Preconditioning Landweber iteration in Hilbert Scales *

Herbert Egger † Andreas Neubauer‡

September 28, 2004

Abstract

In this paper we investigate convergence of Landweber iteration in Hilbert scales for linear and nonlinear inverse problems. As opposed to the usual application of Hilbert scales in the framework of regularization methods, we focus here on the case $s \leq 0$, which (for Tikhonov regularization) corresponds to regularization in a weaker norm. In this case, the Hilbert scale operator L^{-2s} appearing in the iteration acts as a preconditioner, which significantly reduces the number of iterations needed to match a stopping criterion. Additionally, we carry out our analysis under significantly relaxed conditions, i.e., we only require $\|Tx\| \leq \overline{m}\|x\|_{-a}$ instead of $\|Tx\| \sim \|x\|_{-a}$, which is the usual condition for regularization in Hilbert scales. The assumptions needed for our analysis are verified for several examples and numerical results are presented confirming the theoretical ones.

1 Introduction

In this paper we study inverse problems of the form

$$F(x) = y, \tag{1.1}$$

where $F : \mathcal{D}(F) \subset \mathcal{X} \to \mathcal{Y}$ is an operator between Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$. For linear equations, we will use the notation $F(x) = Tx$. We assume that Problem (1.1) is ill-posed, i.e., that solutions do not depend continuously on the data. Thus, they have to be regularized in order to obtain reasonable approximations. Additionally, we suppose that only approximate data y^δ with a known upper bound on the noise level

$$\|y^\delta - y\|_{\mathcal{Y}} \leq \delta \tag{1.2}$$

are available.

*supported by the Austrian Science Foundation (FWF) under grant SFB/F013, Subproject F1308
†Johann Radon Institute for Computational and Applied Mathematics, A-4040 Linz, Austria
‡Industrial Mathematics Institute, Johannes Kepler University, A-4040 Linz, Austria

Tikhonov regularization is certainly the most well-known regularization method for ill-posed problems (cf., e.g., [1, 3]). It has been observed that convergence of Tikhonov regularization can be accelerated, when regularizing in norms stronger than the usual norm in $\mathcal{X}$ (see, e.g., [11, 12]).

Based on spectral theory, more general regularization methods in Hilbert scales for linear inverse problems have been investigated in [1, 14].

For large-scale inverse problems, iterative regularization methods (e.g. Landweber iteration, see [5, 7]) are an attractive alternative to Tikhonov regularization. Like for Tikhonov regularization, the convergence rates for Landweber iteration for nonlinear inverse problems can be improved by performing the iteration in a Hilbert scale (see [13]). Before we formulate this modified iteration process, we shortly recall the definition of a Hilbert scale:

Let L be a densly defined unbounded selfadjoint strictly positive operator in $\mathcal{X}$. Then $(\mathcal{X}_s)_{s\in\mathcal{R}}$ denotes the Hilbert scale induced by L if $\mathcal{X}_s$ is the completion of $\bigcap_{k=0}^{\infty} D(L^k)$ with respect to the Hilbert space norm $\|x\|_s := \|L^s x\|_{\mathcal{X}}$; obviously $\|x\|_0 = \|x\|_{\mathcal{X}}$ (see [8] or [1, Section 8.4] for details).

Considering F' as an operator mapping from $\mathcal{X}_s$ into $\mathcal{Y}$, i.e., taking the adjoint with respect to these spaces, yields the following modified Landweber iteration

$$x_{k+1}^\delta = x_k^\delta + L^{-2s} F'(x_k^\delta)^*(y^\delta - F(x_k^\delta)), \qquad k \geq 0, \tag{1.3}$$

where $F'(x_k^\delta)^*$ denotes the adjoint operator with respect to the spaces $\mathcal{X}$ and $\mathcal{Y}$. In the linear case, $F'(x_k^\delta)$ is simply replaced by T. The iteration (1.3) is performed as long as $k \leq k_*(\delta, y^\delta)$, where k_* is determined from a discrepancy principle

$$\|y^\delta - F(x_{k_*}^\delta)\| < \tau\delta \leq \|y^\delta - F(x_k^\delta)\|, \quad 0 \leq k < k_*, \tag{1.4}$$

for some $\tau > 2$.

In this paper, we investigate convergence rates of Landweber iteration in Hilbert scales (1.3) for linear and nonlinear inverse problems under significantly relaxed conditions on the linear (respectively linearized) operator T ($:= F'(x^\dagger)$ in the nonlinear case), i.e., we only require $\|Tx\| \leq \overline{m}\|x\|_{-a}$ instead of $\|Tx\| \sim \|x\|_{-a}$, which is the usual condition for regularization in Hilbert scales (cf. [13]).

We draw our attention to the choice $s \leq 0$, which we always assume in the sequel. For Tikhonov regularization, this would amount to regularization in a weaker norm. We will show below that the modified iteration (1.3), which is the usual Landweber iteration for the operator F considered as operator from $\mathcal{X}_s$ into $\mathcal{Y}$, is well-defined as iteration in $\mathcal{X}$. As mentioned above, regularization in Hilbert scales was introduced to improve convergence rates, if the exact solution $x^\dagger$ of (1.1) is very smooth. However, if the solution has only poor smoothness properties, it turns out that it suffices to regularize in a weaker norm than the one in $\mathcal{X}$ to still obtain the appropriate rate. Choosing $s \leq 0$ has the advantage that the embedding operator L^{-2s} acts as a preconditioner for the smoothing operator $F'(x_k^\delta)^*$ (cf. Example 4.2, Section 4), yielding a fewer number of iterations to satisfy the stopping criterion (1.4).

In the next Section, we will extend well-known results of regularization in Hilbert scales for linear problems to our more general conditions. Section 3 then deals with convergence rates for nonlinear problems, generalizing the results of [13] to the case $s \leq 0$ and $\|F'(x^\dagger)x\| \leq \overline{m}\|x\|_{-a}$.

We conclude with some examples underlining the importance of the above generalizations and present numerical tests confirming the theoretical results.

2 Linear Problems

As mentioned in the introduction, we consider $T = F(\cdot)$ in (1.1) as operator on $\mathcal{X}_s$ with adjoint $L^{-2s}T^*$ and replace the Landweber method by the modified iteration

$$x_{k+1}^\delta = x_k^\delta + L^{-2s}T^*(y^\delta - Tx_k^\delta), \qquad k = 0, 1, 2, \dots \quad . \tag{2.1}$$

In order to prove convergence rates, we need some basic conditions:

Assumption 2.1

(L1) $Tx = y$ *has a solution* $x^\dagger$.

(L2) $\|Tx\| \leq \overline{m}\|x\|_{-a}$ *for all* $x \in \mathcal{X}$ *and some* $a > 0, \overline{m} > 0$. *Moreover, the extension of* T *to* $\mathcal{X}_{-a}$ *(again denoted by* T*) is injective.*

(L3) $B := TL^{-s}$ *is such that* $\|B\|_{\mathcal{X},\mathcal{Y}} \leq 1$, *where* $s \geq -a$.

Usually, for the analysis of regularization methods in Hilbert scales, a stronger condition than (L2) is used, namely (cf, e.g., [11, 12])

$$\|Tx\| \sim \|x\|_{-a} \qquad \text{for all} \quad x \in \mathcal{X}, \tag{2.2}$$

where the number a can be interpreted as the *degree of ill-posedness*. However, (L2) might still be satisfied, even if (2.2) does not hold. It might also be possible that an estimate from below can be given in a weaker norm, e.g.,

$$\|Tx\| \geq \underline{m}\|x\|_{-\tilde{a}} \qquad \text{for all} \quad x \in \mathcal{X} \quad \text{and some} \quad \tilde{a} \geq a,\ \underline{m} > 0, \tag{2.3}$$

see Example 4.1 below. (L3) is a simple scaling condition. In order to guarantee that the iteration (2.1) is well-defined as iteration in $\mathcal{X}$ for general $y^\delta \in \mathcal{Y}$, we additionally have to assume $s \geq -a/2$ (see Proposition 2.4).

