$\mathcal{Y}$-Scale Regularization*

Herbert Egger

Spezialforschungsbereich SFB F013/F1308, Johann Radon Institute for Computational and Applied Mathematics, A-4040 Linz, Austria
email: herbert.egger@oeaw.ac.at

Abstract

Inverse problems are usually ill-posed in the sence that their solution is unstable with respect to data perturbations and so-called regularization methods have to be used for their stable solution. Two drawbacks of standard regularization methods are

- saturation, i.e., only suboptimal approximations can be found for smooth solutions. This is the case, e.g., for Tikhonov regularization.
- the large number of iterations, e.g., for Landweber iteration.

A framework that allows to overcome both effects for certain classes of inverse problems is regularization in Hilbert scales. There, the solution is searched in a different space out of a scale of spaces (Hilbert scale) over the pre-image space, but convergence is achieved in the original space. Regularization methods in Hilbert scales can be viewed as modified (preconditioned) versions of standard methods.

In order to make the advantages of the Hilbert scale approach applicable to a new class of problems, we propose to use a scale of spaces over the image space instead. This result in a new family of $\mathcal{Y}-$scale regularization methods, whose (optimal) convergence properties are analyzed. One of the key steps in the analysis is the formulation of an adequate a-posteriori stopping rule, which provides optimal convergence rates. The theoretical results are illustrated in several numerical examples.

1 Introduction

Let $\mathcal{X}$ and $\mathcal{Y}$ denote Hilbert spaces and $T \in L(\mathcal{X}, \mathcal{Y})$ be a linear bounded injective operator ($\|T\| \le 1$ w.l.o.g). We consider the solution of inverse problems

$$Tx = y^\delta, \tag{1}$$

from perturbed measurements y^δ satisfying

$$\|y - y^\delta\| \le \delta, \tag{2}$$

where $y = Tx^\dagger \in \mathcal{R}(T)$ denotes the unperturbed data corresponding to the *true* solution $x^\dagger$, i.e., $y = Tx^\dagger$. It is well-known (cf., e.g., [5, 7]) that in case $\mathcal{R}(T)$ is not closed in $\mathcal{Y}$, problem (1) is *ill-posed*, i.e., (1) might not be solvable, and even if it is, a solution may not be unique and does in general not depend stably on the data noise level δ; the Moore-Penrose inverse $T^\dagger$ is only densly defined and unbounded. For a stable (approximate) solution of (1) we apply *regularization* methods of the form

$$x_\alpha^\delta := R_\alpha y^\delta := g_\alpha(T^*T)T^*y^\delta, \tag{3}$$

generated by a family of piecewise continuous (filter-) functions $\{g_\alpha\}_{\alpha \in \mathcal{R}_+}$. A minimal property that ensure convergence of the regularized solutions x_α^δ towards $x^\dagger$ with $\alpha \to 0$ at least for unperturbed

*supported by the Austrian Science Fund "Fonds zur Förderung der wissenschaftlichen Forschung (FWF)" under grant SFB F013, "Numerical and Symbolic Scientific Computing", Project F1308.

data $\delta = 0$ is that the functions $g_\alpha(\lambda) \to 1/\lambda$ as $\alpha \to 0$ pointwise for all $\lambda > 0$. For a convergence (rate) analysis in case of perturbed data, the following stronger assumptions are usually made, cf. [5]:

$$\begin{aligned} \sup_{0\le\lambda\le1} |g_\alpha(\lambda)| &\le c_* \alpha^{-1}, & (4)\\ \sup_{0\le\lambda\le1} \lambda^\mu |r_\alpha(\lambda)| &\le c_\mu \alpha^\mu, \qquad 0 \le \mu \le \mu_0. & (5) \end{aligned}$$

Here $r_\alpha(\lambda) := 1 - \lambda g_\alpha(\lambda)$, and c_μ, c_* are positive constants independent of $\alpha \in \mathbb{R}_+$. A maximal constant μ_0 for which (5) holds is usually called *qualification* of the method. In case of iterative regularization methods, the regularization parameter α will be replaced by $\frac{1}{k}$ respectively $\frac{1}{k^2}$ (see the examples below). Moreover, condition (4) already follows from (5) by Markov's inequality if $g_\alpha =: g_n$ is a polynomial, which is typically the case for iterative regularization methods.

For illustration and later reference, we mention some important examples of regularization methods of this form:

1. *Tikhonov regularization [6, 27]:* The approximate solutions are defined by

$$x_\alpha^\delta = (T^*T + \alpha)^{-1} T^* y^\delta. \tag{6}$$

Tikhonov regularization is of the form (3) with $g_\alpha(\lambda) := (\lambda + \alpha)^{-1}$, and (5) holds for all $\mu \le 1$, i.e., the qualification of Tikhonov regularization is $\mu_0 = 1$.

2. *Landweber iteration [9, 13]:* The approximations x_n^δ are defined by the iteration

$$x_{k+1}^\delta = x_k^\delta + \omega T^*(y^\delta - T x_k^\delta), \qquad x_0^\delta = 0.$$

Replacing α by $\frac{1}{k}$, Landweber iteration has the form (3) with $g_k(\lambda) = \sum_{i=0}^{k-1} (1 - \omega\lambda)^i$ and $r_k(\lambda) = (1 - \omega\lambda)^k$, respectively, and satisfies (5) for all $\mu \ge 0$.

3. *The ν-methods by Brakhage [2, 8]:* (Semi-) iterative methods whose residual polynomials r_k form an orthogonal sequence with respect to some positive weight function satisfy a three-term-recurrence, which also carries over to the iterates, i.e., there exist sequences μ_k and ω_k such that

$$x_k^\delta = x_{k_m}^\delta + \mu_k(x_{k_m}^\delta - x_{k-2}^\delta) + \omega_k T^*(y^\delta - T x_{k_m}^\delta), \quad k \ge 1,$$

with $x_0^\delta = x_{-1}^\delta = x_0$. The choice $\mu_1 = 0$, $\omega_1 = (4\nu + 2)/(4\nu + 1)$ and

$$\begin{aligned} \mu_k &= \tfrac{(k-1)(2k-3)(2k+2\nu-1)}{(k+2\nu-1)(2k+4\nu-1)(2k+2\nu-3)}, \\ \omega_k &= 4\tfrac{(2k+2\nu-1)(k+\nu-1)}{(k+2\nu-1)(2k+2\nu-1)}, \qquad k > 1 \end{aligned}$$

yields the ν-methods by Brakhage. Each ν-method satisfies (5) (with α replaced by $\frac{1}{k^2}$ for $0 \le \mu \le \nu$.

Under condition (4) the regularized solutions x_α^δ converge to $x^\dagger$ with $\delta \to 0$ if $\alpha \to 0$ such that $\delta^2/\alpha \to 0$. In general, this convergence will be arbitrarily slow [23], and convergence rates can be obtained if (and only if) the true solution satisfies certain smoothness requirements, e.g., for $x^\dagger \in \mathcal{R}((T^*T)^\mu)$ the rate

$$\|x_n^\delta - x^\dagger\| = O(\delta^{\frac{2\mu}{2\mu+1}}) \tag{7}$$

holds (and is optimal) for the parameter choice

$$\alpha = \alpha(\delta) \sim \delta^{\frac{2}{2\mu+1}} \tag{8}$$

and for $\mu \le \mu_0$. If $\mu > \mu_0$, only a rate of $\delta^{\frac{2\mu_0}{2\mu_0+1}}$ can be expected; this phenomenon is called *saturation*. Tikhonov regularization, for instance, does not provide a better rate than $\delta^{\frac{2}{3}}$, while for Landweber iteration the rate may be arbitrarily close to δ if μ is sufficiently large. Note that for an iterative method like Landweber iteration the number of iterations yielding opptimal convergence rates is $n_* = O(\delta^{-\frac{2}{2\mu+1}}) \sim O(\delta^{-2})$ if μ is small. For $\delta = 10^{-3}$ one would thus need about 10^6 iterations to stay within the regime of optimal convergence.

