# Fast fully iterative Newton-type methods for inverse problems

Herbert Egger

Center for Computational Engineering Science (CCES)
RWTH Aachen University
D-52074 Aachen, Germany
email: herbert.egger@mathcces.rwth-aachen.de

**Abstract**

We study nonlinear inverse problems of the form

$$F(x) = y\,,$$

and their stable solution via iterative regularization methods, in particular by Newton-type methods, which are well-known for their fast convergence for well-posed problems. A basic step of Newton's method consists of calculating the update $\Delta x_k$ via solution of a linearized equation

$$F'(x_k^\delta)\,\Delta x_k^\delta = (y^\delta - F(x_k^\delta)),$$

which will in general be ill-posed if the nonlinear problem is. Thus, for ill-posed problems, the linearized equations have to be solved by some regularization method. In particular for large scale problems, e.g., inverse problems in partial differential equations, where $F(x)$ is only defined implicitly via the solution of a PDE, iterative methods have to be used for this purpose. In order to keep the overall effort, i.e., the overall number of iterations, as small as possible, appropriate preconditioning has to be applied.

We propose and analyse a general preconditioning strategy in Hilbert scales, and show that the overall number of iterations can be reduced to about the square root by preconditioning. Moreover, in many examples differential operators can be used as preconditioners, and thus preconditioning is almost for free. The theoretical results are illustrated in numerical examples. A comparison with preconditioned Landweber iteration shows that the iteration numbers can be further reduced, if fast iterative methods, e.g., the $\nu-$methods, are used for the solution of the linearized problems.

## 1 Introduction

This paper is concerned with the solution of inverse problems of the form

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

where $F : \mathcal{D}(F) \subset \mathcal{X} \to \mathcal{Y}$ is a continuous, Fréchet differentiable, nonlinear operator between Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$. We assume that a solution $x^\dagger$ of (1) exists, i.e., $F(x^\dagger) = y$. Many practically relevant inverse problems are ill-posed in the sense that their solution depends unstably on data perturbations, i.e., a solution of

$$F(x) = y^\delta, \qquad y^\delta = y + \text{"noise"} \tag{2}$$

might be arbitrarily far away from $x^\dagger$; moreover, a solution of (2) will not even exist in general. In order to recover reasonable approximations to $x^\dagger$ from perturbed data $y^\delta$, so-called regularization methods have to be used, cf., e.g., [3, 13, 15, 28]. For the analysis of the approximation quality of the regularized solutions, we require that a bound on the data noise

$$\|y - y^\delta\| \le \delta \tag{3}$$

is available. As shown by Bakushinskii [1], regularization is in general not possible without such information.

Important classes of inverse problems, e.g., inverse problems in nonlinear evolution or parameter identification governed by partial differential equations, are inherently nonlinear. The probably most investigated regularization method for solving ill-posed problems (2) is *Tikhonov regularization*, cf. [14, 28], where a regularized solution $x_\alpha^\delta$ is defined by

$$\|F(x) - y^\delta\|^2 + \alpha \|x - x^*\|^2 \to \min, \tag{4}$$

where the a-priori guess $x^*$ helps to select a special solution in case of non-uniqueness. One of the main drawbacks of Tikhonov regularization is that an optimal choice of the regularization parameter $\alpha$, e.g., by a *discrepancy principle*, requires to solve a sequence of minimization problems for different regularization parameters $\alpha$, and thus a numerical realization of Tikhonov regularization is often computationally expensive. Moreover, the minimization problem (4) has to be solved iteratively if $F$ is nonlinear. From this point of view, *iterative regularization methods* are a natural alternative to Tikhonov regularization, and parameter choice via the discrepancy principle

$$\|F(x_{k^*}^\delta) - y^\delta\| \le \tau\delta < \|F(x_k^\delta) - y^\delta\|, \qquad \text{for all } 0 \le k < k_* \tag{5}$$

just means to stop the iteration at the right time and requires no additional effort. In [20], nonlinear *Landweber iteration*

$$x_{k+1}^\delta = x_k^\delta + F'(x_k^\delta)(y^\delta - F(x_k^\delta)) \tag{6}$$

has been investigated and optimal convergence rates have been proven if the iteration is stopped according to (5).

Due to their fast convergence for well-posed problems, *Newton-type methods* are very attractive also for the solution of inverse problems (2). The basic step consist in determining the update $\Delta x_k^\delta = x_{k+1}^\delta - x_k^\delta$ by solving the linearized problem

$$F'(x_k^\delta)\, \Delta x_k^\delta = y^\delta - F(x_k^\delta). \tag{7}$$

Note that (7) is typically ill-posed, if (2) is, and thus some regularization method has to be applied. Several variants of Newton's method have been investigated in the framework of regularization methods, among them the iteratively regularized Gauß-Newton (IRGN) [2, 6] and the Levenberg-Marquardt method [18]; in both cases (7) is solved approximately by Tikhonov regularization. Note that for parameter identification problems, the operators $F$ and $F'$ are usually defined only implicitly via the solution of a PDE, cf. Example 1. In particular for large scale problems, only the action of the operators on some element can really be calculated, and therefor the linearized problems (7) have to be solved by some iterative method. The Newton-Landweber [21] or the Newton-CG iteration [19] have this structure. We will recall the definition and review the most important convergence results of a quite general class of Newton-type methods in Section 2.

One of the main drawbacks of iterative methods for ill-posed problems is that usually a large number of iterations is required in order to get optimal reconstructions. This holds true for linear and nonlinear problems as well. Even if only few Newton iterations are required, the iterative solution of the linearized equations (7) still requires many iterations, cf. [22]. The iteration numbers can be reduced in different ways, e.g., by

i) polynomial acceleration, e.g., $\nu-$methods or CG to solve the linearized equations,

ii) preconditioning.

We will investigate a combination of both approaches below.

A quite general strategy for preconditioning of iterative regularization methods has been proposed and analyzed in the framework of *regularization in Hilbert scales* in [11, 12]. It has been shown that the number of iterations for semi-iterative regularization methods for linear problems as well as for nonlinear Landweber iteration can be reduced to the square root, while still optimal convergence rates are obtained.

In this paper we show that Newton-type methods with iterative solution of the linearized equations – *fully iterative Newton-type methods* – can be accelerated by preconditioning in Hilbert scales and that, as for nonlinear Landweber iteration, the total number of (inner) iterations can be reduced to about the square root. Further acceleration can be achieved if (preconditioned) faster semi-iterative methods, e.g., the $\nu-$methods, are used for the iterative solution of the linearized problems.

The outline of this article is as follows: In the next section we formulate a general class of Newton-type methods and present the main convergence results. Section 3 recalls the definition of a Hilbert scale and summarises some auxiliary results needed for the subsequent analysis. In Section 4, we introduce a class of preconditioned Newton-type iterations and present the main convergence results. We conclude with an illustration of the theoretical results by numerical tests.

