}{2\mu}}\, x\| \le \||A|\, x\| \ \text{ for all } x \in X \\ &\Longrightarrow\ \|G\, x\| \le \||A|^{2\mu}\, x\| = \|(A^*A)^\mu\, x\| \ \text{ for all } x \in X\,, \end{aligned} \tag{7.4}$$

which, as well as its version (3.4), in this general form is only true for $0 < \mu \le 1/2$. On the other hand, the proofs of Corollaries 2.3(b) and 2.3(c) show that under the extended operator monotonicity hypotheses on φ^2 with $\varphi(t) = \varrho^{-1}(\sqrt{t})$ and on ϱ^{-1} with $\varphi(t) = \sqrt{\varrho^{-1}(\sqrt{t})}$ of this corollary we have the implication

$$\|\varrho(G)\,x\| \le \||A|\,x\| = \|A\,x\| \;\text{ for all }\; x \in X \;\Longrightarrow\; \mathcal{R}(G) \subset \mathcal{R}(\varphi(A^*A))\,. \tag{7.5}$$

For non-power-type functions φ this implication is, in some sense, a generalization of the Heinz-Kato result. □

8 Commuting operators break up saturation

In this section we show that the implication (7.5) is true for all index functions ϱ whenever the operators G and A^*A commute. Thus, in the case of commuting operators assumption A2 is a consequence of assumption A4.

Proposition 8.1 *Let S and T be strictly positive selfadjoint bounded linear operators on X with $ST = TS$ and spectrum in the interval $[0, \tilde{t}\,]$. Then for every index function ψ on $[0, \tilde{t}\,]$ the implication*

$$\|Sx\| \le \|Tx\| \;\text{ for all }\; x \in X \;\Longrightarrow\; \|\psi(S)x\| \le \|\psi(T)x\| \;\text{ for all }\; x \in X \tag{8.1}$$

is valid.

Proof. Let $\mathcal{L}(X)$ be the C^*-algebra of all bounded linear operators on X and let $\mathcal{A}$ denote the smallest closed subalgebra of $\mathcal{L}(X)$ that contains S, T, and the identity operator. Since S and T are commuting selfadjoint operators, $\mathcal{A}$ is a unital commutative C^*-algebra. We denote by $\mathcal{M}$ the maximal ideal space of $\mathcal{A}$. By the Gelfand-Naimark theorem (see, e.g., [5, Theorem I.3.1]), the Gelfand transform $\Gamma : \mathcal{A} \to C(\mathcal{M})$ is an isometric star-isomorphism that preserves positivity. Put $\sigma = \Gamma S$ and $\tau = \Gamma T$. Then σ and τ are real-valued continuous functions on $\mathcal{M}$ with values in $[0, \tilde{t}\,]$.

Now suppose that $\|Sx\| \le \|Tx\|$ for all $x \in X$. This means that $S^2 \le T^2$, and hence $\sigma^2(\nu) \le \tau^2(\nu)$ for all $\nu \in \mathcal{M}$. Since σ and τ are nonnegative, it follows that $0 \le \sigma(\nu) \le \tau(\nu) \le \tilde{t}$ for all $\nu \in \mathcal{M}$. Taking into account that ψ^2 is monotonically increasing on $[0, \tilde{t}\,]$, we obtain that that $\psi^2(\sigma(\nu)) \le \psi^2(\tau(\nu))$ for all $\nu \in \mathcal{M}$. As $\Gamma^{-1}(\psi^2 \circ \sigma) = \psi^2(S)$ and $\Gamma^{-1}(\psi^2 \circ \tau) = \psi^2(T)$, we arrive at the inequality $\psi^2(S) \le \psi^2(T)$, which, by virtue of the selfadjointness of $\psi(S)$ and $\psi(T)$, implies that $\|\psi(S)x\| \le \|\psi(T)x\|$ for all $x \in X$. □

Theorem 8.2 *Let G and A^*A be commuting operators and let assumption A4 hold. Then the function ψ defined by $\psi(t) = \left(\frac{\varrho}{C}\right)^{-1}(t)$ for $t \in [0, \varrho(\|G\|)/C]$ can be continued to an index function on $[0, \|A\|]$, the function φ defined by $\varphi(t) = \psi(\sqrt{t})$ is an index function on $[0, \|A\|^2]$, and assumption A2 is valid with this function φ.*