One way to overcome saturation or to the reduce the number of iterations is regularization in Hilbert scales introduced by Natterer [18] for Tikhonov regularization, and later analyzed for more general regularization methods in [5, 11, 25] and also for nonlinear problems [21, 22, 26]. We will recall the definition of a Hilbert scale and the most important results on regularization in Hilbert scales in Section 2. The general idea is to relate the operator T to a simpler operator L, e.g., a differential operator. For the standard analysis of regularization in Hilbert scales a condition

$$\underline{m}\|y\| \leq \|L^a T^* y\| \leq \overline{m}\|y\|, \qquad \text{for all} \quad y \in \mathcal{Y} \tag{9}$$

for some a, $\underline{m}$, $\overline{m} > 0$ is required, and instead of (3) the regularized solutions are defined by the modified (preconditionied) method

$$x_\alpha^\delta := g_\alpha(L^{-2s}T^*T)L^{-2s}T^*y^\delta, \qquad s \geq -a/2. \tag{10}$$

For a short illustration of the meaning of condition (9) we give a short example.

Example 1 Let $T : L_2(0,1) \to L_2(0,1)$ denote the solution operator to

$$-qy'' = x, \quad \text{in} \quad [0,1]; \quad u(0) = u(1) = 0, \tag{11}$$

i.e., $Tx = y$. We assume that $q \in L_2(0,1)$ satisfies $0 < q_0 \leq q \leq q_1$ for some positive constants q_0, $q_1 \in \mathbb{R}$. One can show that $\mathcal{R}(T) = H^2(0,1) \cap H_0^1(0,1)$, and by the closed range theorem T is a homeomorphism between the spaces $L_2(0,1)$ and $H^2(0,1) \cap H_0^1(0,1)$.

Now let $L : \mathcal{D}(L) = H^2(0,1) \cap H_0^1(0,1) \to L_2(0,1)$ be defined by $Lx = -x''$. Since L is self-adjoint and positive definit, one can define fractional powers of L and $\|L^s x\|$ are natural norms on a scale of (Sobolev) spaces, cf. Definition 1. Condition (9) cannot hold for any $a \geq 1/4$ if q has jumps. If however $q \in H^2(0,1)$, then also (9) holds for $a = 1$. On the other hand, a condition

$$\underline{m}\|x\| \leq \|L^a T x\| \leq \overline{m}\|x\| \qquad \text{for all with} \quad x \in L_2(0,1) \tag{12}$$

holds with $a = 1$ and for some positive constants $\underline{m}$, $\overline{m}$ also for non-smooth parameters q. According to (12) one might call T smoothing like two times integration, while (9) would only suggest smoothing properties like intgration of power $1/2$ at most.

Similar examples, where (12) characterizes the smoothing properties of T more tightly than (9) arise from integral operators $(Tx)(s) := \int k(s,t)x(t)dt$ with kernels $k(s,t)$ that allow differentiation with respect to s but not with respect to t, see Section 5 for a detailed example.

As shown in [5, 18, 25], saturation effects of a method g_α can be overcome by choosing $s > 0$ in (10); in this case L^{-2s} is smoothing. Conversely, for $s < 0$ the operator L^{-2s} acts as a preconditioner, and in this way iterative regularization methods can be accelerated, cf. [3, 4]. Note, that for $s < 0$ the right inequality in (9) is already needed to ensure well-definedness of the method (10).

Condition (9) characterizes the smoothing properties of the operator T^* in terms of (the scale of spaces over the pre-image space $\mathcal{X}$ generated by) the operator L. Hence, regularization in Hilbert scales is applicable if the range of T^* consists of sufficiently smooth functions. As Example 1 ilustrates, condition (9) may however not reflect the smoothing properties of T correctly, but (12) may be more appropriate. Therefore, we will use (12) to analyse modified regularization methods of the form

$$x_\alpha^\delta := g_\alpha(T^*L^{-2s}T)T^*L^{-2s}y^\delta, \qquad s \geq -a/2. \tag{13}$$

In this way, the advantages of Hilbert scale regularization – overcoming saturation by choosing $s > 0$ respectively preconditioning by setting $s < 0$ – can be made accessible for a new class of problems. Note that for methods (13) L is an operator on the image space $\mathcal{Y}$ and generates a scale of spaces over $\mathcal{Y}$.

The outline of the paper is as follows: We start by recalling the basic results on regularization in Hilbert scales. In Section 3, we discuss regularization methods of the form (13), which we call *$\mathcal{Y}$-scale regularization* in the sequel, and derive the main error estimates. Section 4 is then concerned with the analysis of $\mathcal{Y}-$scale regulariaztion methods under relaxed assumptions with some emphasis on the preconditioning of iterative regularization methods and the design of an order optimal a-posteriori stopping rule. The presentation concludes with some examples and numerical tests illustrating the theoretical results.

2 Regularization in Hilbert scales

Before we quote the most important results concerning regularization in Hilbert scales, we shortly recall the definition of a Hilbert scale, see [12] or [5, Section 8.4] for details:

Definition 1 *Let L be a densly defined, unbounded, selfadjoint and strictly positive operator in a Hilbert space $\mathcal{Z}$, and $\mathcal{M} := \bigcap_{k=0}^{\infty} D(L^k)$. By $\mathcal{Z}_s$ we denote the completion of $\mathcal{M}$ with respect to the norm $\|z\|_s := \|L^s z\|_{\mathcal{Z}}$; $(\mathcal{Z}_s)_{s\in\mathbb{R}}$ is called a Hilbert scale (induced by L).*

Note that obviously $\|z\|_0 = \|z\|_{\mathcal{Z}}$. Two implications of this construction are that $\mathcal{Z}_{-s} = \mathcal{Z}_s'$ and that the following interpolation inequality holds, cf., e.g., [5, Section 8.4]:

$$\|z\|_r \leq \|z\|_q^{\frac{s-r}{s-q}} \|z\|_s^{\frac{r-q}{s-q}}, \tag{14}$$

for $-\infty < q < r < s < \infty$ and $z \in \mathcal{Z}_s$.

Remark 1 Like in Example 1, L is chosen to be a (simple) differential operator in most situations and the spaces $\mathcal{Z}_s$ coincide with standard Sobolev spaces for a certain range of values of s. Note, that the Sobolev spaces H^s do not form a Hilbert scale in general, cf. [20]. Fractional powers of L can then be efficiently implemented by Fourier transfom or multi level techniques.

Another choice of L always yielding (9) with $a = 1$ is $L := (T^*T)^{-\frac{1}{2}}$. At least for preconditioning ($s < 0$, e.g., $s = -a/2$), such a choice is impractical, since applying $(T^*T)^{-\frac{1}{2}}$ in (10) or (13) is at least as difficult as solving the original problem (1). If however T is self-adjoint, then $L^{-2s} = (T^*T)^{-\frac{1}{2}}$ can formally be applied by skipping the terms $L^{-2s}T^* = I$, and the method (3) simplifies to

$$x_\alpha^\delta = g_\alpha(T)y^\delta.$$

The standard assumption for regularization in Hilbert scales is the following norm equivalence condition:

$$\underline{m}\|x\|_{-a} \leq \|Tx\| \leq \overline{m}\|x\|_{-a} \tag{15}$$

for some some $\underline{m}$, $\overline{m}$, $a > 0$. In fact, this condition is equivalent to (9) (cf. [4]; a similar result is derived in Proposition 1). As shown in [5, 25], the Hilbert scale method (10) yields optimal convergence rates

$$\|x_n^\delta - x^\dagger\| = O(\delta^{\frac{u}{a+u}}) \tag{16}$$

under the source condition $x^\dagger \in \mathcal{X}_u$ for $0 < u \leq 2(a+s)\mu_0$, where μ_0 denotes the qualification of the method under consideration and

$$\alpha(\delta) \sim \delta^{\frac{2(a+s)}{a+u}}. \tag{17}$$

By setting $u = 2a\mu$, one can see that the rate (16) coincides with the standard results (7); note that under assumption 15 the source condition $x^\dagger \in \mathcal{X}_u$ is then equivalent to $x^\dagger \in \mathcal{R}((T^*T)^\mu)$ for $0 < u \leq a$. Since the rate (17) holds for $u \leq 2(a+s)\mu_0$, saturation can be overcome by choosing s large enough.