# 2 Iterative regularization of inverse problems

As already mentioned above, assembling the full operator $F$ respectively its linearization $F'$ is not possible in many "real world inverse problems". For illustration, we consider a simple example concerned with parameter identification:

**Example 1 (Reconstruction of a reaction term)** Consider heat conduction in a body $\Omega \subset \mathbb{R}^d$, $d = 2, 3$ with spatially varying reaction term $c$. In the stationary case, the temperature distribution $u = u(x)$ is given as solution of

$$\begin{aligned} -\Delta u + cu &= f, \qquad \text{in} \quad \Omega \\ u &= g, \qquad \text{on} \quad \partial\Omega, \end{aligned} \tag{8}$$

where $f$ denotes interior heat sources and $g$ is the prescribed temperature at the boundary. We define the parameter-to-output mapping $F$ by $F : c \mapsto u(c)$, where $u(c)$ denotes a solution of (8) with parameter $c$. The inverse problem now consist in determining $c$ from measurements of the state $u$.

For further examples of parameter estimation problems, we refer to [5, 13], and the references cited there. Note that the forward operator $F$ is only defined implicitly via the solution of the differential equation (8), which is the typical situation for inverse problems in PDEs. A method for solving the inverse problem should only require applications of the operator $F$ respectively $F'$, which amounts to solutions of PDE problems, and thus iterative methods are a natural choice in this context. In case of Newton-type methods, also the linearized equations (7) have to be solved by iterative (regularization) methods. For later reference, we briefly discuss iterative regularization of linear inverse problems and recall the most important convergence results.

## 2.1 Linear problems

Let $T : \mathcal{X} \to \mathcal{Y}$ be a continuous linear operator. For the iterative solution of $Tx = y^\delta$, we consider a class of *semi-iterative methods* (cf., e.g., [13, 16]) of the form

$$x_k^\delta = x_{k-1}^\delta + \mu_k(x_{k-1}^\delta - x_{k-2}^\delta) + \omega_k T^*(y^\delta - Tx_{k-1}^\delta), \quad k \geq 1. \tag{9}$$

Algorithms of this kind belong to the class of *Krylov-subspace* methods, i.e., the $k$-th iterate $x_k^\delta - x_0$ lies in the $k$-th Krylov subspace $\mathcal{K}_k(T^*T, T^*y^\delta)$, where for some self-adjoint operator $A$

$$\mathcal{K}_k(A, r) := \operatorname{span}\{r, Ar, \ldots, A^{k-1}r\}, \qquad k \geq 1.$$

Consequently, $x_k^\delta$ can be written as

$$x_k^\delta = x_0 + g_k(T^*T)T^*y^\delta, \tag{10}$$

where $g_k$ denotes the *iteration polynomial* of degree $k-1$. For an appropriate choice of the coefficients $\mu_j$ and $\omega_j$ the *residual polynomials* $r_k(\lambda) := 1-\lambda g_k(\lambda)$ satisfy the usual usual properties required in regularization theory, i.e.,

$$\begin{aligned} \sup_{\lambda\in[0,\|T^*T\|]} |\lambda^\mu r_k(\lambda)| \le c_\mu k^{-\sigma\mu}, \quad \text{for} \quad 0 \le \mu \le \mu_0 \\ \sup_{\lambda\in[0,\|T^*T\|]} |g_k(\lambda)| \le C_g\, k^\sigma \end{aligned} \tag{11}$$

for some $\mu_0 > 0$ and $\sigma \in \{1,2\}$, and hence the standard convergence results hold, i.e., for $x^\dagger - x_0 \in \mathcal{R}((T^*T)^\mu)$ for some $0 < \mu \le \mu_0 - 1/2$, one obtains the optimal rates

$$\|x_k^\delta - x^\dagger\| = O(\delta^{\frac{2\mu}{2\mu+1}}) \qquad \text{and} \qquad k_* = O(\delta^{-\frac{2}{\sigma(2\mu+1)}}), \tag{12}$$

if the iteration is stopped according to the linear analogue of the discrepancy principle (5), see [13] for details and proofs.

**Example 2 (Landweber iteration)** By choosing $\mu_{i,j} = 0$ and $\omega_j = 1$, we obtain *Landweber iteration*

$$x_{k+1}^\delta = x_k^\delta + T^*(y^\delta - Tx_k^\delta), \qquad k \ge 0 \tag{13}$$

as a special instance of (9) satisfying (11) with $\sigma = 1$ and $\mu_0 = \infty$.The number of iterations needed to obtain optimal convergence is $k_* = O(\delta^{-\frac{2}{2\mu+1}})$.

**Example 3 ($\nu-$methods)** A class of methods which requires much less iterations are the $\nu-$methods by Brakhage [7, 16], defined by (9) with

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

They satisfy (11) with $\sigma = 2$ and $\mu_0 = \nu$, thus yielding optimal rates of convergence with the stopping index bounded by $k_* = O(\delta^{-\frac{1}{2\mu+1}})$, which is only the square root of iterations than Landweber iteration. We want to mention that for methods satisfying (11) with $\sigma = 2$, the condition on $g_k(\lambda)$ already follows from the one on $r_k(\lambda)$ by Markov's inequality.

Even further acceleration is possible by adapting the iterative method to the data $y^\delta$, which is utilized by conjugate gradient type methods, cf. [17].

## 2.2 Nonlinear problems

The convergence analysis of iterative regularization methods for nonlinear problems is much more involved than in the linear case, and additional assumptions are required in order to guarantee similar convergence results: First, the nonlinearity of the problem has to be restricted. Additionally, due to the nonlinearity, the convergence results typically apply only in a neighborhood of the solution $x^\dagger$. The following assumptions are widely used in the literature, cf, e.g., [20, 22]:

**Assumption 1** (i) $F$ is Fréchet-differentiable in a ball $\mathcal{B}_\rho(x_0)$ and satisfies $\|F'(x)\| \le 1$.

(ii) $F'(x) = R(x,\bar{x})F'(\bar{x})$, for all $x$, $\bar{x} \in \mathcal{B}_\rho(x_0)$ and $\|R(x,\bar{x}) - I\| \le C_R\|x-\bar{x}\|$.

(iii) $F$ has a solution $x^\dagger \in \mathcal{B}_\rho(x_0)$ satisfying

$$x^\dagger - x_0 = \big(F'(x^\dagger)^*F'(x^\dagger)\big)^\mu w,$$

for some $\mu > 0$ and $\|w\|$ sufficiently small.

Note that under condition (ii), the range of the adjoint of the Fréchet-derivative is invariant in a neighborhood of $x_0$, i.e., $\mathcal{R}(F'(x)^*) = \mathcal{R}(F'(\bar{x})^*)$ for $x, \bar{x} \in \mathcal{B}_\rho(x_0)$. We refer to [9] for a discussion of invariance conditions for nonlinear inverse problems.