Proof. Since the $\varrho(t)$ in assumption A4 is assumed to be an index function for $t \in [0, \|G\|]$, the function $\psi(t) = \left(\frac{\varrho}{C}\right)^{-1}(t)$ is well-defined and an index function for $t \in [0, \varrho(\|G\|)/C]$. It is always possible to continue ψ so that the resulting function is continuous and strictly increasing and hence an index function on $[0, \|A\|]$. If G and A^*A commute, then $\varrho(G)/C$

and $|A| = (A^*A)^{1/2}$ also commute. Indeed, as in the proof of Proposition 8.1, let $\mathcal{A}$ be the smallest closed subalgebra of $\mathcal{L}(X)$ that contains G, A^*A, and the identity operator. Then $\mathcal{A}$ is a unital commutative C^*-algebra that is isometrically star-isomorphic to $C(M)$ for some compact Hausdorff space M, the operators G and A^*A correspond to $\sigma \in C(M)$ and $\tau \in C(M)$, and since $\varrho \circ \sigma / C$ and $\tau^{1/2}$ are also in $C(M)$, the operators $\varrho(G)/C$ and $|A|$ belong to the commutative algebra $\mathcal{A}$. Consequently, we may combine assumption A4 and Proposition 8.1 with $S = \varrho(G)/C$, $T = |A|$, $\tilde{t} = \|A\|$ and ψ as introduced in the present theorem to obtain the inequality $\|Gx\| \le \|\varphi(A^*A)x\|$ for all $x \in X$ with φ as defined in the present theorem. From this inequality and Proposition 2.1(b) we deduce the range inclusion $\mathcal{R}(G) \subset \mathcal{R}(\varphi(A^*A))$ and thus arrive at assumption A2. $\Box$

Let ϱ be the power function defined by $\varrho(t) = t^{\frac{1}{2\mu}}$ for some $0 < \mu \le 1$. From Theorem 8.2 and the results of Section 3 we conclude that if the operators G and A^*A commute and one of the assumptions A3 to A5 is fulfilled for this ϱ, then $\|x_\alpha - x^\dagger\| = O(\alpha^\mu)$ as $\alpha \to 0$. Moreover, we see that for commuting G and A^*A the implications (3.4) and (7.4) even hold for all $\mu > 0$. Such a specific property of commuting operators was already observed in the context of regularization in [25, Remark 4] and [31, p. 2125].

Example 8.3 In this example we show that assumption A2 does not necessarily imply assumption A1 even if the operators G and A^*A commute. Let A and G be the multiplication operators on $X = L^2(0,1)$ given by $(Ax)(t) = t\,x(t)$ and $(Gx)(t) = m(t)\,x(t)$ with $m(t) = t\,(1 - \sin(1/t))$. Obviously, G and A^*A are strictly positive bounded (but non-compact) linear operators, they have non-closed ranges, and they commute. Trivially, $\mathcal{R}(G) \subset \mathcal{R}(|A|)$, so that assumption A2 holds with $\varphi(t) = \sqrt{t}$. If assumption A1 would be satisfied with some continuous function φ (possibly different from the φ of assumption A2), we would have $m(t) = \varphi(t^2)$ for all $t \in (0,1)$. But since m fails to be monotone in any right neighborhood of zero, φ fails to be an index function on $[0,\varepsilon]$ for any $\varepsilon > 0$. Hence assumption A1 cannot hold, not even with a function φ that is only required to be an index function in some small right neighborhood of zero. $\Box$

9 Parameter choice and order optimality

If instead of the exact right-hand y we have noisy data y^δ with $\|y - y^\delta\| \le \delta$, we consider the extremal problem

$$\|Ax - y^\delta\|^2 + \alpha\|x\|^2 \to \min,$$

whose unique solution $x = x_\alpha^\delta \in X$ is

$$x_\alpha^\delta = (A^*A + \alpha I)^{-1} A^* y^\delta \,. \tag{9.1}$$

It is well known (and easily seen from (1.3) and (9.1)) that $\|x_\alpha^\delta - x_\alpha\| \le \delta/(2\sqrt{\alpha})$. In summary, we have

$$\|x_\alpha^\delta - x^\dagger\| \le \|x_\alpha - x^\dagger\| + \frac{\delta}{2\sqrt{\alpha}} \,.$$

Hence, from the estimates (1.4) derived in Sections 3 to 7 for index functions φ under very different conditions we find inequalities of the form

$$\|x_\alpha^\delta - x^\dagger\| \le K\varphi(\alpha) + \frac{\delta}{2\sqrt{\alpha}} \,, \tag{9.2}$$

which estimate above the total error of regularization in the case of noisy data.

In this section, we are going to answer the question concerning the best possible worst case error for identifying $x^\dagger$ from noisy data $y^\delta \in Y$. Let $V : Y \to X$ be an arbitrary operator and think of Vy^δ as an approximation to the solution $x^\dagger$. Then, the quantity

$$\Delta(\delta, V) = \sup \left\{ \|V\, y^\delta - x^\dagger\| \ : \ x^\dagger = G\, v\,,\ \|v\| \le R\ ,\ y^\delta \in Y\ ,\ \|y - y^\delta\| \le \delta \right\}$$

is called the worst case error for identifying $x^\dagger$ under the smoothness assumption (1.5). An optimal method $V_{\rm opt}$ is characterized by $\Delta(\delta, V_{\rm opt}) = \inf_V \Delta(\delta, V)$, and this quantity is called the best possible worst case error. If assumption A4 is true as equality, then we know that the best possible worst case error can be estimated from below by