A different motivation for regularization in Hilbert scale has been geiven in [3, 4]: If an iterative regularization method is used for the solution of (1), then optimal regularization parameter choice (17) yields that the number of iterations needed to obtain the optimal convergence rates increases with increasing s. Thus, from a numerical point of view (in particular for non-smooth solutions) it is advantageous to choose s as small as possible, eventually $s < 0$. In this case the operator L^{-2s} in (10) acts as a proconditioner, and by setting $s = -a/2$, the number of iterations needed to guarantee the optimal rates (16) can be reduced to the square root, e.g., $n(\delta) = O(\delta^{-\frac{a}{a+u}})$ for Landweber iteration. Additionally, the analysis can be carried out under the weaker condition

$$\|Tx\| \leq \overline{m}\|x\|_{-a} \quad \text{respectively} \quad \|T^*y\|_a \leq \overline{m}\|y\|_{\mathcal{Y}}. \tag{18}$$

which allows to apply the approach for a significantly wider class of problems.

We will turn now to the analysis of $\mathcal{Y}-$scale methods (13) and show that the basic convergence results for regularization in Hilbert scales can be carried over to problems satisfying (12) instead of (9).

3 $\mathcal{Y}$-scale regularization

Let $\{\mathcal{Y}_s\}_{s\in\mathbb{R}}$ be a Hilbert scale induced by some densely defined selfadjoint strictly positive operator $L : \mathcal{D}(L) \subset \mathcal{Y} \to \mathcal{Y}$ (cf. Definition 1), and let $\|\cdot\|_r$ denote the norm of $\mathcal{Y}_r$, i.e., $\|y\|_r = \|L^r y\|$. In the sequel we require that the operator T satisfies the following assumption:

Assumption 1 *There exist positive real constants $\underline{m}$, $\overline{m}$ and a, such that*

$$\underline{m}\|x\| \leq \|Tx\|_a \leq \overline{m}\|x\| \tag{19}$$

holds uniformly for all $x \in X$.

Note that Assumption 1 already implies injectivity of the operator T.

For the stable solution of the inverse problem (1) we consider modified regularization methods of the form

$$x_\alpha^\delta := g_\alpha(T^*L^{-2s}T)T * L^{-2s}y^\delta = g_\alpha(B^*B)B^*L^{-s}y^\delta, \qquad \text{with} \quad B := L^{-s}T. \tag{20}$$

Here and below we assume that the functions g_α satisfy the standard assumptions (4), (5); $s \in \mathbb{R}$ will be specified later. Throughout, A^* denotes the adjoint of a linear operator A with respect to the spaces $\mathcal{X}$ and $\mathcal{Y}$. Under Assumption 1, the operator B is bounded as operator from $\mathcal{X}$ to $\mathcal{Y}$ for $s \geq -a$. Without loss of generality, we will assume $\|B\| \leq 1$ below, which can always be achieved by proper scaling.

Remark 2 Problem (1) is equivalent to

$$Bx = L^{-s}y^\delta, \tag{21}$$

where B has to be understood as operator between $\mathcal{X}$ and $\mathcal{Y}$, respectively T in (1) as an operator from $\mathcal{X}$ to $\mathcal{Y}_s$. Note that T^*L^{-2s} is the adjoint of T with respect to these spaces. The $\mathcal{Y}$-scale method (20) hence corresponds to applying the regularization method defined by g_α to the solution of (21). Note that in case $s < 0$, the right hand side $L^{-s}y^\delta$ may not be an element of $\mathcal{Y}$. In any case the noise level estimate (2) is not given in the appropriate norm if $s \neq 0$.

For the error analysis of $\mathcal{Y}$-scale methods, we will need some preliminary results:

Proposition 1 *Assumption 1 is equivalent to*

$$\underline{m}\|y\|_{-a} \leq \|T^*y\| \leq \overline{m}\|y\|_{-a}. \tag{22}$$

PROOF. First assume that (19) holds, let $y \in \mathcal{Y}$ and observe that

$$\sup_{y\in\mathcal{Y}} \frac{\|T^*y\|}{\|y\|_{-a}} = \sup_{y\in\mathcal{Y}} \sup_{x\in\mathcal{X}} \frac{\langle T^*y, x\rangle}{\|x\|\,\|y\|_{-a}} = \sup_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} \frac{\langle y, Tx\rangle}{\|x\|\,\|y\|_{-a}} \leq \sup_{x\in\mathcal{X}} \frac{\|Tx\|}{\|x\|} \leq \overline{m}.$$

Hence, T^* is extendable as a bounded linear operator to $\mathcal{Y}_{-a}$. The result than follows similarly as above by noting that

$$\sup_{y\in\mathcal{Y}_{-a}} \frac{\|T^*y\|}{\|y\|_{-a}} = \sup_{x\in\mathcal{X}} \frac{\|Tx\|}{\|x\|}.$$

The other implication follows similarly. □

Proposition 2 *Let Assumption 1 hold and $s \geq -a$. Then for $|\nu| \leq 1$*

$$\underline{c}(\nu)\,\|y\|_{-\nu(a+s)} \leq \|(BB^*)^{\frac{\nu}{2}}y\| \leq \overline{c}(\nu)\,\|y\|_{-\nu(a+s)}$$

holds for $y \in \mathcal{D}((BB^*)^{\frac{\nu}{2}})$ with $\underline{c}(\nu) = \min(\underline{m}^\nu, \overline{m}^\nu)$ and $\overline{c}(\nu) = \max(\underline{m}^\nu, \overline{m}^\nu)$. Moreover,

$$\mathcal{D}((BB^*)^{-\frac{\nu}{2}}) = \mathcal{R}((BB^*)^{\frac{\nu}{2}}) = \mathcal{Y}_{\nu(a+s)},$$

where in case $\nu > 0$, $(BB^*)^{\frac{\nu}{2}}$ is identified with its extension to $\mathcal{Y}$.

Proof. The result follows by the inequality of Heinz, see, e.g., [5, Corollary 8.22] for details. □

It is well-known that in order to obtain convergence rates, the true solution $x^\dagger$ has to satisfy some smoothness assumptions, cf. [23]. As it will become clear from our analysis the appropriate source condition for $\mathcal{Y}$-scale regularization has to be stated in terms of the operator B (see also Remark 2).

Assumption 2 (Source condition) *Let $s > -a$. There exists a real number $u > 0$ and $w \in \mathcal{X}$ such that*

$$x^\dagger = (B^*B)^{\frac{u}{2(a+s)}}w. \tag{23}$$

We use here the notation for standard regularization in Hilbert scales, cf. [18, 25].