The following convergence rates result for nonlinear Landweber iteration (6) can be found in [20]:

**Theorem 1** *Let Assumption 1 hold for some $\rho$ sufficiently small and $\mu \leq 1/2$. If the iteration* (6) *is stopped according to the discrepancy principle* (5) *with $\tau$ sufficiently large, then*

$$k_* = O(\delta^{-\frac{2}{2\mu+1}}) \qquad \text{and} \qquad \|x_k^\delta - x^\dagger\| = O(\delta^{\frac{2\mu}{2\mu+1}}). \tag{14}$$

Next, we consider a solution of (2) is by Newton-type methods. As already mentioned, the linearized problem (7) usually inherits the ill-posedness from the nonlinear problem (2), and hence (7) has to be solved by some regularization method. We consider here a class of methods which solve (7) by regularizing around the initial iterate $x_0$. The iteratively regularized Gauß-Newton method [4, 6] and the Newton-Landweber iteration [21] are of this form. Furthermore, in order to avoid assembling of the linearized operators $F'$, we consider in particular iterative methods for the solution of the linearized systems. Alltogether, we are interested in fully iterative methods of the form

$$x_{n+1}^\delta = x_* + g_{k_n}(F'(x_n^\delta)^* F'(x_n^\delta)) F'(x_n^\delta)^* [y^\delta - F(x_n^\delta) - F'(x_n^\delta)(x_* - x_n^\delta)], \tag{15}$$

where $k_n$ denotes the termination index of the inner iteration in the $n$th Newton step. We only mention that under the above assumptions (i) – (iii), the rates (14) also hold this kind of methods. Under more restrictive assumptions on the iteration polynomials $g_k(\lambda)$, the analysis can be carried out under a weaker nonlinearity condition, namely

(ii$^*$) $F'(x) = R(x, \bar{x})F'(\bar{x}) + Q(x, \bar{x})$, for all $x$, $\bar{x} \in \mathcal{B}_\rho(x_0)$ with $\|R(x, \bar{x}) - I\| \leq C_R\|x - \bar{x}\|$ and $\|Q(x, \bar{x})\| \leq C_Q\|F(x) - F(\bar{x})\|$.

For a comprehensive discussion of various regularized Newton-type methods and their convergence theory, we refer to [22] and the references cited therein. In view of (14) one expects that a rather large number of iterations is required to obtain optimal convergence rates for Newton-type methods as well as for Landweber iteration. In Section 4, we will propose a preconditioned version of a fully iterative Newton-type method, which allows to reduce the number of iterations to about the square root. In order to analyse these methods, we will need some results concerning Hilbert scales.

# 3 Hilbert scales

In the following, we shortly introduce the notion of a Hilbert scale. For details we refer to [13, Section 8.4] and [23]. Further on, let $L$ be a densely defined, unbounded, selfadjoint, strictly positive operator in $\mathcal{X}$, and $\mathcal{M} := \bigcup_{s \in \mathbb{R}} \mathcal{D}(L^s)$.

**Definition 1** *For $x \in \mathcal{M}$ and $s \in \mathbb{R}$, let $\|x\|_s := \|L^s x\|$. Then the Hilbert spaces $\mathcal{X}_s$ are defined as the completion of $\mathcal{M}$ with respect to the norm $\|\cdot\|_s$, and $\{\mathcal{X}_s\}_{s\in\mathbb{R}}$ is called the* Hilbert scale *induced by $L$.*

This construction implies that the interpolation inequality

$$\|x\|_r \leq \|x\|_q^{\frac{s-r}{s-q}} \|x\|_s^{\frac{r-q}{s-q}} \tag{16}$$

holds for $-\infty < q < r < s < \infty$ and $x \in \mathcal{X}_s$. For illustration, we present two short examples:

**Example 4** Let $T : \mathcal{X} \to \mathcal{Y}$ be a compact, injective, linear operator between Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$. Then $L := (T^*T)^{-1}$ induces the Hilbert scale $\{\mathcal{X}_s\}_{s\in\mathbb{R}}$ with

$$\mathcal{X}_s := \mathcal{D}((T^*T)^{-s}) = \mathcal{R}((T^*T)^s)\,.$$

Note that in this case the spaces $\mathcal{X}_\mu$, $\mu \geq 0$, are the usual source sets for regularization in Hilbert spaces.

**Example 5** Let $\Omega \subset \mathbb{R}^n$, $n = 2, 3$, be a bounded domain with sufficiently smooth boundary $\partial\Omega$. Then

$$-\Delta : H^2(\Omega) \cap H_0^1(\Omega) \subset L_2(\Omega) \to L_2(\Omega)$$

satisfies the conditions of Definition 1, i.e., $L := -\Delta$ induces a Hilbert scale $\{\mathcal{X}_s\}_{s\in\mathbb{R}}$. Furthermore, $\mathcal{X}_s = H_0^{2s}(\Omega)$ for $s \in [0, 3/4)$, which means that the Sobolev spaces $H_0^s$ are part of the Hilbert scale $\{\mathcal{X}_s\}_{s\in\mathbb{R}}$, cf. [25].

The following proposition will play an important role in our convergence analysis:

**Proposition 1** *Let $T : \mathcal{X} \to \mathcal{Y}$ be a linear operator and $L$ be as above. Assume that for some $a > 0$ and $\overline{m} > 0$*

$$\|Tx\| \le \overline{m}\|x\|_{-a}, \qquad \text{for all} \quad x \in \mathcal{X} \tag{17}$$

*holds and that the extension of $T$ to $\mathcal{X}_{-a}$ (again denoted by $T$) is injective. Then the following assertions hold: $\mathcal{D}((B^*B)^{-\frac{\nu}{2}}) = \mathcal{R}((B^*B)^{\frac{\nu}{2}}) \subset \mathcal{X}_{\nu(a+s)}$ for all $\nu \in [0, 1]$ with $B := TL^{-s}$ for some $s \ge -a$, and*

$$\|(B^*B)^{\frac{\nu}{2}}x\| \le \overline{m}^{\nu}\|x\|_{-\nu(a+s)} \qquad \text{for all} \quad x \in \mathcal{X}\,, \tag{18}$$

$$\|(B^*B)^{-\frac{\nu}{2}}x\| \ge \overline{m}^{-\nu}\|x\|_{\nu(a+s)} \qquad \text{for all} \quad x \in \mathcal{D}((B^*B)^{-\frac{\nu}{2}})\,. \tag{19}$$

*Condition* (17) *is equivalent to*

$$\mathcal{R}(T^*) \subset \mathcal{X}_a \qquad \text{and} \qquad \|T^*w\|_a \le \overline{m}\|w\| \quad \text{for all} \quad w \in \mathcal{Y}\,. \tag{20}$$

*Now assume that in addition to* (17)

$$\underline{m}\|x\|_{-\tilde{a}} \le \|Tx\|, \qquad \text{for all} \quad x \in \mathcal{X} \tag{21}$$

*holds for some $\underline{m} > 0$, $\tilde{a} > 0$. Then 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*

$$\begin{aligned} \|(B^*B)^{\frac{\nu}{2}}x\| &\ge \underline{m}^{\nu}\|x\|_{-\nu(\tilde{a}+s)} \qquad \text{for all} \quad x \in \mathcal{X}\,, \\ \|(B^*B)^{-\frac{\nu}{2}}x\| &\le \underline{m}^{-\nu}\|x\|_{\nu(\tilde{a}+s)} \qquad \text{for all} \quad x \in \mathcal{X}_{\nu(\tilde{a}+s)}\,. \end{aligned} \tag{22}$$

*Moreover,* (21) *is equivalent to*

$$\begin{aligned} &\mathcal{X}_{\tilde{a}} \subset \mathcal{R}(T^*) \quad \text{and} \quad \|T^*w\|_{\tilde{a}} \ge \underline{m}\|w\| \\ &\qquad\qquad \text{for all} \quad w \in \mathcal{N}(T^*)^{\perp} \text{ with } T^*w \in \mathcal{X}_{\tilde{a}}. \end{aligned} \tag{23}$$

The proof of this result is based on the inequality of Heinz; for details see [12].