$$\inf_V \Delta(\delta, V) \ge R\varphi\left(\Theta^{-1}\left(\frac{C\delta}{R}\right)\right) \tag{9.3}$$

provided $C\delta/R$ is an element of the spectrum of the operator $G\varrho(G)$ (see [24]) and we set $\Theta(t) := \sqrt{t}\,\varphi(t)$ for all $t \ge 0$ under consideration. The error bound on the right-hand side of (9.3) cannot be beaten by any regularization method. Due to this fact, under assumption A4 we call a regularized solution x_α^δ order optimal if

$$\|x_\alpha^\delta - x^\dagger\| \le cR\varphi\left(\Theta^{-1}\left(\frac{C\delta}{R}\right)\right) \quad \text{with some } \ c \ge 1\,.$$

Let us show that the method of Tikhonov regularization provides order optimal error bounds for the a priori parameter choice

$$\alpha(\delta) = \frac{1}{c_2^2}\Theta^{-1}\left(\frac{c_2\delta}{c_1}\right)\,, \tag{9.4}$$

where c_1 and c_2 are positive constants guessing R and C, respectively.

Theorem 9.1 *Let assumption A4 be satisfied and define the index function φ by $\varphi(t) = \varrho^{-1}(\sqrt{t})$ for $t \in [0, \varrho^2(\|G\|)]$. If $\varphi^2(t)$ is concave on $[0, \varrho^2(\|G\|)]$ and $\alpha(\delta)$ is chosen a priori by (9.4), then the regularized solution $x_{\alpha(\delta)}^\delta$ from (9.1) is order optimal. In fact, there is a $\overline{\delta} > 0$ such that*

$$\|x_{\alpha(\delta)}^\delta - x^\dagger\| \le cR\varphi\left(\Theta^{-1}\left(\frac{C\delta}{R}\right)\right) \quad \textit{for all } \ 0 < \delta \le \overline{\delta} \tag{9.5}$$

with

$$c = \left(\frac{C^2}{c_2^2} + \frac{c_1}{2R}\right)\,\max\left\{1, \frac{c_2 R}{c_1 C}\right\}\,. \tag{9.6}$$

Proof. The parameter choice (9.4) can be rewritten in the equivalent form

$$\frac{\delta}{\sqrt{\alpha}} = c_1\varphi\left(c_2^2\alpha\right)\,.$$

Put $\overline{\alpha} = \varrho^2(\|G\|)/C^2$. Since $\varphi^2(t)$ is concave on $(0, \varrho^2(\|G\|)]$, we have $\varphi^2(C^2\alpha)/(C^2\alpha) \le \varphi^2(\alpha)/\alpha$ for $0 < \alpha \le \overline{\alpha}$ or equivalently, $\varphi(C^2\alpha) \le C\varphi(\alpha)$ for $0 < \alpha \le \overline{\alpha}$. This inequality together with (9.2), (9.4), and estimate (7.3) (which holds here with $k = 1$) gives

$$\begin{aligned}
\|x^\delta_{\alpha(\delta)} - x^\dagger\| &\le R\,\varphi\left(C^2\alpha(\delta)\right) + \frac{\delta}{2\sqrt{\alpha(\delta)}} \\
&= R\,\varphi\left(\frac{C^2}{c_2^2}\Theta^{-1}\left(\frac{c_2\delta}{c_1}\right)\right) + \frac{c_1}{2}\varphi\left(\Theta^{-1}\left(\frac{c_2\delta}{c_1}\right)\right) \\
&\le \left[\frac{C^2}{c_2^2} + \frac{c_1}{2R}\right] R\,\varphi\left(\Theta^{-1}\left(\frac{c_2\delta}{c_1}\right)\right). \qquad (9.7)
\end{aligned}$$

Since φ is strictly increasing, we have for arbitrary constants $\kappa > 0$

$$\varphi\left(\Theta^{-1}(\kappa\,\delta)\right) \le \max(1, \kappa)\ \varphi\left(\Theta^{-1}(\delta)\right) \qquad (9.8)$$

for sufficiently small $\delta > 0$ (see [24] and [14, Proof of Theorem 3]). Estimates (9.7) and (9.8) lead us to (9.5) and (9.6). □

Note that for $c_1 = R$ and $c_2 = C$ the constant c in (9.6) is $c = \frac{3}{2}$.

A result analogous to Theorem 9.1 can be established with $C = 1$ if assumption A4 and the concavity of φ^2 is substituted either by assumption A3 and the operator monotonicity of a continuation of φ^2 or by assumption A3 and the requirement that G and A^*A commute.