The total error $\|x_\alpha^\delta - x^\dagger\|$ can be decomposed into the two main components, namely the approximation error $\|x_\alpha - x^\dagger\|$ and the propagated data error $\|x_\alpha^\delta - x_\alpha\|$, where x_α denotes the approximate solution defined by (20) with y^δ replaced by the correct data $y = Tx^\dagger$. We will now derive estimates for the two error components in terms of α and δ:

Theorem 1 *Let Assumption 1 hold, y^δ satisfy* (2), *and x_α^δ, x_α be defined by the $\mathcal{Y}$-scale method* (20) *with some $s \geq -a/2$. Furthermore, let g_α satisfy* (4), (5) *for some $\mu_0 > 0$, and let $x^\dagger$ satisfy Assumption 2 for $u \leq 2(a+s)\mu_0$. Then the following a-priori estimates hold:*

$$\|x_\alpha - x^\dagger\| \;\leq\; C_1\,\alpha^{\frac{u}{2(a+s)}}\,\|w\|, \tag{24}$$

$$\|x_\alpha^\delta - x_\alpha\| \;\leq\; C_2\,\alpha^{-\frac{a}{2(a+s)}}\,\delta, \tag{25}$$

where C_1, C_2 denote generic constants dependending only on c_μ, c_*, $\underline{m}$ and $\overline{m}$. As a consequence

$$\|x^\dagger - x_{\alpha_*}^\delta\| = O(\delta^{\frac{u}{a+u}}) \qquad \textit{for} \quad \alpha_* \sim \delta^{\frac{2(a+s)}{a+u}}. \tag{26}$$

Proof. We start by estimating the approximation error by using (5) and (23):

$$\begin{aligned}\|x_\alpha^\delta - x^\dagger\| &= \|g_\alpha(B^*B)B^*L^{-s}y - x^\dagger\|\\ &= \|r_\alpha(B^*B)(B^*B)^{\frac{u}{2(a+s)}}w\|\\ &\leq c_{\frac{u}{2(a+s)}}\alpha^{\frac{u}{2(a+s)}}\,\|w\|.\end{aligned}$$

Now consider the propagated data error: By Proposition 2 we have $\mathcal{R}(L^{-s}) = \mathcal{Y}_s = \mathcal{R}((BB^*)^{\frac{s}{2(a+s)}})$, hence

$$x_\alpha^\delta - x_\alpha = g_\alpha(B^*B)B^*L^{-s}(y^\delta - y) = g_\alpha(B^*B)B^*(BB^*)^{\frac{s}{2(a+s)}}v,$$

for some $v_\delta \in \mathcal{Y}$ with $\|v_\delta\| \le c_1 \|y - y^\delta\|$. Thus, we can further estimate

$$\begin{aligned} \|x_\alpha^\delta - x_\alpha\| &= \|g_\alpha(B^*B)B^*(BB^*)^{\frac{s}{2(a+s)}} v_\delta\| \\ &\le c_2 \|g_\alpha(B^*B)(B^*B)^{\frac{a+2s}{2(a+s)}}\| \, \|y - y^\delta\| \\ &\le c_3 \, \alpha^{-\frac{a}{2(a+s)}} \, \delta, \end{aligned}$$

where for the last estimate we utilized that

$$\sup_{\lambda \in [0,1]} |g_\alpha(\lambda)\lambda^\mu| \le C_\mu \, \alpha^{\mu-1}, \qquad 0 \le \mu \le 1,$$

which follows readily from (4) and (5). (26) is implied by (24) and (25). □

Remark 3 The estimates of Theorem 1 coincide with the convergence results for regularization in Hilbert scales derived in [5, 18]. However, as indicated by Example 1, $\mathcal{Y}-$scale regularization (13) is applicable to a different class of problems. In contrast to standard regularization in Hilbert scales, the source condition (23) cannot be interpreted in terms of the spaces $\mathcal{Y}_r$ directly. However, the following interpretation is possible: Define z by $x = B^*z$. Then problem (1) is equivalent to

$$BB^*z = L^{-s}y^\delta, \qquad x = B^*z.$$

The source condition (23) can be restated for z by noting

$$B^*z^\dagger = x^\dagger = (B^*B)^{\frac{u}{2(a+s)}} w = B^*(BB^*)^{\frac{u-a-s}{2(a+s)}} v$$

hence $z^\dagger \in \mathcal{R}((BB^*)^{\frac{u-a-s}{2(a+s)}}) = \mathcal{Y}_{u-a-s}$. The last identity is valid only for $u \le a + s$. A similar restriction also holds for standard regularization in Hilbert scales, where the the source condition $x^\dagger \in \mathcal{X}_u$ guarantees optimal convergence rates only for $u \le a + 2s$, cf. [5, Remark 8.24].

Since in practice the smoothness u of a solution $x^\dagger$ is not known a-priori, an optimal parameter choice (26) is not possible in general. Therefor, we next consider a parameter selection criterion based on the discrepancy principle, cf. [5, 17], i.e.,

$$\alpha_* := \sup\{\alpha : \|Tx_\alpha^\delta - y^\delta\| \le \tau\delta\}. \tag{27}$$

The advantage of such a criterion is that it requires no a-priori information about the smoothness of $x^\dagger$. Moreover, the residual $r_n^\delta := Tx_n^\delta - y^\delta$ has usually to be calculated anyway, in particular for iterative regularzation methods.

Theorem 2 *Let the Assumptions of Theorem 1 hold and let $\alpha_* = \alpha_*(\delta)$ be determined by the discrepancy principle* (27)*. Then the rates* (26) *hold for $0 < u \le 2(a+s)\mu_0 - a$.*

Proof. We have to estimate the residual $\|Tx_\alpha^\delta - y^\delta\|$:

$$\begin{aligned} \|Tx_\alpha^\delta - y^\delta\| &= \|L^s(Bg_\alpha(B^*B)B^* - I)L^{-s}y^\delta\| \\ &\le c_1 \|r_\alpha(BB^*)B(B^*B)^{\frac{u-s}{2(a+s)}}\| \, \|w\| + c_2 \, \delta \\ &\le c_{\frac{u+a}{2(a+s)}} \, \alpha^{\frac{u+a}{2(a+s)}} \, \|w\| + c_2 \, \delta, \end{aligned}$$

where we used (5) and $u \le 2(a+s)\mu_0 - a$. Hence,

$$\alpha_* \ge C \, \delta^{\frac{2(a+s)}{a+u}} \tag{28}$$

as long as $\tau > c_2$, which yields a sufficient bound for the propagated data error by (25). It remains to consider the approximation error. First note that we have

$$\|T(x_{\alpha_*} - x^\dagger)\| \le \|Tx_{\alpha_*}^\delta - y^\delta\| + \|T(x_{\alpha_*} - x_{\alpha_*}^\delta)\| + \delta \le C \, \delta.$$

It then follows by the interpolation inequality and (23) that

$$\|x_{\alpha_*} - x^\dagger\| \le c_3 \|T(x_{\alpha_*} - x^\dagger)\|^{\frac{u}{a+u}} \|w\|^{\frac{a}{a+u}} \le c_4 \delta^{\frac{u}{a+u}} \|w\|^{\frac{a}{a+u}},$$

which completes the proof. □

Remark 4 For $s = 0$ the above results coincide with standard regularization theory by noting that $u = 2a\mu$. In Theorem 2 the restriction $u \le 2(a+s)\mu_0 - a$ then amounts to $\mu \le \mu_0 - 1/2$, which is the well-known saturation for the discrepancy principle. As for standard regularization in Hilbert scales, saturation effects of the applied method can be overcome by choosing s large enough.

4 $\mathcal{Y}$-scale regularization under relaxed assumptions - preconditioning

Probably the most severe drawback of $\mathcal{Y}$-scale resgualrization as presented in the previous section is the restrictive assumptions (19) (corresponding to (9) for standard regularization in Hilbert scales). As we will show now, most of the results can still be derived if only one of the estimates in (19) is available. We will treat the case where the right inequality holds and $s \le 0$ has to be chosen in detail below; the corresponding results for the other inequality and the choice $s \ge 0$ are indicated in remarks.

As can be seen from (26), the (optimal) regularization parameter α_* can be increased by choosing $s < 0$. For an iterative regularization method $g_\alpha = g_n$, this means that the corresponding stopping index decreases by setting $s < 0$. Note that, e.g., $k_* \sim 1/\alpha_*$ for Landweber iteration. This motivates to call the operator L^{-2s} in (13) a preconditioner. In order to stay with the notation of the previous section, we continue to write g_α for general regularization methods, and mention some implications for iterative regularization methods in seperate remarks.