**Remark 1** By (20), it follows that $\mathcal{R}((T^*T)^{\frac{1}{2}}) = \mathcal{R}(T^*) \subset \mathcal{X}_a$, and hence

$$\mathcal{R}((T^*T)^{\mu}) \subset \mathcal{X}_{2a\mu}, \qquad \text{for} \quad 0 \le \mu \le \frac{1}{2}.$$

Note that in general, $x^\dagger \in \mathcal{X}_{2a\mu}$ will not imply $x^\dagger \in \mathcal{R}((T^*T)^{\mu})$. If, on the other hand, $\|Tx\| \ge \underline{m}\|x\|_{-\tilde{a}}$ for some $\underline{m} > 0$, then the converse inclusion $\mathcal{R}((T^*T)^{\mu}) \supset \mathcal{X}_{2\tilde{a}\mu}$ holds. Finally, if $T \sim L^{-a}$, i.e.,

$$\underline{m}\|x\|_{-a} \le \|Tx\| \le \overline{m}\|x\|_{-a} \quad \text{for} \quad x \in \mathcal{X}, \tag{24}$$

(which is the usual condition for regularization in Hilbert scales; cf. [13, 24, 27]) holds for some $a > 0$ and $0 < \underline{m} < \overline{m} < \infty$, then the spaces $\mathcal{X}_{2a\mu}$ and $\mathcal{R}((T^*T)^{\mu})$ coincide for $|\mu| \le 1/2$. Such norm equivalence conditions are important for preconditioning also for well-posed problems.

# 4 Preconditioned Newton-type iterations

Before we formulate and analyse a preconditioned version of Newton-type iterations for inverse problems, we shortly outline the framework of regularization in Hilbert scales on the basis of Landweber iteration. We consider the following preconditioned version:

$$x_{k+1}^\delta = x_k^\delta + L^{-2s}T^*(y^\delta - Tx_k^\delta), \qquad k \geq 0. \tag{25}$$

**Remark 2** The iteration (25) can be understood as standard Landweber iteration if $T$ is considered as operator from $\mathcal{X}_s$ to $\mathcal{Y}$. In fact, $T^\sharp := L^{-2s}T^*$ is the adjoint with respect to these spaces. Consequently, a general preconditioned iterative method can be written in the form

$$x_k^\delta = g_k(L^{-2s}T^*T)T^*y^\delta = L^{-s}g_k(B^*B)B^*y^\delta, \quad B := TL^{-s}.$$

Note that (25) can also be seen as standard Landweber iteration for the problem

$$Bz = y^\delta, \qquad x = L^s z, \qquad \text{with} \quad B : \mathcal{X} \to \mathcal{Y}. \tag{26}$$

Here, $L^s$ is considered as an operator from $\mathcal{X}$ to $\mathcal{X}_s$ in the second equation. The natural source condition for (26) is $z^\dagger \in \mathcal{R}((B^*B)^\mu)$ or equivalently

$$x^\dagger = L^{-s}(B^*B)^\mu w, \qquad \text{for some} \quad w \in \mathcal{X}. \tag{27}$$

Convergence (rates) for $x_k^\delta$ in the space $\mathcal{X}_s$ follow immediately by standard results of regularization theory. The important step in the analysis of regularization in Hilbert scales will be to derive optimal rates with respect to the original norm in the space $\mathcal{X}$.

**Remark 3** Originally, regularization in Hilbert scales was introduced for the case $s \geq 0$ in order to overcome saturation effects of Tikhonov regularization and Landweber iteration for nonlinear problems [24, 26]. The potential of preconditioning iterative methods by setting $s < 0$ has been investigated more recently in [11, 12], and optimal convergence rates have been shown under rather weak assumptions on the operator, namely

$$\|L^a T^*\| \leq \overline{m}. \tag{28}$$

Additionally, the optimal rates can be obtained with only the square root of iterations than needed by standard methods.

Preconditioned versions of the Newton-type methods (15) introduced in Section 2 can be constructed by considering $F$ respectively $F'$ as operators from $\mathcal{X}_s$ to $\mathcal{Y}$. Here, we will investigate iterations of the form

$$\begin{aligned} x_{n+1}^\delta &= x_0 + g_{k_n}(L^{-2s}A_n^*A_n)L^{-2s}A_n^*(y^\delta - F(x_n^\delta) + A_n(x_n^\delta - x_0) \\ &= x_0 + L^{-s}g_{k_n}(B_n^*B_n)B_n^*(y^\delta - F(x_n^\delta) + B_nL^s(x_n^\delta - x_0)), \end{aligned} \tag{29}$$

where we use the notations $A_n := F'(x_n^\delta)$ and $B_n := F'(x_n^\delta)L^{-s}$. The stopping indices $k_n$ for the inner iterations are assumed to increase with $n$, i.e., we assume that $k_0 > 0$ and

$$k_{n-1} \leq k_n \leq q\, k_{n-1} \quad \text{for} \quad n \in \mathbb{N}, \quad \text{and} \quad \lim_{n\to\infty} k_n = \infty, \tag{30}$$

for some $q > 1$; cf. [22] for similar conditions. The aim of this section is to give a detailed convergence analysis for the preconditioned iterations (29).

## 4.1 Basic Assumptions

The following conditions will be required for the convergence analysis of methods of the form (29):

**Assumption 2** *Similar to the condition* (28) *for linear problems, we require:*

(A1) $\|F'(x^\dagger)x\| \leq \overline{m}\|x\|_{-a}$ *for all* $x \in \mathcal{X}$, *some* $a > 0$, *and* $\overline{m} > 0$. *Moreover, the extension of* $F'(x^\dagger)$ *to* $\mathcal{X}_{-a}$ *is injective.*

(A2) $B := F'(x^\dagger)L^{-s}$ *is such that* $\|B\|_{\mathcal{X},\mathcal{Y}} \leq 1$, *where* $-a/2 \leq s \leq 0$.

*In order to cope with the nonlinear nature of the problem under consideration, we further require*

(A3) $x_0 \in \mathcal{B}_\rho(x^\dagger) := \{x \in \mathcal{X} : \|x - x^\dagger\| \leq \rho\} \subset \mathcal{D}(F)$ *for some* $\rho > 0$.

(A4) *For all* $x \in \mathcal{B}_\rho(x^\dagger)$ *there exists a linear operator* $R(x, x^\dagger)$ *with* $\|I - R(x, x^\dagger)\| \leq C_R < 1$ *such that*

$$F'(x) = R(x, x^\dagger)F'(x^\dagger)\,.$$

(A5) *There exists an element* $\omega > 0$, $w \in \mathcal{X}$ *with* $\|w\| \leq \omega$ *such that*

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

Before we start our analysis we shortly discuss the conditions above.