Before we state the main result of this section, we draw some conclusions from Assumption 2.1:

Proposition 2.2 *Let Assumption 2.1 hold. Then Condition (L2) is equivalent to*

$$\mathcal{R}(T^*) \subset \mathcal{X}_a \qquad \textit{and} \qquad \|T^*w\|_a \leq \overline{m}\|w\| \quad \textit{for all} \quad w \in \mathcal{Y}. \tag{2.4}$$

Moreover for all $\nu \in [0,1]$ *it holds that* $\mathcal{D}((B^*B)^{-\frac{\nu}{2}}) = \mathcal{R}((B^*B)^{\frac{\nu}{2}}) \subset \mathcal{X}_{\nu(a+s)}$ *and*

$$\|(B^*B)^{\frac{\nu}{2}}x\| \leq \overline{m}^\nu\|x\|_{-\nu(a+s)} \qquad \textit{for all} \quad x \in \mathcal{X} \tag{2.5}$$

$$\|(B^*B)^{-\frac{\nu}{2}}x\| \geq \overline{m}^{-\nu}\|x\|_{\nu(a+s)} \qquad \textit{for all} \quad x \in \mathcal{D}((B^*B)^{-\frac{\nu}{2}}) \tag{2.6}$$

Furthermore, (2.3) *is equivalent to*

$$\begin{array}{ll} \mathcal{X}_{\tilde a} \subset \mathcal{R}(T^*) \quad \textit{and} & \|T^* w\|_{\tilde a} \geq \underline{m}\|w\| \\ & \textit{for all} \quad w \in \mathcal{N}(T^*)^\perp \; \textit{with}\; T^* w \in \mathcal{X}_{\tilde a} \end{array} \tag{2.7}$$

and if (2.3) *holds, than it follows for all* $\nu \in [0,1]$ *that* $\mathcal{X}_{\nu(\tilde a+s)} \subset \mathcal{R}((B^*B)^{\frac{\nu}{2}}) = \mathcal{D}((B^*B)^{-\frac{\nu}{2}})$ *and*

$$\|(B^*B)^{\frac{\nu}{2}} x\| \;\geq\; \underline{m}^{\nu}\,\|x\|_{-\nu(\tilde a+s)} \qquad \textit{for all} \quad x \in \mathcal{X} \tag{2.8}$$

$$\|(B^*B)^{-\frac{\nu}{2}} x\| \;\leq\; \underline{m}^{-\nu}\,\|x\|_{\nu(\tilde a+s)} \qquad \textit{for all} \quad x \in \mathcal{X}_{\nu(\tilde a+s)}. \tag{2.9}$$

Proof The proof follows the lines of Corollary 8.22 in [1] noting that the results there not only hold for $s \geq 0$ but also for $s \geq -a$.

In our convergence analysis the following shifted Hilbert scale will play an important role:

Definition 2.3 *Let* a, s *and* B *be as in Assumption 2.1. We define the shifted Hilbert scale* $\{\mathcal{X}^s_r\}_{r\in\mathcal{R}}$ *by*

$$\begin{array}{l} \mathcal{X}^s_r := \mathcal{D}((B^*B)^{\frac{s-r}{2(a+s)}} L^s) \quad \textit{equipped with the norm} \\ \|x\|_r := \|(B^*B)^{\frac{s-r}{2(a+s)}} L^s x\|_{\mathcal{X}}\,. \end{array} \tag{2.10}$$

In Proposition 3.2 below, we will summarize some properties of this shifted Hilbert scale.

We are now in the position to state the main results of this section:

Proposition 2.4 *Let Assumption 2.1 hold and* $-a/2 \leq s \leq 0$. *Additionally, assume* $x^\dagger - x_0 \in \mathcal{X}^s_u$, *i.e.,*

$$x^\dagger - x_0 = L^{-s}(B^*B)^{\frac{u-s}{2(a+s)}} w, \tag{2.11}$$

for some $w \in \mathcal{X}$ *and* $u > 0$. *Then*

$$\|x^\delta_k - x^\dagger\| \leq c(\delta k^{\frac{a}{2(a+s)}} + k^{-\frac{u}{2(a+s)}}\, \|x^\dagger - x_0\|_u). \tag{2.12}$$

Proof For the propagated data error we get the closed form expression

$$x^\delta_k - x_k = \sum_{j=0}^{k-1} L^{-s}(I - B^*B)^j B^*(y^\delta - y)$$

and similarily for the approximation error

$$x_k - x^\dagger = L^{-s}(I - B^*B)^k L^s(x_0 - x^\dagger).$$

By (2.6) and $-a/2 \leq s \leq 0$, we have $\|L^{-s}v\| \leq \overline{m}^{-\frac{s}{a+s}}\|(B^*B)^{\frac{s}{2(a+s)}} v\|$. The result is then proven in the same way as Theorem 8.23 in [1]. Note that the results of Section 8.5 in [1] apply to Landweber iteration when α is replaced by $1/k$.

As an immediate consequence we have at least convergence if the iteration is stopped at $k \leq k_*(\delta)$ with $\delta \cdot k_*^{\frac{a}{2(a+s)}} \to 0$.

Remark 2.5 Under our assumptions, the source condition $x^\dagger - x_0 \in \mathcal{X}_u^s$ is stronger than the usual source condition $x^\dagger - x_0 \in \mathcal{X}_u$. Therefore, the usual restriction $u \le a + 2s$ can be dropped (cf. [14]). If however (2.2) holds then for $0 < u \le a+2s$ the spaces $\mathcal{X}_u$ and $\mathcal{X}_u^s$ coincide (with equivalent norms). In case only the weaker estimate (2.3) is valid, one still has $\mathcal{X}_u^s \subset \mathcal{X}_{\overline{u}}$, with $\overline{u} = (u-s)\frac{\tilde{a}+s}{a+s} + s$. In particular, since $s \le 0$, an estimate (2.3) from below is only needed to interpret the source condition (2.11) in terms of the Hilbert scale $\{\mathcal{X}_s\}_{s\in\mathcal{R}}$.

In order to derive convergence rates in terms of δ, it remains to bound the number of iterations:

Theorem 2.6 *Let the assumptions of Proposition 2.4 hold. If the iteration* (2.1) *is stopped according to the* a priori *rule* $k^* \sim (\|w\|\delta^{-1})^{\frac{2(a+s)}{a+u}}$ *then*

$$\|x_k^\delta - x^\dagger\| = O(\|w\|^{\frac{a}{a+u}}\delta^{\frac{u}{a+u}}). \tag{2.13}$$

If, alternatively, the iteration is stopped according to the discrepancy principle (1.4) *then*

$$\|x_k^\delta - x^\dagger\| = O(\delta^{\frac{u}{a+u}}). \tag{2.14}$$

Proof Note that $(x^\dagger - x_0) \in \mathcal{X}_u^s$ is equivalent to $(x^\dagger - x_0) \in \mathcal{R}(\tilde{T}^*\tilde{T})^{\frac{u-s}{2(a+s)}})$, where $\tilde{T}$ denotes the extension of T to $\mathcal{X}_s$ ($s<0$). Thus by Theorem 6.5 in [1] we have $k_*(\delta, y^\delta) = O(\delta^{-\frac{2(a+s)}{a+u}})$, when (2.1) is considered as the usual Landweber iteration for the operator $\tilde{T} : \mathcal{X}_s \to \mathcal{Y}$. The rates now follow directly from Proposition 2.4.