Preconditioning of iterative regularization methods in Hilbert scales has been investigated in [3, 4] under the standard assumption (9) and a relaxed version. Here, we are using scales of spaces over $\mathcal{Y}$ and different Assumptions, namely (19) respectively (29). The derivation of the convergence results of the following sections therefor requires different reasoning and a more sophisticated a-posteriori stopping rule.

Let L and $\mathcal{Y}_s$ be defined as above. Throughout this section we require the following assumption:

Assumption 3 *T is injective and there exists a positive real constant a, such that*

$$\|Tx\|_a \le \|x\| \tag{29}$$

holds uniformly for all $x \in X$.

Unlike (19), where the injectivity of the operator T was implied by the lower estimate, it has to be assumed additionally here. Similarly as in the previous section we have the following equivalent characterization of Assumption 3:

Proposition 3 *Assumption 3 is equivalent to T being injective and*

$$\|T^*y\| \le \|y\|_{-a}, \qquad \forall y \in \mathcal{Y}_{-a}. \tag{30}$$

Proof. Let x, y, $\hat{y} \neq 0$ below. Assumption 3 can be restated as

$$\begin{aligned} 1 \quad &\ge \quad \sup_{x\in\mathcal{X}} \frac{\|L^a Tx\|}{\|x\|} = \sup_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} \frac{\langle Tx, L^a y\rangle}{\|x\|\,\|y\|} \\ &= \quad \sup_{\hat{y}\in\mathcal{Y}_{-a}} \sup_{x\in\mathcal{X}} \frac{\langle x, T^*\hat{y}\rangle}{\|x\|\,\|\hat{y}\|_{-a}} = \sup_{\hat{y}\in\mathcal{Y}_{-a}} \frac{\|T^*\hat{y}\|}{\|\hat{y}\|_{-a}}. \end{aligned}$$

The other implication follows in the same way. □

Due to the relaxed assumption (29), Proposition 2 is no longer valid in its general form. Instead, only the following weaker implications hold:

Proposition 4 *Let Assumption 3 hold, and $B = L^{-s}T$ for some $s \geq -a$ with $\|B\| \leq 1$. Then the estimates*

$$\|(BB^*)^{\frac{\nu}{2}}y\| \leq \|y\|_{-\nu(a+s)} \qquad \forall y \in \mathcal{Y}_{-\nu(a+s)} \subset \mathcal{R}((BB^*)^{-\frac{\nu}{2}}), \tag{31}$$

$$\|(BB^*)^{-\frac{\nu}{2}}y\| \geq \|y\|_{\nu(a+s)} \qquad \forall y \in \mathcal{R}((BB^*)^{\frac{\nu}{2}}) \subset \mathcal{Y}_{\nu(a+s)} \tag{32}$$

hold for $0 \leq \nu \leq 1$.

Proof. The results follow from the inequality of Heinz and standard results on Hilbert scales, see [5, Section 8.5] for details. □

Assumption 3 and Proposition 4 are sufficient to prove the following a-priori error estimates corresponding to those of Theorem 1.

Theorem 3 *Let $-a/2 \leq s \leq 0$ and Assumption 3 hold. Additionally, let g_α, $x^\dagger$, y^δ satisfy the assumptions of Theorem 1. Then* (24), (25) *and* (26) *hold.*

Proof. The estimate for the approximation error $x_\alpha^\delta - x^\dagger$ follows as in the proof of Theorem 1. Now observe that $L^{-s}(y^\delta - y) \in \mathcal{R}(L^{-s}) = \mathcal{Y}_s \subset \mathcal{R}((BB^*)^{\frac{s}{2(a+s)}})$ is implied by (31) since $s \leq 0$. Hence, also the estimate for the propagated data error follows as in Theorem 1. □

Remark 5 In case $s \geq 0$, condition (3) has to be replaced by

$$\|x\| \leq \|Tx\|_{\tilde{a}}. \tag{33}$$

Proposition 4 then holds with opposite inclusions and the analogue to Theorem 3 follows verbatim. Again, similar as in the proof of Theorem 1, there exists a $v_\delta \in \mathcal{Y}$ such that $L^{-s}(y - y^\delta) = (BB^*)^{\frac{s}{2(\tilde{a}+s)}} v_\delta$, since for $s \geq 0$ and under condition (33)

$$\mathcal{R}(L^{-s}) = \mathcal{Y}_s \subset \mathcal{R}((BB^*)^{\frac{s}{2(\tilde{a}+s)}})$$

follows from the analogue of Proposition 4 for this case.

4.1 A-posteriori parameter choice

The result of Theorem 3 is not of great use per-se, since it would require exact knowledge of the smoothness of the solution $x^\dagger$. We will therefor consider a-posteriori parameter choice rules in the sequel.

A discrepancy principle: In contrast to the results of the previous section, the discrepancy principle (27) does not provide parameters that guarantee optimal convergence rates if only Assumption 3 is used. To see this, let us estimate the residual

$$\begin{aligned} \|Tx_\alpha^\delta - y^\delta\| &= \|L^s(Bg_\alpha(B^*B)B^* - I)L^{-s}y^\delta\| \qquad (34)\\ &\leq \|L^s r_\alpha(BB^*)B(B^*B)^{\frac{u}{2(a+s)}}w\| + \|L^s g_\alpha(BB^*)BB^*L^{-s}(y^\delta - y)\| + \|y^\delta - y\| \\ &\leq c_1\alpha^{\frac{u+a+s}{2(a+s)}} + c_2\alpha^{\frac{s}{2(a+s)}}\delta, \qquad (35) \end{aligned}$$

for some positive constants c_1 and c_2. A careful inspection of the proof of Theorem 2 reveals that the second term may disturb the convergence of the residual $Tx_\alpha^\delta - y^\delta$, i.e., one cannot guarantee that $\inf\{\|Tx_\alpha^\delta - y^\delta\| : \alpha > 0\} \leq \tau\delta$ for any τ (independent of δ and y^δ), for all possible data y^δ. Hence, the discrepancy principle (27) might not yield a parameter at all. A partial result can however be found for the following modified criterion: Let α_{max} be defined by

$$\alpha_{max} := \sup\{\alpha : \|Tx_\alpha^\delta - y^\delta\| \leq \tau\delta\alpha^{\frac{s}{2(a+s)}}\}, \tag{36}$$

for some appropriate $\tau > 1$. We can estimate α_{max} from below by

$$\tau\delta\alpha_{max}^{\frac{s}{2(a+s)}} \le \|Tx_\alpha^\delta - y^\delta\| \le c_1\, \alpha^{\frac{u+a+s}{2(a+s)}} + c_2\, \alpha^{\frac{s}{2(a+s)}}\, \delta$$

for all $\alpha > \alpha_{max}$. Hence, assuming that $\tau > c_2$, we obtain with an appropriate constant $\overline{C} > 0$ that

$$\alpha_{max} \ge \overline{C}\delta^{\frac{2(a+s)}{a+u}}. \tag{37}$$

In view of Theorems 1 and 2 a parameter choice $\alpha_* = \alpha_{max}$ cannot provide optimal convergence rates in general, but at least the following sub-optimal result holds:

Proposition 5 *Let the iterative method* (13) *be stopped at* $\alpha_* = \alpha_{max}$. *Then the following (sub-optimal) rate holds:*

$$\|x_\alpha^\delta - x^\dagger\| = O(\delta^{\frac{u+\frac{us}{2(a+s)(a+u)}}{a+u}}). \tag{38}$$

Proof. In the same manner as (34) one obtains that

$$\|Tx_\alpha - y\| \le \alpha^{\frac{s}{2(a+s)}}\delta.$$

The result the follows similar as in the proof of Theorem 2. □

Remark 6 For $s = 0$ the rate (38) coincides with the optimal rate $O(\delta^{\frac{u}{a+u}})$. If $s < 0$, the convergence rate observed for this parameter choice is not optimal, e.g., for $s = -a/2$ equation (38) yields

$$\|x_\alpha^\delta - x^\dagger\| = O(\delta^{\frac{u}{2(a+u)}}),$$

which is only half of the optimal rate (26).