**Remark 4** Condition (A1) amounts to (28) in the linear case. Note that only a one-sided estimate is needed, i.e., in view of Proposition 1 it suffices that $F'(x)^*$ is sufficiently smoothing in terms of the Hilbert scale $\{\mathcal{X}_s\}_{s\in\mathbb{R}}$. (A2) is a simple scaling condition. (A3) is a standard closeness condition, which is typically required for nonlinear problems. Condition (A4) is the nonlinearity condition (ii) already used for the analysis of nonlinear Landweber iteration, cf. Section 2 and [20]. Finally, as outlined in Remark 2, condition (A5) is the natural source condition for regularization in Hilbert scales. If the smoothness of the operator $F'(x^\dagger)$ can be estimated from below, i.e., if there exist $\tilde{a}$ and $\underline{m}$ such that

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

then by Proposition 1, the source condition (A5) can be interpreted as condition in the Hilbert scale $\{\mathcal{X}_r\}$, respectively as standard condition $x^\dagger \in \mathcal{R}((T^*T)^\mu)$ for some appropriate $\mu$. We refer to [10, 11, 12] for a detailed discussion.

## 4.2 A-priori estimates

We start with investigating the convergence behavior for the iteration (29) when stopped by an *a-priori* rule. In our analysis, we will use the following auxiliary result:

**Lemma 1** *Let* $A$, $B$, $R$ *be bounded linear operators between Hilbert spaces* $\mathcal{X}$ *and* $\mathcal{Y}$. *If* $B = RA$ *with* $\|I - R\| < 1$, *then for every* $|\nu| \leq 1/2$ *and* $w \in \mathcal{X}$ *there exist positive constants* $\underline{c}, \overline{c}$ *and an element* $v \in \mathcal{X}$ *such that*

$$(A^*A)^\nu w = (B^*B)^\nu v,$$

*with* $\underline{c}\|w\| \leq \|v\| \leq \overline{c}\|w\|$.

*Proof.* Observing that $\mathcal{R}((A^*A)^{1/2}) = \mathcal{R}(A^*) = \mathcal{R}(B^*) = \mathcal{R}((B^*B)^{1/2})$, the result follows by the inequality of Heinz and duality arguments (cf. [22] for details). □

**Proposition 2** *Let Assumption 2 hold with* $C_R$, $\omega$ *sufficiently small, and (A7) hold for some* $0 < u \leq a + 2s$. *Furthermore, let* $y^\delta \in \mathcal{Y}$ *satisfy* (3), *and let* $x_n^\delta$ *denote the iterates defined by* (29) *with* $r_k$ *satisfying* (11) *for some* $\mu_0 \geq 1$. *Finally, let* $k_n$ *satisfy* (30), *and for* $\eta > 0$ *let* $N(\delta)$ *denote the largest integer such that*

$$k_n \leq \left(\frac{1}{\eta\omega}\delta\right)^{-\frac{2(a+s)}{a+u}} \tag{31}$$

*for all* $0 \leq n \leq N(\delta)$.

*Then there exists a positive constant $C_\eta$ such that the estimates*

$$\|x_n^\delta - x^\dagger\| \leq C_\eta\, k_n^{-\frac{\sigma u}{2(a+s)}}\omega, \tag{32}$$

$$\|F'(x^\dagger)(x_n^\delta - x^\dagger)\| \leq C_\eta\, k_n^{-\frac{\sigma(u+a)}{2(a+s)}}\omega \tag{33}$$

*hold and $x_n^\delta \in \mathcal{B}_\rho(x^\dagger)$ for $0 \leq n \leq N(\delta)$.*

*Proof.* We prove the assertions by induction: By assumption $\|x_0 - x^\dagger\| \leq \omega$, and hence (32), (33) holds for $n = 0$ if $C_\eta$ is chosen large enough.

Now let (32), (33) hold for some $0 < n < N(\delta)$ and assume that $x_n^\delta \in \mathcal{B}_\rho(x^\dagger)$. Then with the notation $e_n^\delta := x_n^\delta - x^\dagger$ and (29), we get the closed form representation

$$e_{n+1}^\delta = L^{-s} r_{k_n}(B_n^*B_n)L^s(x_0 - x^\dagger) + L^{-s} g_{k_n}(B_n^*B_n)B_n^*(y^\delta - y + l_n),$$

with $l_n := \int_0^1 [F'(x^\dagger + te_n^\delta) - F'(x_n^\delta)]\, e_n^\delta\, dt$. By the nonlinearity condition (A4) and Lemma 1, there exists a $w_n \in \mathcal{X}$ with $\|w_n\| \sim \|w\|$ such that

$$(B^*B)^{\frac{u-s}{2(a+s)}} w = (B_n^*B_n)^{\frac{u-s}{2(a+s)}} w_n$$

and hence

$$\begin{aligned}\|e_{n+1}^\delta\| \leq\ & c\,\|(B_n^*B_n)^{\frac{u}{2(a+s)}} r_{k_n}(B_n^*B_n)w_n\| \\ & + c\,\|(B_nB_n^*)^{\frac{a+2s}{2(a+s)}} g_{k_n}(B_nB_n^*)(y^\delta - y + l_n)\|\end{aligned}$$

for some $c > 0$. With (11) we obtain

$$c\,\|(B_n^*B_n)^{\frac{u}{2(a+s)}} r_{k_n}(B_n^*B_n)\, w_n\| \leq c_1 k_n^{-\frac{\sigma u}{2(a+s)}}\omega.$$

Next we estimate

$$\begin{aligned}\|l_n\| &= \|\int_0^1 [F'(x^\dagger + te_n^\delta) - F'(x_n^\delta)]\, e_n^\delta\, dt\| \\ &\leq 2C_R\|(F'(x^\dagger)e_n^\delta\| \leq 2C_RC_\eta k_n^{-\frac{\sigma(u+a)}{2(a+s)}}\omega\,,\end{aligned}$$

which together with (11) yields

$$c\,\|(B_nB_n^*)^{\frac{a+2s}{2(a+s)}} g_{k_n}(B_nB_n^*)(y^\delta - y + l_n)\| \leq c_2 k_n^{\frac{\sigma a}{2(a+s)}}\delta + c_3C_RC_\eta k_n^{-\frac{\sigma u}{2(a+s)}}\omega.$$

In order to establish (32) for $n+1$, one has to choose $C_\eta$ such that

$$C_\eta(1 - c_3C_R) \geq c_1 + c_2\eta,$$

where we used the bound (31). Note that such a choice is always possible and independent of $n$ as long as $c_3C_R < 1$. Finally, if $\omega$ is sufficiently small, e.g., so small that $C_\eta\omega k_0^{-\frac{\sigma u}{2(a+s)}} \leq \rho$, then (32) yields $x_{n+1}^\delta \in \mathcal{B}_\rho(x^\dagger)$.