Remark 2.7 In analogy to Theorem 8.25 in [1] it is even possible to derive $o(\cdot)$-bounds in (2.14). If (2.2) is valid, the rates are optimal, i.e., the best possible worst case error bounds under the given source condition.

Note, that the convergence rates do not depend on s, while the stopping index k_* does. This suggests to choose s as small as possible, i.e., $s = -a/2$, in which case the number of iterations is bounded by $k_* \sim \delta^{-\frac{a}{a+u}}$. If the stronger condition (2.2) holds, we have $\mathcal{X}_u^s = \mathcal{X}_u = \mathcal{R}((T^*T)^\mu)$ with $\mu = \frac{u}{2a}$ as long as $u \le \min(a, a+2s)$. Thus, we get $k_* \sim \delta^{-\frac{1}{2\mu+1}}$, which is the square-root of the number of iterations needed for the standard Landweber iteration and is of the same order than the optimal number of iterations for accelerated Landweber methods (cf. [1, Section 6]), e.g., the $\nu-$methods (see [4]).

In case $s = -a/2$ and if (2.2) holds, the backprojection operator $L^{-2s}T^*$ is not smoothing any more; to be more specific, we have

$$\|L^{-2s}T^*y\| \sim \|(T^*T)^{-\frac{1}{2}}T^*y\| \sim \|y\|_{\mathcal{Y}}.$$

This means that L^{-2s} is an optimal preconditioner for T^* and the operator M_s appearing in the preconditioned normal equation

$$M_s x := L^{-2s}T^*Tx = L^{-2s}T^*y^\delta \tag{2.15}$$

has the same smoothing properties as the operator T in the original equation $Tx = y$, while being selfadjoint when considered as operator on $\mathcal{X}_s$.

Note, that it is not possible to choose $s = -a$, in which case one would have $\|M_s x\| \sim \|x\|$. In particular, if $s < -a/2$, the iteration (2.1) is not even well-defined for general $y^\delta \in \mathcal{Y}$.

3 Nonlinear Problems

Convergence of Landweber iteration for nonlinear inverse problems has been investigated in [5]. There, convergence rates are proven under the range invariance condition

$$F'(x) = R_x F'(x^\dagger), \quad x \in \mathcal{B}_\rho(x_0), \tag{3.1}$$

with $\|R_x - I\| \le C\|x - x^\dagger\|$ for all $x \in \mathcal{B}_\rho(x_0)$. Additionally, the following source condition is assumed:

$$x^\dagger - x_0 = [F'(x^\dagger)^* F'(x^\dagger)]^\mu w, \tag{3.2}$$

with some $\mu > 0$ and $\|w\|$ small enough.

We will require similar, slightly more general conditions for our analysis below. For the solution of nonlinear inverse problems (1.1), we consider the iteration

$$x_{k+1}^\delta = x_k^\delta + L^{-2s} F'(x_k^\delta)^* (y^\delta - F(x_k^\delta)), \qquad k = 0, 1, 2, \dots \quad . \tag{3.3}$$

Similar to the conditions in Assumption 2.1 for linear problems, we require the following conditions:

Assumption 3.1

(N1) $F : \mathcal{D}(F)(\subset \mathcal{X}) \to \mathcal{Y}$ *is continuous and Fréchet-differentiable in* $\mathcal{X}$*.*

(N2) $F(x) = y$ *has a solution* $x^\dagger$*.*

(N3) $\|F'(x^\dagger)x\| \le \overline{m}\|x\|_{-a}$ *for all* $x \in \mathcal{X}$ *and some* $a > 0, \overline{m} > 0$*. Moreover, the extension of* $F'(x^\dagger)$ *to* $\mathcal{X}_{-a}$ *is injective.*

(N4) $B := F'(x^\dagger)L^{-s}$ *is such that* $\|B\|_{\mathcal{X},\mathcal{Y}} \le 1$*, where* $s \ge -a$*.*

Under this Assumption, Proposition 2.2 holds verbatim for the linearized operator $T := F'(x^\dagger)$. The next proposition summarizes basic properties of the shifted Hilbert scale defined in (2.10):

Proposition 3.2 *Let Assumption 3.1 hold and let* $(\mathcal{X}_r^s)_{r\in\mathcal{R}}$ *be defined as in* (2.10)*.*

(i) The space $\mathcal{X}_q^s$ *is continuously embedded in* $\mathcal{X}_p^s$ *for* $p < q$*, i.e., for* $x \in \mathcal{X}_q^s$

$$|\!|\!| x |\!|\!|_p \le \gamma^{p-q} |\!|\!| x |\!|\!|_q \,, \tag{3.4}$$

where γ *is such that*

$$\langle\, (B^*B)^{-\frac{1}{2(a+s)}} x, x \,\rangle \ge \gamma \|x\|^2 \quad \text{for all} \quad x \in \mathcal{D}((B^*B)^{-\frac{1}{2(a+s)}})\,.$$

(ii) The interpolation inequality holds, i.e., for all $x \in \mathcal{X}_r^s$

$$|\!|\!| x |\!|\!|_q \le |\!|\!| x |\!|\!|_p^{\frac{r-q}{r-p}} \, |\!|\!| x |\!|\!|_r^{\frac{q-p}{r-p}} \,, \qquad p < q < r\,. \tag{3.5}$$

(iii) The following estimate holds:

$$\|x\|_r \le \overline{m}^{\frac{r-s}{a+s}} |\!|\!| x |\!|\!|_r \qquad \text{for all} \quad x \in \mathcal{X}_r^s \subset \mathcal{X}_r \quad \text{if} \quad s \le r \le a + 2s. \tag{3.6}$$

In particular, we obtain

$$\|x\|_0 \le \overline{m}^{\frac{-s}{a+s}} \|x\|_0 \qquad \text{for all} \quad x \in \mathcal{X}_0^s \subset \mathcal{X}_0 \quad \text{if} \quad -a/2 \le s \le 0\,. \tag{3.7}$$

Moreover,

$$\|x\|_{-a} = \|F'(x^\dagger)x\| \qquad \text{for all} \quad x \in \mathcal{X}\,. \tag{3.8}$$

(iv) If in addition (2.3) *holds, the following estimates hold with* $p = s + \frac{r-s}{a+s}(\tilde{a}+s)$*:*

$$\|x\|_p \ge \underline{m}^{\frac{r-s}{a+s}} \|x\|_r \qquad \text{for all} \quad x \in \mathcal{X}_p \subset \mathcal{X}_r^s \quad \text{if} \quad s \le r \le a+2s. \tag{3.9}$$

Proof The proof follows from Proposition 8.19 in [1] and Proposition 2.2.

Note that in general $\mathcal{X}_{-b}^s$ is not the dual space of $\mathcal{X}_b^s$, as would be the case for a Hilbert scale. Thus the spaces $\mathcal{X}_r^s$ are no Hilbert scale in general.

For the following convergence rate analysis for nonlinear problems, we need some smoothness conditions on the solution $x^\dagger$ and additional conditions on the Fréchet-derivative of F:

Assumption 3.3

(N5) $x_0 \in \widetilde{\mathcal{B}}_\rho(x^\dagger) := \{x \in \mathcal{X} \,:\, x - x^\dagger \in \mathcal{X}_0^s \wedge \|x - x^\dagger\|_0 \le \rho\} \subset \mathcal{D}(F)$ *for some* $\rho > 0$.