We will show next, how optimal rates can be restored:

A Lepskij principle: In view of Theorem 1, α_{max} is a reasonable upper bound for appropriate parameters yielding optimal convergence rates. On the other hand, according to Theorem 1 optimal rates are always obtained for the choice

$$\alpha_* = \alpha_{opt} := \overline{C}\delta^{\frac{2(a+s)}{a+u}} \ge \overline{C}\delta^{\frac{2(a+s)}{a}} =: \alpha_{min}, \tag{39}$$

with $\overline{C}$ as in (37). Hence, α_{min} is a in any case a reasonable lower bound for interesting regularization parameters. A specific parameter yielding optimal convergence rates can be found in the range $[\alpha_{min}, \alpha_{max}]$ by means of the Lepskij principle, cf. [14, 16]; we use here a formulation similar to the one presented in [1]:

Let α_{min}, α_{max} be determined according to (36), (39), and denote by $\alpha_n := q^n \cdot \alpha_{max}$ for some $0 < q < 1$. We define the set $\mathcal{M}$ (of admissible parameters) by

$$\mathcal{M} := \left\{ \alpha_n \in [\alpha_{min}, \alpha_{max}] : \begin{array}{l} \|x_{\alpha_n}^\delta - x_{\alpha_m}^\delta\| \le 4C\alpha_n^{-\frac{a}{2(a+s)}}\delta \\ \text{for all } m = n+1, \dots, N \end{array} \right\}, \tag{40}$$

where N is the largest integer such that $\alpha_N > \alpha_{min}$, and C is a generic constant such that

$$\max\{\|x_\alpha^\delta - x_\alpha\|, \|x_\alpha - x^\dagger\|\} \le C\alpha^{-\frac{a}{2(a+s)}}\delta \qquad \text{for all} \quad \alpha \le \alpha_{opt}. \tag{41}$$

Note that by the error estimates (24), (25), such a constant can be chosen depending on (a bound on) $\|w\|$, but without knowledge of u or α_{opt}. As appropriate parameter, we then choose

$$\alpha_* := \begin{cases} \arg\max \mathcal{M} & \text{if } \mathcal{M} \ne \{\}, \\ N. \end{cases} \tag{42}$$

Theorem 4 *Let the assumptions of Theorem 3 be satisfied and let* α_* *be defined by* (42)*. Then the rate*

$$\|x_{\alpha_*} - x^\dagger\| = O(\delta^{\frac{u}{a+u}}) \tag{43}$$

holds.

Proof. By definition of α_{opt} we have $\alpha_{opt} \in [\alpha_{min}, \alpha_{max}]$. Additionally, it follows from the error estimates (24), (25), and (41) that $\alpha_* \geq \alpha_{opt}$. Hence, we obtain togehter with (40) that

$$\|x^\delta_{\alpha_*} - x^\dagger\| \leq \|x^\delta_{\alpha_{opt}} - x^\dagger\| + \|x^\delta_{\alpha_*} - x^\delta_{\alpha_{opt}}\| = O(\delta^{\frac{u}{a+u)}}).$$

□

Remark 7 A disadvantage of the Lepskij principle is that its realization requires a higher computational effort than the standard discrepance principle (27). Note that assembling of the set (40) requires to generate solution for a wide range of regularization parameters, which ensures that a parameter providing optimal convergence rates is in the set $\mathcal{M}$. Additionally, the solution $x^\delta_{\alpha_n}$ have to be stored. By using the upper bound α_{max} on admissible parameters, the number of solutions $x^\delta_{\alpha_n}$ which have to be stored can be reduced significantly; note, that usually $\alpha_{max} \sim 1$ is chosen, cf. [1, 16].

The $\mathcal{Y}-$scale approach allows to choose the smallest regularization parameter α_{min} relatively large, e.g., $\alpha_{min} \sim \delta$ for $s = -a/2$, whereas one would need $\alpha_{min} \sim \delta^2$ for standard regularization methods (3), cf. [1, 16].

As already mentioned, setting $\alpha_* = \alpha_{max}$, which can be determined by the modified discrepancy principle (36), will only guarantee the sub-optimal rates (38). Even if the standard discrepancy principle (27) provides a parameter, i.e., if there exists an α such that $\|Tx^\delta_\alpha - y^\delta\| \leq \tau\delta$, such a choice will in general only yield sub-optimal rates similar to (38).

Remark 8 The results derived for general regularization methods (3) hold with obvious modifications also for iterative regularization methods. Under the strong condition (19), and with $s = -a/2$, Theorem 2 yields optimal convergence rates when the method is stopped at iteration N_* defined by

$$\|Tx^\delta_{N_*} - y^\delta\| \leq \tau\delta < \|Tx^\delta_n\|$$

As can be seen from (26) respectively (17), the number of iterations needed to get the optimal rates (16) can be reduced to about the square root by proper preconditioning ($s = -a/2$).

Under the weaker assumption (29), and with $s = -a/2$, one has to perform $N_{max} := 1/\alpha_{min} \sim \delta^{-1}$ iterations in order to apply the Lepskij principle (42). Note that if $x^\dagger$ is not very smooth, i.e., if μ in (8) is small, which is the case we are interested in, then one would expect $N_* = O(\delta^{\frac{2}{2\mu+1}}) \sim O(\delta^{-2})$ iterations for (the standard version of) Landweber iteration. Hence also in this case, the number of iterations can be reduced to about the square root by choosing $s = -a/2$.

Remark 9 If instead of Assumption 3 condition (33) holds and $s \geq 0$, then the results concerning the discrepancy principle are slightly different to those presented here. E.g., the standard discrepancy principle (27) can be used to determine α_{max}. However, as in the case $s \leq 0$, a choice $\alpha_* = \alpha_{max}$ does not yield optimal convergence rates in general. The result for the Lepskij principle follows as above with obvoius modification. For $s \geq 0$ and if iterative regularization methods are applied, the number of iterations will be increased by application of L^{-2s}, which in case $s \geq 0$ is a smoothing operator. Such choice seems to be disadvantageous from a numerical point of view, at least for iterative methods.

We only mention that if a condition (33) holds for some $\tilde{a} \geq a$ additionally to (29) (cf. Example 3), then some of the previous results can be strengthened. In particular, for $\tilde{a} = a$, the results of Section 3 apply.

5 Examples and numerical tests

In this section we verify Assumptions 1 respectively 3 for several examples, and present the results of some numerical tests illustrating the theoretical results of the previous section.

Example 2 (Radon transform) Let $f \in C_0^\infty(\Omega^n)$, for some compact domain $\Omega^n \subset \mathbb{R}^n$, $n \geq 2$. The Radon transform Rf of f is defined by

$$Rf(\omega, s) = \int_{x\cdot\omega=s} f(x)dx.$$

The Radon transform is important in many applications, e.g., in computerized tomography, and has been studied extensively in the literature. The following very general stability estimate holds, cf. [19, 24]:

$$c(\alpha, n)\|f\|_{H_0^\alpha(\Omega^n)} \leq \|Rf\|_{H^{\alpha+(n-1)/2}(Z)} \leq C(\alpha, n)\|f\|_{H_0^\alpha(\Omega^n)} \tag{44}$$

where Z denotes the unit cylinder in $\mathbb{R}^n$, and

$$\|g\|_{H^{\alpha+(n-1)/2}(Z)} := \int_{S^{n-1}} \int_{\mathbb{R}} (1+\sigma^2)^\alpha |\hat{g}(\theta, \sigma)|^2 d\sigma\, d\theta.$$

Here, the Fourier transform is only taken with respect to the second variable. Due to (44) the result of Section 3 are applicable with $Ly(\theta, s) = -\partial_{ss}y(\theta, s) + y(\theta, s)$, and fractional powers of the operator L can be realized by Fourier transform.