The second estimate (33) follows in a similar manner. □

Proposition 2 immediately implies the following convergence rates in terms of $\delta$:

**Corollary 1** *Let the assumptions of Proposition 2 be valid and let $N(\delta)$ be chosen as in* (31)*. Then the following rates hold:*

$$\|x_{N(\delta)}^\delta - x^\dagger\| = O(\delta^{\frac{u}{a+u}}) \quad \textit{and} \quad \|F'(x^\dagger)(x_{N(\delta)}^\delta - x^\dagger)\| = O(\delta). \tag{34}$$

*Proof.* The assertion follows immediately from the previous proposition. □

**Remark 5** The rates (34) even hold in the slightly stronger norm $|||x||| := \|(B^*B)^{\frac{s}{2(a+s)}} L^s x\|$ and are order optimal with respect to this norm and the source condition (A5); we refer to [11, 12] for a detailed discussion. For the choice $k_n \sim k_0 q^n$ with $q > 0$, the number of (outer) Newton iterations is $O(1 + |\log \delta|)$ and hence the overall number of inner iterations is bounded by

$$k_* = \sum_{n=0}^{N(\delta)} k_n = O(\delta^{-\frac{2(a+s)}{a+u}}), \quad \text{respectively} \quad k_* = \sum_{n=0}^{N(\delta)} k_n = O(\delta^{-\frac{a+s}{a+u}})$$

for the preconditioned Newton-Landweber iteration and the preconditioned Newton-$\nu$-methods, respectively. With $s = -a/2$, the resulting iteration numbers can be reduced to the square root by preconditioning, which corresponds to the acceleration effect already observed for linear problems and for nonlinear Landweber iteration [11, 12]. Moreover, the preconditioned Newton-$\nu$-methods yield optimal convergence with only the square root of iterations that would be needed for the preconditioned Landweber iteration and only the fourth root of iterations needed for the standard Landweber iteration. We will demonstrate this substantial speed-up in numerical examples in Section 5.

## 4.3 A-posteriori stopping

The *a-priori* results of Proposition 2 and Corollary 1 are not of great use *per-se*, since in general one does not know the smoothness of the solution, i.e., for which $u$ the source condition (A5) holds. However, the estimate of Proposition 2 will be used to prove convergence rates, when the iteration is stopped according to the following generalized discrepancy principle (cf. [22]):

**Stopping rule:** For some $\tau > 1$ let $n_* = n(\delta, y^\delta)$ be the smallest integer such that

$$\max\{\|y^\delta - F(x^\delta_{n_*-1})\|, \|y^\delta - F(x^\delta_{n_*})\|\} \le \tau\delta. \tag{35}$$

According to (35), the (outer) Newton iteration is stopped, when the first time two consecutive residuals are less than $\tau\delta$. The following proposition guarantees stability of our class of preconditioned Newton-type methods (29) equipped with the above criterion:

**Proposition 3** *Let the assumptions of Proposition 2 be valid, and the iteration* (29) *be stopped according to* (35) *with $\tau$ sufficiently large. Then, the iteration is well-defined and $n_* \le N(\delta)$ with $N(\delta)$ as in* (31).

*Proof.* For some $\eta > 0$, Proposition 2, (3) and (A4) imply that

$$\|F(x^\delta_n) - y^\delta\| \le \delta + (1 + C_R)\|F'(x^\dagger)e^\delta_n\| \le \delta + (1 + C_R)\, C_\eta k_n^{-\frac{u+a}{2(a+s)}} \omega$$

for all $0 \le n \le N(\delta)$. This together with (30) and (31) yields the estimate

$$\|F(x^\delta_n) - y^\delta\| \le \delta\,[1 + (1 + C_R)C_\eta \eta^{-1} q^{-\frac{a+u}{a+s}}] \tag{36}$$

for $n = N(\delta) - 1$ and $n = N(\delta)$. Thus, if $\tau$ is larger than the constant in square brackets above, then $n_* \le N(\delta)$ and, due to Proposition 2, the iteration is well-defined up to the stopping index $n_*$. □

We are now in the position to prove the following convergence rates result:

**Theorem 2** *Let the assumptions of Proposition 2 be satisfied, and the iteration* (29) *be stopped after $n_* = n(\delta, y^\delta)$ steps according to the stopping rule* (35) *with some $\tau$ sufficiently large. Then*

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

*Proof.* We use the notation $A = F'(x^\dagger)$ and $A_n = F'(x_n^\delta)$. Observe, that by (3) and (A4) the following estimate holds for $n = n_*$ and $n = n_* - 1$:

$$\begin{aligned}\|A_n e_n^\delta\| &\le (1 + C_R)\|A e_n^\delta\| \\ &\le 2\,\|y - F(x_n^\delta) - \int_0^1 [F'(x_n^\delta - t e_n^\delta) - F'(x^\dagger)]\, e_n^\delta\, dt\| \\ &\le 2\,[\delta + \|F(x_n^\delta) - y^\delta\| + \tfrac{C_R}{1-C_R}\|A_n e_n^\delta\|],\end{aligned}$$

and hence with (35), $C_R < 1/2$ and (36)

$$\|A_n e_n^\delta\| \le c_1 \delta, \qquad n \in \{n_* - 1, n_*\} \tag{37}$$

for some positive constant $c_1$. Next, by (29), and denoting $n = n_* - 1$, $B_n = A_n L^{-s}$ we have

$$A_n e_{n^*}^\delta = B_n r_{k_n}(B_n^* B_n) L^s (x_0 - x^\dagger) + B_n g_{k_n}(B_n^* B_n) B_n^* [y^\delta - F(x_n^\delta) + A_n e_n^\delta)]\,.$$

Thus, we obtain with (A4), (37), and (35) that

$$\begin{aligned}&\|B_n r_{k_n}(B_n^* B_n) L^s (x_0 - x^\dagger)\| \\ &\quad = \|A_n e_{n_*}^\delta - B_n B_n^* g_{k_n}(B_n^* B_n)[y^\delta - F(x_n^\delta) + A_n(x_n^\delta - x^\dagger)]\| \\ &\quad \le \tfrac{1+C_R}{1-C_R}\|A_{n_*} e_{n_*}^\delta\| + c_2(\|y^\delta - F(x_n^\delta)\| + \|A_n e_n^\delta\|) \le c_3 \delta,\end{aligned}$$

for some $c_2$, $c_3 > 0$. Finally, using the above estimates, the representation (29), (A5), and $n = n_* - 1$ the error can be estimated as follows:

$$\begin{aligned}\|e_{n_*}^\delta\| &\le \|L^{-s} r_{k_n}(B_n^* B_n)(B^* B)^{\frac{u-s}{2(a+s)}} w\| + \|L^{-s} g_{k_n}(B_n^* B_n) B_n^* (y^\delta - F(x_n^\delta) + A_n e_n^\delta)\| \\ &\le c_4\big(\|(B_n^* B_n)^{\frac{u}{2(a+s)}} r_{k_n}(B_n^* B_n) w_n\| \\ &\qquad + \|g_{k_n}(B_n B_n^*)(B_n B_n^*)^{\frac{a+2s}{2(a+s)}}\|(\tau\delta + \|A_n e_n^\delta\|\big) \\ &\le c_5\big(\|r_{k_n}(B_n^* B_n)(B_n^* B_n)^{\frac{u}{2(a+s)}} w_n\| + k_n^{\frac{\sigma a}{2(a+s)}}\delta\big)\end{aligned}$$

for some constants $c_4$, $c_5 > 0$. Using the interpolation inequality, the above estimates and (A5), we obtain

$$\begin{aligned}\|(B_n^* B_n)^{\frac{u}{2(a+s)}} r_{k_n}(B_n^* B_n) w_n\| &\le c_6 \|(B_n^* B_n)^{\frac{u+a}{2(a+s)}} r_{k_n}(B_n^* B_n) w_n\|^{\frac{u}{a+u}} \|w_n\|^{\frac{a}{a+u}} \\ &\le c_6 \|B_n r_{k_n}(B_n^* B_n) L^s (x_0 - x^\dagger)\|^{\frac{u}{a+u}} \|w_n\|^{\frac{a}{a+u}} \\ &\le c_7 \delta^{\frac{u}{a+u}} \omega^{\frac{a}{a+u}}\end{aligned}$$

for some positive constants $c_6$ and $c_7$. This together with Proposition 3 and (31) completes the proof. □