(N6) $\|F'(x^\dagger) - F'(x)\|_{\mathcal{X}_{-b}^s,\mathcal{Y}} \le c\|x^\dagger - x\|_0^\beta$ *for all* $x \in \widetilde{\mathcal{B}}_\rho(x^\dagger)$ *and some* $b \in [0,a]$, $\beta \in (0,1]$, *and* $c > 0$.

(N7) $x^\dagger - x_0 \in \mathcal{X}_u^s$ *for some* $\frac{a-b}{\beta} < u \le b + 2s$, *i.e., there is an element* $v \in \mathcal{X}$ *so that*

$$L^s(x^\dagger - x_0) = (B^*B)^{\frac{u-s}{2(a+s)}} v\,. \tag{3.10}$$

Before we start our analysis we want to discuss the conditions above.

Remark 3.4 First of all we want to mention that, if (2.2) holds, then the conditions in Assumptions 3.1 and 3.3 are equivalent to the ones in Assumption 2.1 in [13].

Note that, due to (3.8), $F'(x^\dagger)$ has a continuous extension to $\mathcal{X}_{-a}^s \supset \mathcal{X}_{-a}$. Therefore, condition (N6) implies that $F'(x)$ has at least a continuous extension to $\mathcal{X}_{-b}^s \supset \mathcal{X}$ in a neighborhood of $x^\dagger$. By definition of the space $\mathcal{X}_{-b}^s$, this condition is equivalent to

$$\|(B^*B)^{-\frac{b+s}{2(a+s)}} L^{-s}(F'(x^\dagger)^* - F'(x)^*)\|_{\mathcal{Y},\mathcal{X}} \le c\|x^\dagger - x\|_0^\beta. \tag{3.11}$$

By virtue of (2.4) and Proposition 3.2 (iii), this implies that $L^{-2s}F'(x_k^\delta)^*$ maps $\mathcal{Y}$ at least into $\mathcal{X}_{b+2s}^s \subset \mathcal{X}_{b+2s}$ and hence $F'(x_k^\delta)^*$ maps $\mathcal{Y}$ at least into $\mathcal{X}_b$ while $F'(x^\dagger)^*$ maps $\mathcal{Y}$ into $\mathcal{X}_a$.

Note that, if $s = 0$, (N6) reduces to

$$\|(F'(x^\dagger)^*F'(x^\dagger))^{-\frac{b}{2a}}(F'(x^\dagger)^* - F'(x)^*)\|_{\mathcal{Y},\mathcal{X}} \le c\|x^\dagger - x\|_0^\beta\,,$$

compare to [5, (3.18)]. Moreover, if $b = a$ and $\beta = 1$, this condition is equivalent to (3.1) with $\|\mathcal{R}_x - I\|$ replaced by $\|(\mathcal{R}_x - I)Q\|$, where Q is the orthogonal projector from $\mathcal{Y}$ onto $\overline{\mathcal{R}(F'(x^\dagger))}$.

Condition (N7) is a smoothness condition for the exact solution comparable to (3.2). In case of regularization in Hilbert scales and under the usual assumption $\|Tx\| \sim \|x\|_{-a}$, this coincides with $x^\dagger - x_0 \in \mathcal{X}_u$.

If $b = a$, then $u \leq a + 2s$ is allowed, which is the usual restriction for regularization in Hilbert scales. For $s = 0$ and if (2.2) is valid, $u \leq a + 2s$ reduces to $\mu \leq 1/2$, which is the known restriction for optimal convergence of Landweber iteration for nonlinear problems under condition (3.1).

We will now state the main results of this section (cf. [7]):

Proposition 3.5 *Let Assumptions 3.1 and 3.3 hold. Additionally, let* $k_* = k_*(\delta, y^\delta)$ *be chosen according to the stopping rule* (1.4) *with* $\tau > 2$ *and let* $\|x^\dagger - x_0\|_u$ *be sufficiently small. Then*

$$\|x_k^\delta - x^\dagger\|_r \leq \tfrac{4(\tau-1)}{\tau-2} \|x^\dagger - x_0\|_u (k+1)^{-\frac{u-r}{2(a+s)}} \tag{3.12}$$

for $-a \leq r \leq 0$ *and*

$$\|y^\delta - F(x_k^\delta)\| \leq \tfrac{2\tau^2}{\tau-2} \|x^\dagger - x_0\|_u (k+1)^{-\frac{a+u}{2(a+s)}} \tag{3.13}$$

for all $0 \leq k < k_*$. *Moreover, for* $\delta > 0$,

$$k_* \leq \left(\tfrac{2\tau}{\tau-2} \|x^\dagger - x_0\|_u \, \delta^{-1} \right)^{\frac{2(a+s)}{a+u}} \tag{3.14}$$

In the case of exact data ($\delta = 0$), *the estimates above hold for all* $k \geq 0$.

The proof of these statements is given in the appendix. For a slightly more general statement see [7].

Combining the results of Proposition 3.5 and (3.7) yields the following

Theorem 3.6 *Under the assumptions of Proposition 3.5 the following estimate holds for* $-a \leq r \leq 0$ *and some positive constant* c_r*:*

$$\|x_{k_*}^\delta - x^\dagger\|_r \leq c_r \|x^\dagger - x_0\|_u^{\frac{a+r}{a+u}} \delta^{\frac{u-r}{a+u}} . \tag{3.15}$$

In particular, if $s \leq 0$,

$$\|x_{k_*}^\delta - x^\dagger\|_0 = O(\delta^{\frac{u}{a+u}}). \tag{3.16}$$

4 Examples and numerical tests

In this section we give some examples where the results of Sections 2 and 3 can be applied and give sufficient conditions for the validity of Assumption 2.1 respectively Assumptions 3.1 and 3.3. For the numerical tests of Example 4.2-4.4, we chose a very fine discretization by standard piecewise linear finite elements. In order to ensure that discretization effects have no significant influence, we compared the results for different discretization levels.

We start with two examples concerning linear problems:

Example 4.1 The first problem under consideration is the solution of a linear Fredholm integral equation of the first kind. Let $T : L^2[0,1] \to L^2[0,1]$ be defined by

$$(Tx)(s) = \int_0^1 k(s,t)x(t)dt, \tag{4.1}$$

with kernel

$$k(s,t) = s^{1/2} \begin{cases} s(1-t), & 0 \le s < t \le 1 \\ t(1-s), & 0 \le t \le s \le 1. \end{cases}$$

It is well-known (see, e.g., [9]) that Fredholm integral operators of the first kind with L^2 kernel are compact on $L^2[0,1]$, and thus the solution of (4.1) is ill-posed. We next give sufficient conditions for Assumption 2.1 to hold: Let us assume that $y = Tx^\dagger$ (L1); we show (L2) for an appropriate choice of a Hilbert scale: With

$$(T^*y)(t) = (1-t)\int_0^t s^{3/2}y(s)ds + t\int_t^1 s^{1/2}(1-s)y(s)ds$$

we get $(T^*y)(0) = (T^*y)(1) = 0$. Differentiation yields

$$(T^*y)'(t) = -\int_0^t s^{3/2}y(s)ds + \int_t^1 s^{1/2}(1-s)y(s)ds$$

and

$$(T^*y)''(t) = -t^{1/2}y(t).$$

Thus, we have $\mathcal{R}(T^*) \subset H^2 \cap H_0^1$. More specifically, it is easy to see that

$$\mathcal{R}(T^*) = \{w \in H^2[0,1] \cap H_0^1[0,1] : t^{-1/2}w''(t) \in L^2[0,1]\}.$$

As Hilbert scale operator, we choose

$$L^s x := \sum_{n=1}^\infty (n\pi)^s \langle x, x_n \rangle x_n, \qquad x_n := \sqrt{2}\sin(n\pi\cdot), \tag{4.2}$$

and $L^2 x = -x''$.