Example 9.2 We consider the Hölder rate situation with $\varphi(t) = t^\mu$ discussed in Examples 3.5 and 4.2 in the case of noisy data. The a priori parameter choice (9.4) now amounts to

$$\alpha(\delta) = \hat{K}\,\delta^{\frac{1}{2\mu+1}} \ \text{ with some constant } \ \hat{K} < \infty\,,$$

and the order optimal convergence rate of Theorem 9.1 becomes

$$\|x^\delta_{\alpha(\delta)} - x^\dagger\| = O\left(\delta^{\frac{2\mu}{1+2\mu}}\right) \qquad \text{as} \quad \delta \to 0\,. \qquad (9.9)$$

However, we have to restrict the value of μ to the interval $0 < \mu \le 1/2$ in Theorem 9.1 in order to obtain concave functions φ^2. Consequently, the best rate for noisy data that can be derived from Theorem 9.1 is $O(\sqrt{\delta})$ for $\mu = 1/2$. This specific saturation melts in the case of commuting operators G and A^*A, where $O(\delta^{2/3})$ for $\mu = 1$ is the best rate as a consequence of Theorem 8.2 whenever one of the asssumptions A3 to A5 holds with the index function $\varrho(t) = t^{\frac{1}{2\mu}}$ and $1/2 < \mu \le 1$.

In the non-commuting case, this specific saturation can be prevented if the stabilization term $\alpha\,\|x\|^2$ in the Tikhonov functional (1.2) is replaced by the stronger stabilization term $\alpha\,\|G^{-1}x\|^2$ (or more generally, by $\alpha\,\|G^{-s}x\|^2$ with $s \ge 1$), see [6], [25], [31] for the Hölder case and [24] for the general case. The remaining restriction comes from the qualification of the method of Tikhonov regularization which tells us that the best possible rate for $\|G^{-1}(x^\delta_\alpha - x^\dagger)\|$ is $O(\delta^{2/3})$. This second type of saturation can be prevented by using regularization methods with higher qualification. □

Example 9.3 The logarithmic rates (4.2) of Example 4.3 for $0 < p \le 1/2$, which were also briefly mentioned in Sections 5 and 6 for the more general case $p > 0$, can be combined with Theorem 9.1 to lead to the order optimal convergence rate

$$\|x^\delta_{\alpha(\delta)} - x^\dagger\| \;=\; O\left((-\ln\delta)^{-p}\right) \qquad \text{as} \quad \delta \to 0 \tag{9.10}$$

for the a priori parameter choice (9.4) and even for the more simple choice $\alpha \sim \delta^\theta$ with exponents $0 < \theta < 2$ (see [21, p. 208], [14, Section 5]). This convergence rate seems to be rather slow, but for severely (exponentially) ill-posed problems (1.1) (see [15] and the references therein) source conditions $x^\dagger = (A^*A)^\mu$ leading to Hölder rates are not realistic, whereas assumptions A3 to A5 with exponential functions $\varrho(t) = K_1 \exp\left(-K_2\, t^{-1/p}\right)$ with some positive constants K_1 and K_2 are well-interpretable in a couple of situations. □

10 Applications

Example 10.1 (Heat conduction backward in time) Let us illustrate our general theory by the following heat conduction problem backward in time, which has been studied, e.g., in [1]. This inverse problem consists in identifying a two-dimensional temperature profile $x^\dagger = u(\kappa_1, \kappa_2, 0) \in L^2(\mathbf{R}^2)$ at time $t = 0$ from given noisy temperature data $y^\delta \in L^2(\mathbf{R}^2)$ which correspond with exact data $y = u(\kappa_1, \kappa_2, 1)$ at time $t = 1$ such that $\|y^\delta - y\|_{L^2(\mathbf{R}^2)} \le \delta$. The associated linear forward operator A maps in $L^2(\mathbf{R}^2)$ and assigns the function $u(\kappa_1, \kappa_2, 1)$ to the function $u(\kappa_1, \kappa_2, 0)$, where u satisfies the heat equation

$$u_t - \Delta u = 0 \quad \text{for} \quad (\kappa_1, \kappa_2, t) \in \mathbf{R}^2 \times (0, \infty)\,.$$

Let us transform the operator equation (1.1) with $X = Y = L^2(\mathbf{R}^2)$ into the frequency domain by means of the unitary Fourier transform. Let $\widehat{x}(\xi_1, \xi_2) = \mathcal{F}\left(x(\kappa_1, \kappa_2)\right)$ be the Fourier transform of x, that is,

$$\widehat{x} = \mathcal{F}(x) = \frac{1}{2\pi}\int_{\mathbf{R}^2} x(\kappa_1, \kappa_2)\mathrm{e}^{-\mathrm{i}(\kappa_1\xi_1 + \kappa_2\xi_2)}\mathrm{d}\kappa_1 \mathrm{d}\kappa_2\,.$$

It results that in the frequency domain we have the equivalent operator equation $\widehat{A}\,\widehat{x} = \widehat{y}$ of the form

$$\mathrm{e}^{-(\xi_1^2 + \xi_2^2)}\widehat{x} = \widehat{y}\,. \tag{10.1}$$

From this representation we realize that the operator $\widehat{A}$ is a multiplication operator with spectrum in the segment $[0, 1]$ and that hence both $\widehat{A}$ and $A = \mathcal{F}^{-1}\widehat{A}\mathcal{F}$ are non-compact bounded linear strictly positive selfadjoint injective operators with non-closed ranges and with $\|A\| = \|\widehat{A}\| = 1$. Consequently, the operator equation (1.1) of this example is ill-posed of type I.