**Remark 6** We want to mention that the above results can be strengthened in several directions: First, the rates can be proven in a slightly stronger norm, cf. Remark 5, and then are order optimal; see [11, 12]. For nonlinear Landweber iteration, the convergence rates can proven under more flexible nonlinearity conditions, e.g.,

$$F'(x) = R(\bar{x}, x) F'(\bar{x}) + Q(\bar{x}, x)$$

with

$$\|I - R(x, x^\dagger)\| \le C_R < 1, \qquad \|Q(x, x^\dagger)\|_{\mathcal{X}_{-b}^s, \mathcal{Y}} \le C_Q \,|\!|\!| x - x^\dagger |\!|\!|_{b-a},$$

for some $b \in [0, a]$, $\beta \in (0, 1]$, and $C_R, C_Q > 0$, where $\mathcal{X}_b^s$ is the space equipped with the norm $\|x\|_{\mathcal{X}_r^s} = |\!|\!| x |\!|\!|_r = \|(B^* B)^{\frac{s-r}{2(a+s)}} L^s x\|$, cf. [10, 12]. We conjecture that at least for the preconditioned IRGN and Newton-Landweber method, the convergence rates can be proven under this assumption instead of (A4). The generalized discrepancy principle (35) was already used in [22] for the proof of optimal convergence rates of the IRGN method.

**Remark 7** Another very efficient fully iterative Newton-type method can obtained by solving the linear Tikhonov regularization problems of IRGN with a conjugate gradient method. We only mention that our theory applies literally to IRGN by replacing $g_k$ in (29) by $g_\alpha$. As a consequence of Propositions 2 and 3 one obtains that the regularization parameter $\alpha_*$, where the discrepancy principle is reached, is about the square-root from the one without preconditioning. A larger value of $\alpha$ implies that the spectrum of the operator $(L^{-2s}T^*T+\alpha I)$ (which appears in the preconditioned iteration) clusters at a larger value and thus fewer iterations are required to solve the linearized systems approximately. We refer to [10] for an analysis of CGNE in Hilbert scales.

# 5 Numerical test examples

In this Section we verify the Assumptions (A1)-(A5) needed in our convergence for some test examples. Moreover, we illustrate the theoretical results by numerical experiments. For the numerical tests, the corresponding equations are discretized by standard finite elements. In order to avoid inverse crimes, the data $y$ are computed on a finer grid and additionally perturbed by uniformly distributed random noise.

**Example 6 (A nonlinear Volterra-Hammerstein integral equation)** The following integral equation of the first kind is a special case from an example discussed in [20, 26]: 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

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

and

$$\|F'(x^\dagger)^* w\|_{H^3} \sim \|w\|, \quad \text{for all} \quad w \in \mathcal{Y}.$$

As a Hilbert scale we choose the one induced by $L^2 x := -x'' + x$ over the space $\mathcal{X} = H^1[0,1]$ with $\mathcal{X}_1 = \{x \in H^2[0,1] : x'(0) = x(1) = 0\}$. With this choice, we have

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

and hence, by Proposition 1, (A1) holds with $a = 2$. Therefore, we set $s = -1$, which yields

$$L^{-2s}F'(x)^* w = 2x(\cdot)\int_\cdot^1 w(t)dt; \tag{38}$$

in particular, we have

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

with

$$\|R(x, x^\dagger) - I\| \leq C\|x - x^\dagger\|_0,$$

which proves (A4) with $C_R$ arbitrarily small for $x$ close to $x^\dagger$.

Note that in this example the application of the Hilbert scale operator $L^{-2s}$ in fact makes the iteration even simpler, i.e., application of $A^{-1}$, which is a main part or the numerical effort for calculating $F'(x)^*$, can be avoided.

For a numerical test, we set

$$x^\dagger(t) := \begin{cases} 1/2+t, & 0 \le t \le 1/2 \\ 3/2-t, & 1/2 < t \le 1 \end{cases} \quad \text{and} \quad x_0 = 1/2,$$

and compare the performance of Landweber iteration, the Newton-Landweber and the Newton-$\nu$ method, and their preconditioned equivalents. The iteration numbers realized in the numerical tests are listed in Table 1.

| $\delta/\|y\|$ | lw | new-lw | new-$\nu$ | hs-lw | new-hs-lw | new-hs-$\nu$ |
|---|---|---|---|---|---|---|
| 0.02 | 52 | 6(126) | 5(62) | 5 | 3(14) | 3(14) |
| 0.01 | 110 | 7(254) | 5(62) | 11 | 4(30) | 3(14) |
| 0.005 | 278 | 9(1022) | 6(126) | 18 | 4(30) | 3(14) |
| 0.002 | 711 | 10(2046) | 7(254) | 29 | 5(62) | 4(30) |
| 0.001 | 1615 | 11(4094) | 7(254) | 40 | 6(126) | 4(30) |
| $\eta$ | -1.14 | -1.18 | -0.56 | -0.67 | -0.67 | -0.31 |

Table 1: Iteration numbers (outer and total inner iterations) for iterative regularization methods and their Hilbert-scale equivalents and the corresponding rates $k_* = O(\delta^\eta)$; parameters $\tau = 2.1$, $\nu = 2$ and $N_t = 501$.

Note, that as predicted by the theory the rates of the overall iteration numbers are significantly smaller for the preconditioned iterations than for the standard iterations. We only mention that all iterations yield similar reconstructions and convergence rates $\|x_{k^*}^\delta - x^\dagger\| \sim \delta^{0.2}$.

**Example 7 (Parameter identification)** In this example, which is taken from [20], we try to identify the parameter $c$ in

$$\begin{aligned} -\Delta u + cu &= f \quad \text{in} \quad \Omega, \\ u &= g \quad \text{in} \quad \partial\Omega, \end{aligned} \tag{39}$$

from distributed measurements of the state $u$. Here, $\Omega$ is an interval in $\mathbb{R}$ or a bounded domain in $\mathbb{R}^2$ or $\mathbb{R}^3$ with smooth boundary or a parallelepiped. The right hand side is assumed to satisfy $f \in L_2(\Omega)$ and the boundary data $g \in H^{3/2}(\partial\Omega)$. If $u$ would be known exactly, then one could reconstruct $c$ by

$$c = \frac{f + \Delta u}{u}, \tag{40}$$

which in case of noisy measurements $u^\delta$ is unstable due to differentiation. (40) already reveals another possible source of instability, namely division by $u$, which may cause noise amplification where $u$ is close to zero. Note that if $u = 0$ on a subdomain, then $c$ is not uniquely determined by (39) there, and hence cannot be reconstructed.