With this choice, we get $\mathcal{R}(T^*) \subsetneq \mathcal{X}_2 := H^2[0,1] \cap H_0^1[0,1]$ and additionally, $\mathcal{R}(T^*) \supset \mathcal{X}_{2.5} := \{w \in H^{2.5}[0,1] \cap H_0^1[0,1] : \rho^{-1/2}w'' \in L^2[0,1]\}$, with $\rho(t) = t(1-t)$. By Theorem 11.7 in [10], we have

$$\|w\|_{2.5}^2 \sim \|w''\|_{H^{1/2}}^2 + \|\rho^{-1/2}w''\|_{L^2}^2.$$

This yields

$$\begin{aligned} \|T^*y\|_{2.5}^2 &\sim \|(\cdot)^{1/2}y\|_{H^{1/2}}^2 + \|\rho^{-1/2}(\cdot)^{1/2}y\|_{L^2}^2 \\ &\ge \|(\cdot)^{1/2}y\|_{L^2}^2 + \|\rho^{-1/2}(\cdot)^{1/2}y\|_{L^2}^2 \\ &= c\left(\int_0^1 ty(t)^2dt + \int_0^1 (1-t)^{-1}y(t)^2dt\right) \ge c\|y\|_{L^2}^2 \end{aligned}$$

Summarizing, we have

$$\underline{m}\|x\|_{-2.5} \leq \|Tx\| \leq \overline{m}\|x\|_{-2}.$$

Note, that Proposition 3.2 yields $\|x^\dagger\|_u \leq c\|x^\dagger\|_{s+\frac{u-s}{2+s}(2.5+s)}$ for $s \leq u \leq 2a+s$, which gives a sufficient condition for (N7) in terms of the Hilbert scale $\{\mathcal{X}_s\}_{s\in\mathcal{R}}$.

Example 4.2 With the second example we want to demonstrate that even for exponentially ill-posed problems preconditioning with a Hilbert scale operator L^{-2s} may be numerically advantageous. Consider the backwards heat equation $Tx = y$, with $T : L^2[0,1] \to L^2[0,1]$ defined by $(Tg)(x) = y(x) := u(x,\bar{t})$ with some $\bar{t} > 0$ and

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

A solution of the heat equation has the Fourier expansion

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

with the basis functions x_n as in Example 4.1. Let L^s be defined by (4.2). Then we have

$$\|T^*y\|_r \leq c(r)\,\|y\|_0 \qquad \text{for all} \quad r \in \mathcal{R}.$$

For the numerical tests we set $s = -1$. We shortly discuss, in which sense $L^{-2s} = -(\cdot)''$ works as a preconditioner. For this purpose, let us approximate T by a truncated singular value expansion, i.e.,

$$T \sim T_N := \sum_{n=1}^{N} \exp(-q\bar{t}\pi^2 n^2)\langle\, \cdot, x_n \,\rangle x_n.$$

Assume that $g_N^\dagger \in X_N := \overline{\{x_n : n = 1,..,N\}}$ is the solution to $T_N g_N^\dagger = y_N \in X_N$. The number of iterations for an iterative solution of $T_N g = y_N$ is coupled to the condition number of the iteration operator, which is $T_N^* T_N$ for Landweber iteration, $L^{-2s}T_N^*T_N$ for the Hilbert scale version, and T_N for the method of successive approximation (also known as Richardson iteration; investigated by Fridman [2] for compact selfadjoint positive semidefinite operators), note that $T^* = T$ in our example. Using the singular value expansion of T and denoting $a = q\bar{t}\pi^2$, we get the following condition number estimates:

$$\mathrm{cond}(T_N^*T_N) \leq e^{a(2N^2-1)}, \qquad \mathrm{cond}(T_N) \leq e^{a(N^2-1)}.$$

and

$$\mathrm{cond}(L^{-2}T_N^*T_N) \leq \begin{cases} e^{2a(N^2-1))}/N^2, & 1/2 \leq a, \\ e^{2a-1}/(2a), & a < 1/2, e^{2a(N^2-1)} < N^2, \\ e^{2aN^2-1}/(2aN^2), & \text{else.} \end{cases}$$

Note that although for $N \to \infty$, the condition numbers of Landweber iteration and the Hilbert scale version have the same assymptotic behaviour, for small N the

Hilbert scale problem may even be better conditioned than the successive approximation hence yielding a fewer number of iterations. This means that for reconstructing low frequency components, the Hilbert scale iteration may be faster than successive approximation (and much faster than usual Landweber iteration). The following numerical test demonstrates this behaviour:

We consider the problem of reconstructing the initial condition

$$g_N^\dagger = \sin(2\pi x) + \sin N\pi x, \qquad N = 3, \ldots, 7.$$

The data y are calculated analytically (with $q = 0.01$, $\bar{t} = 1$) and additional noise is added, such that $\|y - y^\delta\| \leq 0.001$. The iteration is stopped according to the discrepancy principle (1.4) with $\tau = 2.1$.

N	(lw)	cond	(hs)	cond	(sa)	cond
3	17	4.85	6	4.54	6	2.20
4	63	19.31	6	2.74	12	4.39
5	292	114.14	21	10.37	25	10.68
6	1652	1000.97	87	63.13	50	31.63
7	> 5000	13027.51	357	603.62	83	114.14

Table 1. Iteration numbers and condition number estimates for Landweber iteration (lw), the Hilbert scale method (hs) and successive approximation (sa).

In the second table, we compare the numerically observed convergence rates and iteration numbers for the example

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

δ	$k_*(lw)$	$\frac{\|g_{k_*}^\delta - g^\dagger\|}{\|g^\dagger\|}$	$k_*(hs)$	$\frac{\|g_{k_*}^\delta - g^\dagger\|}{\|g^\dagger\|}$
0.008	26	0.504	4	0.479
0.004	46	0.487	5	0.474
0.002	180	0.471	9	0.464
0.001	962	0.429	34	0.429
0.0005	1784	0.418	63	0.417

Table 2. Iteration numbers and relative error for Landweber iteration (lw) and the Hilbert scale method (hs).

Although $x^\dagger \notin \mathcal{X}_u^s$ for any $u > 0$, we still observe the following convergence rates numerically: $k_* \sim \delta^{-1.66}$ and $\|g_{k_*}^\delta - g^\dagger\| \sim \delta^{0.07}$ for Landweber iteration and $k_* \sim \delta^{-1.07}$ and $\|g_{k_*}^\delta - g^\dagger\| \sim \delta^{0.05}$ for the Hilbert scale version. We emphasize that usually only logarithmic rates can be expected for exponentially ill-posed problems (see, e.g.,

[6]) under reasonable source conditions, which explains the small exponents in the numerical example.

Since the eigenvalues of T decrease like $\exp(-an^2)$, only components corresponding to $N^2 \leq \log(1/\delta)/(2a)$ can be reconstructed reliably and until the stopping criterion is reached, the operator T behaves essentially like T_N. For $\delta = 0.0005$ we have $N \sim 6$ (compare to the iteration numbers of Table 1).

Example 4.3 *A nonlinear Hammerstein integral equation.* The following example is taken from [13]. It has been observed there that convergence of Landweber iteration can be accelerated when the iteration is performed in a Hilbert scale $\mathcal{X}_s$ with $s \geq 0$ and $x^\dagger - x_0$ is sufficiently smooth. However, if $x^\dagger - x_0$ is not very smooth, optimal convergence can occur even for $s < 0$.