In order to formulate our solution smoothness (1.5) we introduce the classical Sobolev scale $(H_r)_{r\in[0,\infty)}$ by $H_0 = X = L^2(\mathbf{R}^2)$ and $H_r = \{x(\kappa_1, \kappa_2) \in H_0 : \|x\|_r < \infty\}$ where

$$\|x\|_r = \left(\int_{\mathbf{R}^2} (1 + \xi_1^2 + \xi_2^2)^r\, |\widehat{x}(\xi_1, \xi_2)|^2 \mathrm{d}\xi_1 \mathrm{d}\xi_2\right)^{1/2}.$$

We pick some $p > 0$, define the operator $\widehat{G}$ by

$$\widehat{G}\,\widehat{v}(\xi_1, \xi_2) = (1 + \xi_1^2 + \xi_2^2)^{-p/2}\widehat{v}(\xi_1, \xi_2)\,, \tag{10.2}$$

and put $G = \mathcal{F}^{-1}\widehat{G}\mathcal{F}$. With this G, condition (1.5) is equivalent to the condition $\widehat{x}^\dagger = \widehat{G}\widehat{v}$ with $\|\widehat{v}\| = \|v\| \le R$ and thus to the condition $x^\dagger \in H_p$ and $\|x^\dagger\|_p \le R$, which is clearly a smoothness condition. Note that the operator $\widehat{G}$ is a multiplication operator with spectrum in the segment $[0, 1]$ and that both G and $\widehat{G}$ are non-compact bounded linear strictly positive selfadjoint injective operators with non-closed ranges and with $\|G\| = \|\widehat{G}\| = 1$.

Let us consider the method of Tikhonov regularization in the frequency domain. By (9.1), the regularized solution $\widehat{x}_\alpha^\delta$ is given by

$$\widehat{x}_\alpha^\delta(\xi_1, \xi_2) = \frac{e^{-(\xi_1^2+\xi_2^2)}\,\widehat{y}^\delta(\xi_1, \xi_2)}{e^{-2(\xi_1^2+\xi_2^2)} + \alpha}\,.$$

Using the inverse Fourier transform we construct the regularized solution x_α^δ of our backward heat conduction problem according to $x_\alpha^\delta = \mathcal{F}^{-1}\left(\widehat{x}_\alpha^\delta\right)$. Now we are ready for the illustration of our general theory.

Due to formulas (10.1) and (10.2) we see that assumption A1 is satisfied with the function

$$\varphi(t) = \left[\ln\frac{e}{\sqrt{t}}\right]^{-p/2} \quad \text{for } 0 < t \le 1 \text{ and } \varphi(0) = 0.$$

Thus φ maps $[0, ||A||^2] = [0, 1]$ into $[0, 1]$. Moreover, the operators $\widehat{A}^*\widehat{A}$ and $\widehat{G}$ and hence A^*A and G are commuting operators. In order to apply Corollary 3.4 we have to check conditions (a) to (e) of Proposition 3.3.

(a) From $\left(\frac{t}{\varphi(t)}\right)' = \frac{1}{2}\left[1 - \frac{1}{2}\ln t\right]^{p/2-1}(2 - p/2 - \ln t)$ we conclude that $\frac{t}{\varphi(t)}$ is strictly increasing on $[0, e^{2-p/2}]$. Therefore, $\frac{t}{\varphi(t)}$ is strictly increasing on $[0, \|A\|^2] = [0, 1]$ for $p \in (0, 4]$.

(b) From $\varphi''(t) = -\frac{p}{8t^2}\left[1 - \frac{1}{2}\ln t\right]^{-p/2-2}(1 - p/2 - \ln t)$ we conclude that $\varphi(t)$ is concave on $[0, e^{1-p/2}]$. Thus, $\varphi(t)$ is concave on $[0, \|A\|^2] = [0, 1]$ for $p \in (0, 2]$.

(c) φ is operator monotone on $[0, 1]$ for $p \in (0, 2]$ (see [20]).

(d) Due to part (a), $\frac{t}{\varphi(t)}$ is increasing on $(0, \hat{t}\,]$ with $\hat{t} = e^{2-p/2}$. We conclude that (3.2) holds with $k = 1$ for $p \in (0, 4]$ and $k = \varphi(\|A\|^2)/\varphi(\hat{t}) = 1/\varphi\left(e^{2-p/2}\right) = (p/4)^{p/2}$ for $p > 4$.

(e) By part (b), $\varphi(t)$ is concave on $(0, \hat{t}\,]$ with $\hat{t} = e^{1-p/2}$. Hence we conclude that (3.2) is valid with $k = 1$ for $p \in (0, 2]$ and $k = \varphi(\|A\|^2)/\varphi(\hat{t}) = 1/\varphi\left(e^{1-p/2}\right) = [(2+p)/4]^{p/2}$ for $p > 2$.