We consider the inverse problem as abstract operator equation, and define the nonlinear parameter-to-solution mapping

$$F : \mathcal{D}(F) \subset L_2(\Omega) \to L_2(\Omega)$$

with $F(c) = u(c)$, and $u = u(c)$ denoting the solution of (39) with parameter $c$. One can show (cf. [8]) that the operator $F$ is well-defined and Fréchet differentiable on

$$\mathcal{D}(F) := \{c \in L_2(\Omega) : \|c - \bar{c}\| \le \gamma \text{ for some } \bar{c} \ge 0 \text{ a.e.}\}$$

where $u(c)$ denotes the solution of (39), and $\gamma > 0$ has to be sufficiently small. By standard arguments one can show that

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

where $A(c) : H^2(\Omega) \cap H_0^1(\Omega) \to L_2(\Omega)$ is defined by $A(c)u = -\Delta u + cu$.

Next we choose an appropriate Hilbert scale namely the one induced by $L^2 = -\Delta$ over $\mathcal{X} = L_2(\Omega)$ with $\mathcal{X}_2 := H^2(\Omega) \cap H_0^1(\Omega)$. This yields

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

which already proves (A1). If furthermore $u^\dagger \geq \gamma > 0$ a.e., then we even have

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

and according to Remark 4, (A5) can be interpreted in terms of the spaces $\mathcal{X}_u$. In order to show (A4), we use that

$$\begin{aligned} [F'(c)^* - F'(d)^*]w &= u(c)[A(c)^{-1} - A(d)^{-1}]\, w + [u(d) - u(c)]A(d)^{-1} w \\ &=: r_1 + r_2. \end{aligned}$$

The terms $r_1$ and $r_2$ can be further estimated by

$$\begin{aligned} \|r_1\|_2 &\leq C\|u(c)\|_{H^2}\|[A(c)^{-1} - A(d)^{-1}]w\|_{H^2} \\ &\leq C_1\|u(c)\|_{H^2}\|c - d\|_{L_2}\|w\|_{L_2} \end{aligned}$$

and

$$\begin{aligned} \|r_2\|_2 &\leq \|u(d) - u(c)\|_{H^2}\|A(d)^{-1} w\|_{H^2} \\ &\leq C_2\|c - d\|_{L_2}\|w\|_{L_2}. \end{aligned}$$

Here we used that $A(c)$ is an isomorphism between $L_2(\Omega)$ and $H^2(\Omega) \cap H_0^1(\Omega)$. If $u^\dagger \geq \gamma > 0$, this yields (A4) with $C_R \leq C\|c_0 - c^\dagger\|$.

In our numerical test, we try to reconstruct the reaction term

$$c^\dagger = \operatorname{sign}(x - 0.5) \cdot \operatorname{sign}(y - 0.5)$$

on $\Omega = [0, 1]^2$ and start with the initial guess $c_0 = 0$. Obviously $c^\dagger$ is non-smooth and hence one expects that a large number of iterations is necessary for a reasonable reconstruction. For comparison, we apply the Newton-Landweber and the Newton-$\nu$ method with their preconditioned ($s = -1$) equivalents. The iteration number achieved in the numerical tests are summarized in Table 2.

| $\frac{\delta}{\|u^\delta - u(c_0)\|}$ | new-lw | new-$\nu$ | hs-new-lw | hs-new-$\nu$ |
|---|---|---|---|---|
| 0.08 | 5(70) | 3(25) | 1(5) | 1(5) |
| 0.04 | 8(266) | 5(70) | 2(13) | 2(13) |
| 0.02 | 11(931) | 7(173) | 4(43) | 3(25) |
| 0.01 | 13(2114) | 8(266) | 5(70) | 3(25) |
| 0.005 | 16(7174) | 9(406) | 6(111) | 4(43) |
| $\eta$ | -1.63 | -0.99 | -1.13 | -0.71 |

Table 2: Iteration numbers for iterative regularization methods and their Hilbert-scale equivalents and the corresponding rates $k_* = O(\delta^\eta)$; parameters $\tau = 2.1$, $\nu = 2$.

As expected, the overall iteration numbers can be reduced substantially by preconditioning. The numerical reconstructions are similar for all methods, and the observed convergence rates are approximately $\|x_{k^*}^\delta - x^\dagger\| = O(\delta^{0.2})$.

# References


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

[2] A. B. Bakushinskii and A. V. Goncharskii. *Iterative methods for the solution of incorrect problems.* Nauka, Moscow, 1989.

[3] A. B. Bakushinskii and A. V. Goncharskii. *Ill-Posed Problems: Theory and Applications.* Kluwer, Dordrecht, 1994.

[4] A. W. Bakushinskii. The problem of the convergence of the iteratively regularized Gauß-Newton method. *Comput. Math. Math. Phys.*, 32:1353–1359, 1992.

[5] H. Banks and K. Kunisch. *Parameter Estimation Techniques for Distributed Systems.* Birkhäuser, Braunschweig, 1989.

[6] B. Blaschke, A. Neubauer, and O. Scherzer. On convergence rates for the iteratively regularized Gauss-Newton method. *IMA Journal of Numer. Anal.*, 17:421–436, 1997.

[7] 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.

[8] F. Colonius and K. Kunisch. Stability for parameter estimation in two point boundary value problems. *J. Reine Angew. Math.*, 370:1–29, 1986.

[9] P. Deuflhard, H. W. Engl, and O. Scherzer. A convergence analysis of iterative methods for the solution of nonlinear ill-posed problems under affinely invariant conditions. *Inverse Problems*, 14:1081–1106, 1998.

[10] H. Egger. *Preconditioning Iterative Regularization methods in Hilbert Scales.* PhD thesis, Johannes Kepler Universität, Linz, 2005.

[11] H. Egger. Semiiterative regularization in Hilbert scales. *Siam J. Numer. Math.*, 44:66–81, 2006.

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

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

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

[15] C. W. Groetsch. *Inverse Problems in the Mathematical Sciences.* Vieweg, Braunschweig, 1993.

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

[17] M. Hanke. *Conjugate Gradient Type Methods for Ill-Posed Problems.* Number 327 in Pitman Research Notes in Mathematics. Longman Scientific & Technical, Harlow, UK, 1995.

[18] M. Hanke. A regularization Levenberg-Marquart scheme, with application to inverse groundwater filtration problems. *Inverse Problems*, 13:79–95, 1997.

[19] M. Hanke. Regularizing properties of a truncated Newton-CG algorithm for nonlinear inverse problems. *Numer. Func. Anal. Optim.*, 18:971–993, 1997.

[20] 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.

[21] B. Kaltenbacher. Some Newton-type methods for the regularization of nonlinear ill-posed problems. *Inverse Problems*, 13:729–753, 1997.

[22] B. Kaltenbacher. A-posteriori parameter choice strategies for some Newton type methods for the regularization of nonlinear ill-posed problems. *Numer. Math.*, 79:501–528, 1998.

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

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

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

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

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

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