Let $F : H^1[0,1] \to L^2[0,1]$ be defined by

$$(F(x))(s) = \int_0^s x(t)^2 dt.$$

The adjoint of the Fréchet derivative is then given by

$$(F'(x)^* w) = 2A^{-1}\left[x(\cdot)\int_\cdot^1 w(t)dt\right],$$

where $A : \mathcal{D}(A) = \{\psi \in H^2[0,1] : \psi'(0) = \psi'(1) = 0\} \to L^2[0,1]$ is defined by $A\psi := -\psi'' + \psi$; note that A^{-1} is the adjoint of the embedding operator from $H^1[0,1]$ in $L^2[0,1]$. Assuming that $x^\dagger \geq \gamma > 0$ a.e., we get (see [13, Section 4] for details)

$$\mathcal{R}(F'(x^\dagger)^*) = \{w \in H^3[0,1] : w'(0) = w'(1) = 0, w(1) = w''(1)\}.$$

We choose the Hilbert scale induced by $L^2 x := -x'' + x$, and $\mathcal{X}_0 = H^1[0,1]$. With this choice, we have

$$\mathcal{R}(F'(x^\dagger)^*) \subset \mathcal{X}_2$$

and therefore can set $s = -1$. This yields

$$L^{-2s}F'(x)^* w = 2x(\cdot)\int_\cdot^1 w(t)dt,$$

in particular we have

$$F'(x) = R_x(x^\dagger)F'(x^\dagger),$$

with

$$\|R_x(x^\dagger) - I\| \leq C\|x - x^\dagger\|_0 \leq c\|x - x^\dagger\|_0.$$

Thus, (N6) holds wit $b = a$ and $\beta = 1$.

For the first numerical test we set $x^\dagger(t) := 3/2 - |\text{-erf}(4t-2)|$, where $\text{erf}(\cdot)$ denotes the standard error function, and $x_0 = 1/2$. To avoid inverse crimes, the data are calculated on a finer grid.

δ	$k_*(lw)$	$\frac{\|x_{k_*}^\delta - x^\dagger\|}{\|x^\dagger - x_0\|}$	$k_*(hs)$	$\frac{\|x_{k_*}^\delta - x^\dagger\|}{\|x^\dagger - x_0\|}$
0.02	479	1.895	10	1.085
0.01	1504	0.815	21	0.620
0.005	4001	0.744	37	0.402
0.0025	12333	0.530	56	0.365

Table 3. Iteration numbers and relative errors for Landweber iteration (lw) and the Hilbert scale method (hs).

The numerical test gives the following rates: $k_* \sim \delta^{-1.54}$ and $\|x_k^\delta - x_0\| \sim \delta^{0.56}$ for standard Landweber and $k_* \sim \delta^{-0.83}$ and $\|x_k^\delta - x_0\| \sim \delta^{0.53}$ for the Hilbert scale version, i.e., the number of iterations for the Hilbert scale iteration is approximately the square-root of that for standard Landweber, as predicted by the theory. Note that by setting $s = -1 = -a/2$, we have $u \le a + 2s = 0$. Thus we can actually proof convergence rates only in spaces $\mathcal{X}_r$ with $-a \le r < 0$, e.g., in $H^s[0,1]$ for $0 \le s < 1$.

Note that in this example the application of the Hilbert scale operator L^{-2} in fact makes the iteration even simpler, i.e., application of A^{-1}, which is the main numerical effort in the Landweber iteration, can be avoided while simultaneously the number of iterations is reduced.

The case, when the restriction $u \le a + 2s$ becomes active is demonstrated in the following example, which is taken from [13]: Let

$$x^\dagger(t) = t + 10^{-6}(196145 - 41286t^2 + 19775t^4 + 70t^6 + 436t^7).$$

It was shown in [13] that standard Landweber iteration yields a convergence rate of $\|x_k^\delta - x_0\| = O(\delta^{-\frac{1}{2}})$. For the Hilbert scale iteration with $s = -1$ we can not guarantee convergence in $\mathcal{X}_0 = H^1[0,1]$. However, Theorem 3.6 yields convergence rates in $\mathcal{X}_{-1} = L^2[0,1]$, which is also observed numerically.

δ	$k_*(lw)$	$\|e_*\|_0$	$\|e_*\|_{-1}$	$k_*(hs)$	$\|e_*\|_0$	$\|e_*\|_{-1}$
0.004	1	0.0356	0.0107	16	0.740	0.0586
0.002	9	0.0301	0.00883	42	0.918	0.0489
0.001	39	0.0177	0.00474	118	1.170	0.0401
0.0005	70	0.0128	0.00293	338	1.509	0.0322
0.00025	113	0.0102	0.00189	922	1.927	0.0262

Table 4. Iteration numbers and relative errors $e_* = x_{k_*}^\delta - x^\dagger$ for Landweber iteration (left) and the Hilbert scale method (right) in $\mathcal{X}_0 = H^2[0,1]$ and $\mathcal{X}_{-1} = L^2[0,1]$.

The corresponding convergence rates for Landweber iteration are: $\|x_k^\delta - x^\dagger\|_0 \sim \delta^{0.48}$ and $\|x_k^\delta - x^\dagger\|_{-1} \sim \delta^{0.66}$; for the Hilbert scale version we have $\|x_k^\delta - x^\dagger\|_0 \sim \delta^{-0.35}$ respectively $\|x_k^\delta - x^\dagger\|_{-1} \sim \delta^{0.30}$.

Note, that the Hilbert scale iterates are not bounded in $\mathcal{X}_0$. This could only be guaranteed, if

$$x^\dagger - x_0 \in \mathcal{X}_0^s = \mathcal{R}(L^{-2}T^*) \subset \{w \in H^1[0,1] : w(1) = 0\},$$

which is not the case for our choice of the Hilbert scale. Instead, we only have $x^\dagger - x_0 \in H^{1/2-\epsilon}[0,1] \subset \mathcal{X}^s_{-1/2-\epsilon}$. Thus, with Theorem 3.6 and $u < -1/2$, one can expect at most a rate of $\delta^{\frac{1}{3}}$ for the error in $\mathcal{X}_{-1} = L^2[0,1]$, which was also observed numerically.

Example 4.4 *Parameter identification.* In this example, which is taken from [5], we want to estimate the parameter c in

$$\begin{aligned} -\Delta u + cu &= f \quad \text{in} \quad \Omega, \\ u &= g \quad \text{in} \quad \partial\Omega, \end{aligned} \tag{4.3}$$

where Ω is an intervall in $\mathcal{R}^1$ or a bounded domain in $\mathcal{R}^2$ or $\mathcal{R}^3$ with smooth boundary (or a parallelepiped), $f \in L^2(\Omega)$ and $g \in H^{3/2}(\partial\Omega)$. The nonlinear operator $F : \mathcal{D}(F) \subset L^2(\Omega) \to L^2(\Omega)$ is defined as the parameter-to-solution mapping $F(c) = u(c)$, which is well-defined and Fréchet differentiable on

$$\mathcal{D}(F) := \{c \in L^2(\Omega) : \|c - \overline{c}\| \le \gamma \text{ for some } \overline{c} \ge 0 \text{ a.e.}\}$$

where $u(c)$ denotes the solution of (4.3) and $\gamma > 0$ has to be sufficiently small. With this setting, we have

$$F'(c)^* w = u(c) A(c)^{-1} w,$$

where $A(c) : H^2(\Omega) \cap H^1_0(\Omega) \to L^2(\Omega)$ is defined by $A(c)u = -\Delta u + cu$. Using $\mathcal{X}_2 := H^2(\Omega) \cap H^1_0(\Omega)$ and $L^2 = -\Delta$ we get

$$\mathcal{R}(F'(c)^*) \subset \mathcal{X}_2.$$

If $u(c^\dagger) \ge \kappa > 0$ a.e. in Ω, then for all c with $\|c - c^\dagger\| \le \rho \le \gamma$ (see [5] for details)

$$F'(c)^* = F'(c^\dagger) R_c(c^\dagger),$$

with

$$\|R_c(c^\dagger) - I\| \le C\|c - c^\dagger\|_0 \le c\|c - c^\dagger\|_0.$$

The last inequality follows by Proposition 3.2 (iii). This proofs (N6); additionally we have in this case

$$\|F'(c^\dagger)^* w\|_2 \sim \|w\|_0,$$

and thus (N7) reduces to $x^\dagger - x_0 \in \mathcal{X}_u$ ($= H_0^{2u}(\Omega)$ for $0 \le u < 3/4$).