Consequently, we may have recourse to Corollary 3.4 for arbitrary $p \in (0, \infty)$. Combining Proposition 3.2 and Proposition 3.3(d) we get the error estimate

$$\|\widehat{x}_\alpha - \widehat{x}^\dagger\| \le kR\varphi(\alpha) = kR\left[\ln\frac{e}{\sqrt{\alpha}}\right]^{-p/2}$$

with $k = 1$ for $p \in (0, 4]$ and $k = (p/4)^{p/2}$ for $p \ge 4$. For the a priori parameter choice

$$\frac{\delta}{\sqrt{\alpha}} = c_1\left[\ln\frac{e}{\sqrt{\alpha}}\right]^{-p/2} \tag{10.3}$$

we obtain for the total error $\|x_\alpha^\delta - x^\dagger\| = \|\widehat{x}_\alpha^\delta - \widehat{x}^\dagger\|$ the order optimal error estimate

$$\|x_\alpha^\delta - x^\dagger\| \le cR\left[\ln \frac{eR}{\delta}\right]^{-p/2} (1+o(1)) \quad \text{as} \quad \delta \to 0 \tag{10.4}$$

whenever $\|x^\dagger\|_p \le R$ with $p \in (0,\infty)$.

By Corollary 2.3(a) we know that our example with assumption A1 also satisfies assumption A3 with

$$\varrho(t) = \sqrt{\varphi^{-1}(t)} = \exp\left(1 - t^{-2/p}\right) \quad \text{for} \;\; 0 < t \le 1 \;\; \text{and} \;\; \varrho(0) = 0\,,$$

so that $\varrho : [0, ||G||] = [0,1] \to [0,1]$. However, the results of Sections 4 and 5 impose constraints. Namely, $\varphi^2(t) = 2^p \left[\ln\left(\frac{e^2}{t}\right)\right]^{-p}$ with $\varphi^2(0) = 0$ is only operator monotone if $0 < p \le 1$, so that Proposition 4.1(a) is applicable only for that interval of the exponent p. In the same manner, Theorem 5.1 applies only if φ^2 is concave on $[0, \varrho^2(\|G\|)] = [0,1]$. Now φ^2 is concave just on the interval $[0, e^{1-p}]$. This interval is a subset of $[0,1]$ if and only if $0 < p \le 1$. □

Example 10.2 (Interplay of compact and non-compact operators) Here we consider the linear operator equation (1.1) in the case where $X = Y = L^2(0,1)$ and A is a multiplication operator, $(Ax)(t) = m(t)x(t)$. We suppose that $m \in L^\infty(0,1)$ and that $m(t) > 0$ for almost all $t \in (0,1)$. If $\text{essinf}_{t\in[0,1]} m(t) = 0$, then A has non-closed range. Clearly, A is a non-compact strictly positive selfadjoint bounded linear operator.

In [11] it was proved (even for the more general case of increasing rearrangements of m) that a single essential zero of m at $t = 0$ with a limited decay rate for $m(t) \to 0$ as $t \to 0$ of the form

$$m(t) \;\ge\; \underline{c}\, t^\kappa \;\; \text{a.e. on} \;\; [0,1] \;\; \text{with an exponent} \;\; \kappa > 1/4 \tag{10.5}$$

and a constant $\underline{c} > 0$ provides an order optimal estimate

$$\|x_\alpha - x^\dagger\| \;=\; O\left(\alpha^{\frac{1}{4\kappa}}\right) \quad \text{as} \quad \alpha \to 0 \tag{10.6}$$

for the Tikhonov regularization whenever

$$x^\dagger \in L^\infty(0,1)\,. \tag{10.7}$$

Hence, the order optimal convergence rate does not depend on the smoothing properties of the linear operator A as in the case of compact Fredholm integral operators, but basically on the behavior of the multiplier function $m(t)$ in a neighborhood of zero (see also [10]).

Now we consider the family of multiplier functions

$$m(t) \;=\; t^\kappa \quad \text{for} \quad \kappa > 1/4 \tag{10.8}$$

and a Hilbert scale $(H_r)_{r\in\mathbf{R}}$ such that $H_r = W^{r,2}(0,1)$ for all $0 \le r \le n$ and some fixed but arbitrarily large integer n. This means that the scale elements H_r and the usual Hilbertian Sobolev spaces $W^{r,2}(0,1)$ of order r on the unit interval coincide for $r \in [0,n]$.

In [27] it is shown that a Hilbert scale with this property exists. Due to the compactness of the embedding operator from $W^{1,2}(0,1)$ into $L^2(0,1)$ the selfadjoint and strictly positive linear operator G with $\mathcal{R}(G^r) = H_r$ for all $0 \le r \le n$ that generates the Hilbert scale is compact. With this G, condition (1.5) takes the form $x^\dagger \in W^{1,2}(0,1)$.