Example 3 (An integral equation allowing a lower bound) Consider the Fredholm integral equation $Tx = y$ with operator $T : L_2(0,1) \to L_2(0,1)$ defined by

$$Tx(s) = \int_0^1 k(s,t)dt \quad \text{with} \quad k(s,t) = \sqrt{t}\begin{cases} s(1-t)\,, & 0 \leq s < t \leq 1\,, \\ t(1-s)\,, & 0 \leq t \leq s \leq 1\,. \end{cases}$$

Since $k(s,t) \in L_2([0,1]^2)$, it follows that T is compact and $Tx = y$ is ill-posed. With

$$(Tx)(s) = (1-s)\int_0^s t^{3/2}x(t)dt + s\int_s^1 t^{1/2}(1-t)x(t)dt$$

we get $(Tx)(0) = (Tx)(1) = 0$. By twice differentiation we obtain that

$$(Tx)''(s) = -s^{1/2}x(s)\,,$$

which yields

$$\mathcal{R}(T) = \{y \in H^2(0,1) \cap H_0^1(0,1) : s^{-1/2}y''(s) \in L_2(0,1)\}\,.$$

As operator L generating a $\mathcal{Y}$-scale, we choose

$$L^s y := \sum_{n=1}^\infty (n\pi)^s \langle x, x_n \rangle y_n\,, \qquad \text{with} \quad y_n := \sqrt{2}\sin(n\pi\cdot)\,, \tag{45}$$

i.e., $L^2x = -x''$ and $\mathcal{Y}_2 = H^2(0,1) \cap H_0^1(0,1)$. This choice yields $\mathcal{R}(T) \subsetneq \mathcal{Y}_2$ and $\mathcal{R}(T) \supset \mathcal{Y}_{2.5} = \{y \in H^{2.5}(0,1) \cap H_0^1(0,1) : \rho^{-1/2}y'' \in L_2(0,1)\}$ with $\rho(t) = t(1-t)$. By Theorem 11.7 in [15], we get

$$\underline{m}\|x\| \leq \|Tx\|_{2.5} \quad \text{and} \quad \|Tx\|_2 \leq \overline{m}\|x\|$$

for some positive constants $\underline{m}$, $\overline{m}$, thus we can apply the results of Section 4.

Next we apply Tikhonov regularization, Landweber iteration, and the ν-methods by Brakhage [2, 8] for the numerical solution of $Tx = y$. As parameter choice we use the Lepskij principle (42). We shortly describe how an implementation of this stopping rule is implemented for Landweber iteration: As soon as the modified discrepancy principle (36) is satisfied, we set $\mathcal{M}_x := \{x_{k_{min}}^\delta\}$, where k_{min}

plays the role of α_{max} in (36). At each iteration $k_n := 2^n \cdot k_{min}$, $n \in \mathbb{N}$, we add the element $x^\delta_{k_n}$ to the set $\mathcal{M}_x$ and check if the condition

$$\|x^\delta_{k_n} - x^\delta_{k_m}\| \leq 4Ck_n^{\frac{a}{2(a+s)}}\delta \tag{46}$$

for all $x^\delta_{k_n}$, $x^\delta_{k_m}$ in the set $\mathcal{M}_x$ with $m > n$. If (46) does not hold for some n, we eliminate all elements $x^\delta_{k_m}$ with $m \leq n$ and repeat until the condition holds for all elements in $\mathcal{M}_x$ or the set $\mathcal{M}_x$ consist of only one element. This procedure is continued until the final iteration $k_{max} = C\delta^{-1}$ is reached.

For a concrete numerical test, we set $x^\dagger(s) = \text{sign}(1-2s)$. With $a = 2$, $s = -1$ and the above choice of a Hilbert scale, we have $x^\dagger \in \mathcal{R}(B^*B)^{\frac{u}{2(a+s)}} = \{x \in \mathcal{Y}_2 : \frac{1}{\sqrt{s}}(\frac{x}{\sqrt{s}})'' \in L_2(0,1)\}$ for $0 \leq u < 1$. The correct data $y = Tx^\dagger$ are perturbed by uniformly distributed random noise such that $\|y - y^\delta\| = \delta$ for several values of delta. In a first test, we set $\delta = 0.002$, $\tau = 1.1$ and compare the two error components (approximation error and propagated data error) for Landweber iteration with preconditioning ($s = -a/2$). According to Figure 1, the optimal stopping index is 230, whereas the standard discrep-

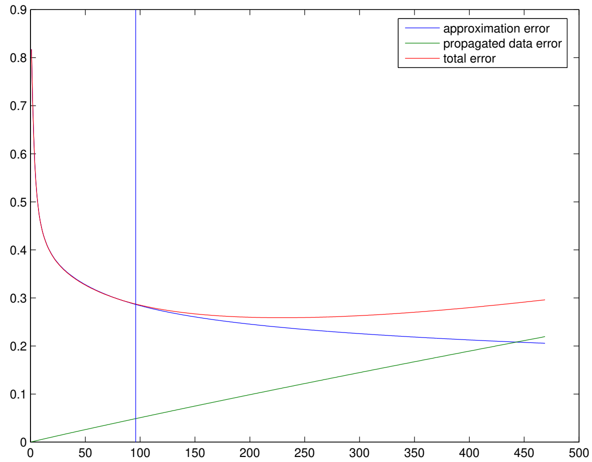

Figure 1: Approximation, propagated data and total error, Example 3 with $\delta = 0.002$, $\tau = 1.1$ and iterates obtained by the preconditioned Landweber iteration. The vertical line denotes the actual stopping index n_* determined by the Lepskij rule.

ancy principle stoppes only after 469 iterations, cf. (34) and Remark 7. The Lepskij principle (42), with $C = 0.25$ yields $N = 96$. The corresponding errors $e_n = \|x^\delta_n - x^\dagger\|$ are $e_{230} = 0.26$, $e_{469} = 0.30$, and $e_{96} = 0.27$.

In order to illustrate the effect of preconditioning, we list the iteration numbers and errors for several values of δ obtained for Landweber iteration and the ν-methods by Brakhage [2, 8] with and without preconditioning in Table 1. Note that the numerical effort for on iteration step is about the same for all methods and consists of application of T and T^*. The preconditioned iteration requires further application of a differential operator, which is cheap in comparison to application of the integral operator T.

The corresponding convergence rates are $e_n := \|x^\delta_n - x^\dagger\| = O(\delta^{0.18})$ for all iterations. The iteration numbers behave like $n_*^{(lw)} \sim \delta^{-1.7}$, $n_*^{(nu)} \sim \delta^{-0.87}$ for the standard Landweber iteration and the ν-method. In both cases the stopping indices can be reduced to the square root by preconditioning, i.e., $n_*^{(hslw)} \sim \delta^{-0.90}$ and $n_*^{(hsnu)} \sim \delta^{-0.48}$.