For the numerical test we set $s = -1$. Note, that in this case we get the restriction $0 \le u \le a + 2s = 0$. Thus, the results of Section 3 do not guarantee convergence rates in $L^2(\Omega)$. However, as in the first test case of Example 3, convergence rates are observed in the numerical results. In order to ensure convergence rates also theoretically, one could alternatively set $s > -1$ and use a multilevel technique to implement L^{-2s}.

We consider $c^\dagger = \text{sign}(x - 0.5) \cdot \text{sign}(y - 0.5)$ and $\Omega = [0,1]^2$ and start with $x_0 = 1$.

δ	$k_*(lw)$	$\frac{\|x^\delta_{k_*}-x^\dagger\|}{\|x^\dagger-x_0\|}$	$k_*(hs)$	$\frac{\|x^\delta_{k_*}-x^\dagger\|}{\|x^\dagger-x_0\|}$
0.004	6	0.937767	2	0.844723
0.002	18	0.750728	4	0.697871
0.001	33	0.690766	7	0.612904
0.0005	118	0.615639	11	0.553232
0.00025	410	0.517045	21	0.472961

Table 5. Iteration numbers and relative errors for Landweber iteration (lw) and the Hilbert scale method (hs).

The corresponding rates are $k_* \sim \delta^{-1.63}$ and $\|x_k^\delta - x^\dagger\| \sim \delta^{-0.21}$ for Landweber iteration respectively $k_* \sim \delta^{-0.86}$ and $\|x_k^\delta - x^\dagger\| \sim \delta^{0.2}$ for the Hilbert scale version. Note, that we only have $x^\dagger - x_0 \in H_0^{1/2-\epsilon} = \mathcal{X}^s_{1/2-\epsilon}$, thus the optimal convergence rate $\frac{u}{a+u} = \frac{1}{5} - O(\epsilon)$ under the given source condition is realized numerically.

This numerical result and the first test of Example 3 suggest that the restriction $u \le a + 2s$, which is only needed in the nonlinear case, can possibly be relaxed in some cases (cf. [5, (3.18)]).

Appendix

For the proof of Proposition 3.5 we need the following Lemma (cf. [13])

Lemma A.1 *Let Assumptions 3.1 and 3.3 hold. Moreover, let $k_* = k_*(\delta, y^\delta)$ be chosen according to the stopping rule (1.4) with $\tau > 2$, and assume that $\|e_j^\delta\|_0 \le \rho$ and that $\|e_j^\delta\|_u \le \rho_u$ for all $0 \le j < k \le k_*$ and some $\rho_u > 0$, where $e_j^\delta := x_j^\delta - x^\dagger$. Then there is a positive constant γ_1 (independent of k and δ) such that for all $0 \le k \le k_*$ the following estimates hold:*

$$\begin{aligned} \|e_k^\delta\|_0 \quad \le \quad & \|x^\dagger - x_0\|_u (k+1)^{-\frac{u}{2(a+s)}} + \delta k^{\frac{a}{2(a+s)}} \\ & + \gamma_1 \sum_{j=0}^{k-1} (k-j)^{-\frac{a+2s}{2(a+s)}} \|e_j^\delta\|_{-a}^{\frac{b}{a}} \|e_j^\delta\|_0^{\frac{a(1+\beta)-b}{a}} \\ & + \gamma_1 \sum_{j=0}^{k-1} (k-j)^{-\frac{b+2s}{2(a+s)}} \left(\|e_j^\delta\|_{-a}^{\frac{b}{a}} \|e_j^\delta\|_0^{\frac{a(1+2\beta)-b}{a}} + \|e_j^\delta\|_{-a} \|e_j^\delta\|_0^{\beta} \right) \end{aligned} \tag{A.1}$$

$$\begin{aligned} \|e_k^\delta\|_{-a} \quad \le \quad & (k+1)^{-\frac{a+u}{2(a+s)}} \|x^\dagger - x_0\|_u + \delta \\ & + \gamma_1 \sum_{j=0}^{k-1} (k-j)^{-1} \|e_j^\delta\|_{-a}^{\frac{b}{a}} \|e_j^\delta\|_0^{\frac{a(1+\beta)-b}{a}} \\ & + \gamma_1 \sum_{j=0}^{k-1} (k-j)^{-\frac{b+a+2s}{2(a+s)}} \left(\|e_j^\delta\|_{-a}^{\frac{b}{a}} \|e_j^\delta\|_0^{\frac{a(1+2\beta)-b}{a}} + \|e_j^\delta\|_{-a} \|e_j^\delta\|_0^{\beta} \right) \end{aligned} \tag{A.2}$$

Proof From (3.3) we immediately obtain the representation

$$e_{k+1}^\delta = (I - L^{-2s}F'(x^\dagger)^*F'(x^\dagger))e_k^\delta + L^{-2s}F'(x^\dagger)^*(y^\delta - y - q_k^\delta) + L^{-2s}p_k^\delta$$

with

$$\begin{aligned} q_k^\delta &:= F(x_k^\delta) - F(x^\dagger) - F'(x^\dagger)e_k^\delta && \text{(A.3)} \\ p_k^\delta &:= (F'(x_k^\delta)^* - F'(x^\dagger)^*)(y^\delta - F(x_k^\delta)) && \text{(A.4)} \end{aligned}$$

and furthermore the closed expression

$$e_k^\delta = L^{-s}(I-B^*B)^k L^s(x_0 - x^\dagger) + \sum_{j=0}^{k-1} L^{-s}(I-B^*B)^{k-j-1}(B^*(y^\delta - y - q_j^\delta) + L^{-s}p_j^\delta)\,.$$

Together with (1.2), (2.10), and (3.10) we now obtain the following estimates

$$\begin{aligned} \|e_k^\delta\|_0 \;\le\; & \|(B^*B)^{\frac{u}{2(a+s)}}(I - B^*B)^k\|\,\|v\| \\ & + \sum_{j=0}^{k-1} \|(B^*B)^{\frac{a+2s}{2(a+s)}}(I - B^*B)^{k-j-1}\|(\delta + \|q_j^\delta\|) && \text{(A.5)} \\ & + \sum_{j=0}^{k-1} \|(B^*B)^{\frac{b+2s}{2(a+s)}}(I - B^*B)^{k-j-1}\|\,\|(B^*B)^{-\frac{b+s}{2(a+s)}}L^{-s}p_j^\delta\| \end{aligned}$$

and

$$\begin{aligned} \|e_k^\delta\|_{-a} \;\le\; & \|(B^*B)^{\frac{a}{2(a+s)}}(I - B^*B)^k\|\,\|v\| + \\ & + \sum_{j=0}^{k-1} \|(B^*B)(I - B^*B)^{k-j-1}\|(\delta + \|q_j^\delta\|) && \text{(A.6)} \\ & + \sum_{j=0}^{k-1} \|(B^*B)^{\frac{a+b+2s}{2(a+s)}}(I - B^*B)^{k-j-1}\|\,\|(B^*B)^{-\frac{b+s}{2(a+s)}}L^{-s}p_j^\delta\|_0\,. \end{aligned}$$