Certainly, our assumption A1 can never be satisfied if one of the operators A^*A and G is compact and the other is non-compact. Hence assumption A1 cannot hold. So let us analyze whether assumptions A5 and A2 are satisfied for the operators A generated by the multipliers (10.8). In this context, we set $\mu_0 = 1/(2n)$ and consider two cases: (a) $1/4 < \kappa < 1/2$ and (b) $\kappa \ge 1/2$.

Case (a). If $x(t)$ is a function in $L^\infty(0,1)$ and $0 < \kappa < 1/2$, then $x(t)/t^\kappa$ is a function in $L^2(0,1)$. This implies that

$$L^\infty(0,1) \subset \mathcal{R}(|A|)\,.$$

We have $\mathcal{R}(G^{\frac{1}{2\mu}}) = W^{\frac{1}{2\mu},2}(0,1) \subset L^\infty(0,1)$ for all $0 < \mu_0 \le \mu < 1$ and also $\mathcal{R}(G^{\frac{1}{2\mu}}) \subset L^\infty(0,1)$ for all $0 < \mu < \mu_0$, since the Sobolev spaces of order $r > 1/2$ contain only continuous functions. So we arrive at assumption A5 in the form

$$\mathcal{R}(\varrho(G)) \subset \mathcal{R}(|A|) \quad \text{for} \quad \varrho(t) = t^{\frac{1}{2\mu}} \quad \text{and} \quad 0 < \mu < 1\,. \tag{10.9}$$

Theorem 6.1 is applicable if $0 < \mu < 1/2$ and Theorem 7.1 can be used for $0 < \mu \le 1/2$. They deliver the convergence rate $O(\alpha^\mu)$, which is not better than $O(\sqrt{\alpha})$. From (3.4) and (10.9) we obtain assumption A2 in the form

$$\mathcal{R}(G) \subset \mathcal{R}(\varphi(A^*A)) \quad \text{with} \quad \varphi(t) = t^\mu\,, \tag{10.10}$$

but again only for $0 < \mu \le 1/2$. Thus, from Example 3.5 we get once more at most the convergence rate $O(\sqrt{\alpha})$. For $1/2 < \mu < 1$ we cannot say anything. If, however, $x^\dagger \in \mathcal{R}(G) = W^{1,2}(0,1)$, then (10.7) holds and hence we have (10.6), which is a higher convergence rate, since $\frac{1}{4\kappa} > \frac{1}{2}$. Thus, there is a gap between the optimal convergence rate and the convergence rate obtained from our assumptions A5.

On the other hand, we remark that assumption A2 in the form (10.10) is valid if $0 < \mu < \frac{1}{4\kappa}$, since if $x \in L^\infty(0,1)$ and $0 < 4\mu\kappa < 1$, then $x(t)/t^{2\mu\kappa} \in L^2(0,1)$. So the validity of our assumption A2 in the form (10.10) provides us with the convergence rates $O(\alpha^{\frac{1}{4\kappa}-\varepsilon})$ with arbitrarily small $\varepsilon > 0$, which makes the gap to the optimal order arbitrarily small.

Since constant functions are in $\mathcal{R}(G)$, but $\mathrm{const}/t^{2\mu\kappa}$ is not a function in $L^2(0,1)$ for $\mu \ge \frac{1}{4\kappa}$, we have

$$\mathcal{R}(G) \not\subset \mathcal{R}(|A|^{2\mu}) \quad \text{for} \quad \mu \ge \frac{1}{4\kappa}\,.$$

Hence, for $\varphi(t) = t^\mu$ with μ from the interval $[\frac{1}{4\kappa}, 1)$ and the corresponding functions $\varrho(t) = t^{\frac{1}{2\mu}}$ assumption A5 does not imply assumption A2. This, however, would be the case for commuting operators because of Theorem 8.2. Consequently, the operators A^*A and G cannot commute here.

Case (b). The constant functions are in $\mathcal{R}(G^{\frac{1}{2\mu}})$ for $0 < \mu_0 \le \mu \le 1$, but if $\kappa \ge 1/2$, then const/t^κ is not a function in $L^2(0,1)$. This implies that

$$\mathcal{R}(\varrho(G)) \not\subset \mathcal{R}(|A|) \quad \text{for} \quad \varrho(t) = t^{\frac{1}{2\mu}} \quad \text{and} \quad 0 < \mu_0 \le \mu \le 1\,,$$

and hence assumption A5 cannot hold for power functions ϱ of this family. On the other hand, in case (b) assumption A2 in the form (10.10) just holds for $0 < \mu < \frac{1}{4\kappa} < \frac{1}{2}$. As in case (a), (10.10) allows us to derive the convergence rates $O(\alpha^{\frac{1}{4\kappa}-\varepsilon})$. These rates are almost optimal, since (10.6) is optimal under assumptions (10.5) and (10.7) and since we have the converse results mentioned in Example 3.5.