The following two examples shall serve as models for inverse problems governed by partial differential equations. Note that in 2-D, or 3-D, an assembly of the full operator is usually impossible for

δ	$n_*^{(lw)}$	$e_n^{(lw)}$	$n_*^{(hslw)}$	$e_n^{(hslw)}$	$n_*^{(nu)}$	$e_n^{(nu)}$	$n_*^{(hsnu)}$	$e_n^{(hsnu)}$
0.02	46	0.4299	20	0.4244	13	0.4099	9	0.4078
0.01	177	0.3920	38	0.3655	21	0.3655	12	0.3634
0.005	819	0.3307	81	0.3312	43	0.3333	18	0.3300
0.002	3488	0.2690	166	0.2704	112	0.2587	29	0.2589

Table 1: Iteration numbers and relative errors for Landweber iteration (lw), the ν-Method with $\nu = 2$ and the preconditioned versions, Example 3.

reasonable discretization levels, and even the application of the operator T, which involves the solution of the governing equation may be rather expensive. Thus such problems will usually be solved by some iterative method, and the number of iterations should be kept as small as possible. As we will illustrate, this can effectively be achieved by preconditioning in $\mathcal{Y}$-scales:

Example 4 (Source reconstruction) Let $\Omega \subset \mathbb{R}^n$, $n = 2, 3$ be a bounded convex domain and let $T : L_2(\Omega) \to L_2(\Omega)$ be defined by $Tf = u$ where u satisfies

$$-q\Delta u + cu = f, \qquad u\big|_{\partial\Omega} = 0,$$

and c, $q \in L_2(\Omega)$ is a known function bounded from below by some constants $\underline{c} \geq 0$, $\underline{q} > 0$. We consider the $\mathcal{Y}$-scale induced by

$$L^2 : \mathcal{D}(L^2) \subset L_2(\Omega) \to L_2(\Omega) \quad \text{with} \quad L^2 u = -\Delta u,$$

and domain $\mathcal{D}(L^2) = H^2(\Omega) \cap H_0^1(\Omega) =: \mathcal{Y}_2$. It follows from standard regularity results for elliptic equations that the solution u satisfies $u \in \mathcal{Y}_2$, i.e., $\mathcal{R}(T) \subset \mathcal{Y}_2$. Moreover, if q is additionally bounded from above, than also the opposite inclusion holds, i.e., Assumption 1 is valid with $a = 2$.

For a numerical test, we consider $\Omega = (0, 1)^2$, $q = 2 + \text{sign}(1 - 2x)$, $c = 1$. Let the true solution be defined by $f^\dagger = \text{sign}(1 - 2y)$ and choose $s = -a/2 = -1$. In view of Theorem 2, we can apply the standard discrepancy principle for this example. The results of the numerical tests are listed in Table 2.

δ	$n_*^{(lw)}$	$e_n^{(lw)}$	$n_*^{(hslw)}$	$e_n^{(hslw)}$	$n_*^{(nu)}$	$e_n^{(nu)}$	$n_*^{(hsnu)}$	$e_n^{(hsnu)}$
0.02	424	0.451	34	0.587	55	0.456	14	0.538
0.01	1639	0.374	52	0.445	116	0.375	18	0.425
0.005	4723	0.307	91	0.359	192	0.310	26	0.355
0.0025	12794	0.251	161	0.282	321	0.253	33	0.275

Table 2: Iteration numbers and relative errors for Landweber iteration (lw), the ν-Method with $\nu = 2$ and the preconditioned versions, Example 4.

Example 5 (Exponentially ill-posed problems) As a final example we consider the backwards heat equation as a model problem for exponentially ill-posed problems: Let $T : L_2(0, 1) \to L_2(0, 1)$ defined by $(Tg)(x) = y(x) := u(x, \bar{t})$ with some $\bar{t} > 0$ and

$$-u_t + qu_{xx} = 0\,, \qquad u(0, t) = u(1, t) = 0\,, \qquad u(x, 0) = g\,.$$

The solution of this equation has the Fourier expansion

$$u(x, \bar{t}; g) = \sum_{n=1}^{\infty} \exp(-q\bar{t}\pi^2 n^2)\langle\, g, y_n \,\rangle y_n\,,$$

with the basis functions y_n as in Example 3. Let L^s be defined by (45). Then we have

$$\|Tg\|_r \leq c(r)\,\|g\| \qquad \text{for all} \quad r \in \mathbb{R},$$

thus Assumption 3 can be satisfied for arbitrary a (up to scaling by a constant).

For a numerical test, we let $q = 0.01$, set $s = -2$ for preconditioning, and try to reconstruct the initial condition

$$g^\dagger = 2x - \mathrm{sign}(2x-1)) - 1$$

from measurements of u at time $\bar{t} = 1$.

δ	$n_*^{(lw)}$	$e_n^{(lw)}$	$n_*^{(hslw)}$	$e_n^{(hslw)}$	$n_*^{(nu)}$	$e_n^{(nu)}$	$n_*^{(hsnu)}$	$e_n^{(hsnu)}$
0.02	13	0.501	9	0.498	6	0.492	5	0.486
0.01	21	0.488	12	0.486	9	0.482	7	0.476
0.005	37	0.482	15	0.482	15	0.473	8	0.474
0.0025	402	0.435	58	0.433	76	0.434	21	0.432

Table 3: Iteration numbers and relative errors for Landweber iteration (lw), the ν-Method with $\nu = 2$ and the preconditioned versions, Example 5.

Since the problem under consideration is exponentially ill-posed, only logarithmic convergence rates can be expected, cf. [10]. The solution of all considered methods look almost identical. Note that for this problem, our theory does actually not imply smaller stopping indices of the discrepancy for the preconditioned methods. However, as can be seen from Table 1, the number of iterations is reduced significantly for finite noise levels $\delta > 0$, cf. also [4] for a discussion of this phenomenon.

References

[1] F. Bauer and T. Hohage. A Lepskij-type stopping rule for regularized Newton methods. *Inverse Problems*, 21:1975–1991, 2005.

[2] H. Brakhage. On ill-posed problems and the method of conjugate gradients. In H. W. Engl and C. W. Groetsch, editors, *Inverse and Ill-posed Problems*, pages 165–175, Boston, New York, London, 1987. Academic Press.

[3] H. Egger. Semiiterative regularization in Hilbert scales. *SIAM J. Numer. Anal.*, 2005. to appear.

[4] H. Egger and A. Neubauer. Preconditioning Landweber iteration in Hilbert scales. *Numer. Math.*, 101:643 – 662, 2005.

[5] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems.* Kluwer Academic Publishers, 1996.

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

[7] C. W. Groetsch. *Generalized Inverses of Linear Operators.* Dekker, New York, Basel, 1977.

[8] M. Hanke. Accelerated Landweber iterations for the solution of ill-posed equations. *Numer. Math.*, 60:341–373, 1991.

[9] M. Hanke, A. Neubauer, and O. Scherzer. A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. *Numer. Math.*, 72:21–37, 1995.

[10] T. Hohage. *Iterative Methods in Inverse Obstacle Scattering: RegularizationTheory of Linear and Nonlinear Exponentially Ill-Posed Problems.* PhD thesis, University of Linz, 1999.

[11] J. Janno and U. Tautenhahn. On Lavrientiev regularization for ill-posed problems in Hilbert scales. *Numer. Funct. Anal. Optim.*, 24:531–555, 2003.

[12] S. G. Krein and J. I. Petunin. Scales of Banach spaces. *Russian Math. Surveys*, 21:85–160, 1966.

[13] L. Landweber. An iteration formula for Fredholm integral equations of the first kind. *Amer. J. Math.*, 73:615–624, 1951.

[14] O. V. Lepskij. A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatn. Primen.*, 35:459–70, 1990.

[15] J. L. Lions and E. Magenes. *Non-Homogeneous Boundary Value Problems and Applications: Volume I.* Springer, Berlin - Heidelberg, 1972.

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

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

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

[19] F. Natterer. *The Mathematics of Computerized Tomography.* Teubner, Stuttgart, 1986.

[20] A. Neubauer. When do Sobolev spaces form a Hilbert scale? *Proc. Amer. Math. Soc.*, 103:557–562, 1988.

[21] A. Neubauer. Tikhonov regularization of nonlinear ill-posed problems in Hilbert scales. *Appl. Anal.*, 46:59–72, 1992.

[22] A. Neubauer. On Landweber iteration for nonlinear ill-posed problems in Hilbert scales. *Numer. Math.*, 85:309–328, 2000.

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

[24] K. T. Smith, D. C. Solomon, and S. L. Wagner. Practical and mathematical aspects of the problem of reconstructing a function from radiographs. *Bull. AMS*, 83:1227–1270, 1977.

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

[26] U. Tautenhahn. On a general regularization scheme for nonlinear ill-posed problems: II. Regularization in Hilbert scales. *Inverse Problems*, 14:1607–1616, 1998.

[27] A. N. Tikhonov. Regularization of incorrectly posed problems. *Soviet Math. Dokl.*, 4:1624–1627, 1963.