Next we derive estimates for $\|q_j^\delta\|$ and $\|(B^*B)^{-\frac{b+s}{2(a+s)}}L^{-s}p_j^\delta\|$. Assumption (N6), (3.5), and (A.3) imply that

$$\begin{aligned} \|q_j^\delta\| &\le \int_0^1 \|F'(x_j^\delta + \xi(x^\dagger - x_j^\delta)) - F'(x^\dagger))e_j^\delta\|\,d\xi && \text{(A.7)} \\ &\le \tfrac{c}{\beta+1}\|e_j^\delta\|_{-b}\|e_j^\delta\|_0^\beta \le \tfrac{c}{\beta+1}\|e_j^\delta\|_{-a}^{\frac{b}{a}}\|e_j^\delta\|_0^{\frac{a(1+\beta)-b}{a}}\,. \end{aligned}$$

Since $\tau > 2$, (1.2) and (1.4) imply that for all $0 \le k < k_*$

$$\|y^\delta - F(x_k^\delta)\| \;<\; 2\|y - F(x_k^\delta)\|\,.$$

Thus, we obtain together with (3.8), (3.11), (A.3), (A.4), and $F(x^\dagger) = y$ (cf. Assumption (N2)) that

$$\begin{aligned} \|(B^*B)^{-\frac{b+s}{2(a+s)}}L^{-s}p_j^\delta\|_0 &\le 2c\|y - F(x_j^\delta)\|\,\|e_j^\delta\|_0^\beta \\ &\le 2c(\|q_j^\delta\| + \|e_j^\delta\|_{-a})\|e_j^\delta\|_0^\beta && \text{(A.8)} \\ &\le \tilde\gamma(\|e_j^\delta\|_{-a}^{\frac{b}{a}}\|e_j^\delta\|_0^{\frac{a(1+2\beta)-b}{a}} + \|e_j^\delta\|_{-a}\|e_j^\delta\|_0^\beta) \end{aligned}$$

for all $0 \leq j < k$.

Combining the estimates, using spectral theory and Lemma 2.9 in [13] now yield the assertions (A.1) and (A.2).

We are now in the position to proof Proposition 3.5:

Proof[Proof of Proposition 3.5]

We proceed similar as in the proof of Theorem 2.3 in [13] and show by induction that

$$\|e_j^\delta\|_0 \leq \eta(j+1)^{-\frac{u}{2(a+s)}} \|x^\dagger - x_0\|_u \,, \qquad 0 \leq j \leq k_* \,, \tag{A.9}$$

and

$$\|e_j^\delta\|_{-a} \leq \eta(j+1)^{-\frac{a+u}{2(a+s)}} \|x^\dagger - x_0\|_u \,, \qquad 0 \leq j < k_* \,, \tag{A.10}$$

hold if $\|x^\dagger - x_0\|_u$ is sufficiently small and

$$\eta = \frac{4(\tau-1)}{\tau-2} \,. \tag{A.11}$$

The assertion holds for $j = 0$, if $\|x^\dagger - x_0\|_u$ is small enough. Furthermore, if $\|x^\dagger - x_0\|_u$ is so small that $\gamma^{-u}\eta\|x^\dagger - x_0\|_u \leq \rho$ then by (3.4) $x_j \in \widetilde{\mathcal{B}}_\rho(x^\dagger)$ and the iteration (3.3) is well-defined. Now assume (A.9), (A.10) are valid for $0 \leq j < k \leq k_*$. Then by virtue of Lemma A.1 the estimates

$$\begin{aligned} \|e_k^\delta\|_0 &\leq (1+\gamma_2\|x^\dagger - x_0\|_u^\beta)\|x^\dagger - x_0\|_u (k+1)^{-\frac{u}{2(a+s)}} + \delta k^{\frac{a}{2(a+s)}} & \text{(A.12)} \\ \|e_k^\delta\|_{-a} &\leq (1+\gamma_2\|x^\dagger - x_0\|_u^\beta)\|x^\dagger - x_0\|_u (k+1)^{-\frac{a}{2(a+s)}} + \delta & \text{(A.13)} \end{aligned}$$

hold for some $\gamma_2 > 0$ (independent of k). Here, we also have used the restriction $\frac{a-b}{\beta} < u \leq b + 2s$.

We will now derive an estimate for k in terms of δ: Similar to (A.8) we get

$$(\tau-1)\delta \leq \tilde{c}\|e_j^\delta\|_{-a}^{\frac{b}{a}}\|e_j^\delta\|_0^{\frac{a(1+\beta)-b}{a}} + \|e_j^\delta\|_{-a} \tag{A.14}$$

for all $0 \leq j < k \leq k_*$ and hence (A.9) and (A.10) for $j = k-1$ yield that

$$\delta \leq \tfrac{\tau}{2(\tau-1)}\eta k^{-\frac{a+u}{2(a+s)}} \|x^\dagger - x_0\|_u \tag{A.15}$$

provided that $\tilde{c}\eta^\beta \|x^\dagger - x_0\|_u^\beta \leq \frac{\tau-2}{2}$. This already proofs (3.14).

Together with (A.11) and (A.12) we obtain

$$\begin{aligned} \|e_k^\delta\|_0 &\leq \|x^\dagger - x_0\|_u (1+\gamma_2\|x^\dagger - x_0\|_u^\beta + \tfrac{\tau}{2(\tau-1)}\eta)(k+1)^{-\frac{u}{2(a+s)}} \\ &\leq \eta\|x^\dagger - x_0\|_u (k+1)^{-\frac{u}{2(a+s)}} \end{aligned}$$

if $\gamma_2\|x^\dagger - x_0\|_u^\beta \leq 1$ which we again assume to hold in the following. Similarly, we obtain that

$$\begin{aligned} \|e_k^\delta\|_{-a} &\leq \tfrac{2(\tau-1)}{\tau-2}(k+1)^{-\frac{a+u}{2(a+s)}} \|x^\dagger - x_0\|_u (1+\gamma_2\|x^\dagger - x_0\|_u^\beta) \\ &\leq \eta(k+1)^{-\frac{a+u}{2(a+s)}} \|x^\dagger - x_0\|_u . \end{aligned}$$

The estimate (A.10) follows by similar to (A.8). Thus, if $\|x^\dagger - x_0\|_u$ is sufficiently small, then the assertion holds for all $j \leq k_*$. In the case of exact data ($\delta = 0$), the estimates hold for all $k \geq 0$, since then Lemma A.1 holds for all $k \geq 0$.

References

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

[2] V. Fridman. Methods of successive approximations for Fredholm integral equations of the first kind. *Uspekhi Mat. Nauk*, 11:233–234, 1956. in Russian.

[3] C. W. Groetsch. *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind.* Pitman, Boston, 1984.

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

[5] 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.

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

[7] B. Kaltenbacher, A. Neubauer, and O. Scherzer. *Iterative Regularization Methods for Nonlinear Problems.* in preparation.

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

[9] R. Kreß. *Linear Integral Equations.* Springer, Berlin, 1989.

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

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

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

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

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