Note that the relative severity of ill-posedness of (1.1) with respect to the operator G grows in this example when κ grows. A consequence of this fact is the violation of assumption A5 for power functions ϱ in case (b) in contrast to case (a). If the operators G defining the solution smoothness and A^*A defining the degree of ill-posedness of the problem (1.1) are sufficiently different in character, then assumption A5 is rarely satisfied. Regardless of this fact convergence rates of Tikhonov regularization may occur accidentally if the specific $x^\dagger$ fits the smoothing property of A as in this case. $\square$

References

[1] Ang, D. D., Gorenflo, R., Le, V. K. and Trong, D. D. (2002): *Moment Theory and Some Inverse Problems in Potential Theory and Heat Conduction*, Lect. Notes Math. 1792, Berlin: Springer.

[2] Bakushinsky, A. and Goncharsky, A. (1994): *Ill-Posed Problems: Theory and Applications*, Dordrecht: Kluwer Academic Publishers.

[3] Baumeister, J. (1987): *Stable Solution of Inverse Problems*, Braunschweig: Vieweg.

[4] Bhatia, R. (1997): *Matrix Analysis,* New York: Springer.

[5] Davidson, K. R. (1996): *C^*-Algebras by Example.* Fields Institute Monographs, Providence: American Mathematical Society.

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

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

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

[9] Hofmann, B. (1986): *Regularization of Applied Inverse and Ill-Posed Problems*, Leipzig: Teubner.

[10] Hofmann, B. (2005): *Approximate source conditions in Tikhonov-Phillips regularization and consequences for inverse problems with multiplication operators*, Paper submitted.

[11] Hofmann, B. and Fleischer, G. (1999): *Stability rates for linear ill-posed problems with compact and non-compact operators*, Z. Anal. Anw. **18**, 267–286.

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

[13] Hofmann, B. and von Wolfersdorf, L. (2005): *Some results and a conjecture on the degree of ill-posedness for integration operators with weights*, Inverse Problems **21**, 427–433.

[14] Hofmann, B. and Yamamoto, M. (2005): *Convergence rates for Tikhonov regularization based on range inclusions*, Inverse Problems **21**, 805–820.

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

[16] Kirsch, A. (1996): *An Introduction to the Mathematical Theory of Inverse Problems*, New York: Springer.

[17] Kress, R. (1989): *Linear integral equations*, Berlin: Springer.

[18] Louis, A. K. (1989): *Inverse und schlecht gestellte Probleme*, Stuttgart: Teubner.

[19] Mair, B. A. (1994): *Tikhonov regularization for finitely and infinitely smoothing operators*, SIAM J. Math. Anal. **25**, 135–147.

[20] Mathé, P. and Pereverzev, S. V. (2002): *Moduli of continuity for operator valued functions*, Numer. Funct. Anal. Optim. **23**, 623–631.

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

[22] Mathé, P. and Pereverzev, S. V. (2004): *Regularization of some linear ill-posed problems with discretized random noisy data*, Paper submitted.

[23] Morozov, V. A. (1966): *On the solution of functional equations by the method of regularization*, Soviet Math. Dokl. **7**, 414–417.

[24] Nair, M. T., Pereverzev, S. V. and Tautenhahn, U. (2005): *Regularization in Hilbert scales under general smoothing conditions*, RICAM Report 2005-09, Linz (Austria): Johann Radon Institute of Computational and Applied Mathematics.

[25] Natterer, F. (1984): *Error bounds for Tikhonov regularization in Hilbert scales*, Appl. Anal. **18**, 29–37.

[26] Nashed, M. Z. (1987): *A new approach to classification and regularization of ill-posed operator equations.* In: *Inverse and Ill-posed Problems* (Eds.: H. W. Engl and C. W. Groetsch), Orlando: Acad. Press 1987, 53–75.

[27] Neubauer, A. (1988): *When do Sobolev spaces form a Hilbert scale?*, Proc. Amer. Math. Soc. **102**, 587–592.

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

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

[30] Tanabe, H. (1979): *Equations of Evolution*, London: Pitman.

[31] Tautenhahn, U. (1996): *Error estimates for regularization methods in Hilbert scales*, SIAM J. Numer. Anal. **33**, 2120–2130.

[32] Tautenhahn, U. (1998): *Optimality for ill-posed problems under general source conditions*, Numer. Funct. Anal. & Optimiz. **19**, 377–398.

[33] Tikhonov, A. N. (1963): *Solution of incorrectly formulated problems and the regularization method*, Soviet Mathematics Doklady **4**, 1035–1038.

[34] Tikhonov, A. N. and Arsenin, V. Y. (1977): *Solution of Ill-Posed Problems*, New York: Wiley.

[35] Weidmann, J. (2000): *Lineare Operatoren in Hilberträumen, Vol. I, Grundlagen*, Stuttgart: Teubner.

[36] Vainikko, G. M. and Veretennikov, A. Y. (1986): *Iteration Procedures in Ill-Posed Problems*, Moscow: Nauka (In Russian).