Accelerated Newton-Landweber Iterations for Regularizing Nonlinear Inverse Problems*

Herbert Egger

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

Abstract

In this paper, we investigate the convergence behaviour of a class of regularized Newton methods for the solution of nonlinear inverse problems.

In order to keep the overall numerical effort as small as possible, we propose to solve the linearized equations by certain *semiiterative* regularization methods, in particular, iterations with *optimal speed of convergence*.

Our convergence rate analysis of this class of *accelerated Newton-Landweber* methods contributes to the analysis of Newton-type regularization methods in two ways: first, we show that under standard assumptions, accelerated Newton-Landweber iterations yield optimal convergence rates under appropriate *a priori* stopping criteria. Secondly, we prove inproved convergence rates for $\mu > 1/2$ under an adequate *a posteriori* stopping rule, thus extending existing results. Our theory naturally applies to a wider class of Newton-type regularization methods.

We conclude with several examples and numerical tests confirming the theoretical results, including a comparison to the Gauß-Newton method and the Newton-Landweber iteration.

1 Introduction

Many mathematically and physically relevant problems are concerned with determining causes for a desired or an observed effect. As a general model for such inverse problems, we consider the abstract operator equation

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

where $F : \mathcal{D}(F) \subset \mathcal{X} \to \mathcal{Y}$ is a (nonlinear) operator between Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$. We are especially interested in the case, when a solution of (1) is ill-posed in the

*supported by the Austrian Science Fund (FWF) under grant SFB/F013, "Numerical and Scientific Computing", Project F1308

sense of Hadamard, in particular if a solution does not depend continuously on the right hand side. If only perturbed data y^δ with a bound on the noise level

$$\|y - y^\delta\| \leq \delta \tag{2}$$

are available, the solution of (1) has to be regularized in order to get reasonable (stable) approximations.

Tikhonov regularization (cf., e.g., [3, 4]) is probably the most well known regularization method for linear as well as nonlinear inverse problems. However, when it comes to an implementation of Tikhonov regularization for nonlinear or even large scale linear problems, iterative methods are often or even have to be used for finding a minimizer of the Tikhonov functional. A more direct approach is offered by iterative methods, which have been investigated in the framework of regularization of nonlinear problems more recently (see, e.g., [1, 6, 8, 9, 12]). Beside simple gradient-type iterations, like Landweber iteration (cf. [14]), Newton-type methods seem to be especially attractive, due to their well known, fast convergence properties for well-posed problems. In order to take into account the ill-posed nature of the problem (1), several variants of Newton's method have been proposed for the stable, iterative solution of inverse problems, e.g., the iteratively regularized Gauß-Newton method [1, 10, 11, 12], the Levenberg-Marquardt method [6], a Newton-CG algorithm [7] or a Newton-Landweber method (cf. [9, 10]).

For the analysis of regularization methods for nonlinear problems certain restrictions on the nonlinearity of F are needed. Similar to [8, 9], we require the following nonlinearity conditions on F, which we assume to be Fréchet differentiable in a neighbourhood $\mathcal{B}_\rho(x^\dagger) := \{x \in \mathcal{X} : \|x - x^\dagger\| < \rho\}$ of a solution $x^\dagger$ to (1):

$$F'(x) = R(\bar{x}, x)F'(\bar{x}) + Q(\bar{x}, x), \tag{3}$$

with

$$\|I - R(\bar{x}, x)\| \leq C_R < 1 \tag{4}$$

and

$$\|Q(\bar{x}, x)\| \leq C_Q \|F'(\bar{x})(\bar{x} - x)\| \tag{5}$$

for all $x, \bar{x} \in \mathcal{B}_\rho(x^\dagger)$. Additionally, we assume,

$$\|F'(x)\| \leq 1, \quad x \in \mathcal{B}_\rho(x^\dagger), \tag{6}$$

which can always be achieved by a proper scaling.

It is well known (cf. [17]), that a rate of convergence can only be expected under an additional *source condition* on the difference of the *a priori* guess x_0 and the true solution $x^\dagger$, i.e., if there exist a $\mu > 0$ and $w \in \mathcal{N}(F'(x^\dagger)^\perp)$ such that

$$x^\dagger - x_0 = (F'(x^\dagger)^* F'(x^\dagger))^\mu w. \tag{7}$$

In case this difference is sufficiently smooth, i.e., (7) holds for some $\mu \geq 1/2$, only Lipschitz continuity of the Fréchet-derivative of F, i.e.,

$$\|F'(\bar{x}) - F'(x)\| \leq L \|\bar{x} - x\|, \tag{8}$$

for some $L > 0$ and all $x, \bar{x} \in \mathcal{B}_\rho(x^\dagger)$, is required instead of (3)-(5) for some of our results. For linear problems, the *source condition* (7) can even be shown to be necessary for a convergence rate of Hölder type (cf., e.g., [3]).

Iterative methods are turned into regularizing algorithms by stopping the iteration after an adequate number of steps. In contrast to *a priori* stopping criteria, which explicitly use information like (7) for determinig the stopping index, *a posteriori* stopping rules, e.g., the discrepancy principle (cf. [3, 15])

$$\|y^\delta - F(x^\delta_{k_*})\| < \tau\delta \leq \|y^\delta - F(x^\delta_k)\|, \quad 0 \leq k < k_*, \tag{9}$$

for some $\tau > 1$, allow to obtain optimal convergence rates without such information.

In this paper, we investigate a class of regularized Newton methods, i.e., iterations of the form

$$x_{n+1} = x_0 + g_{k_n}(A_n^* A_n) A_n^*(y - F(x_n) + A_n(x_n - x_0)), \quad A_n := F'(x_n). \tag{10}$$

where g_{k_n} is an appropriate regularization method used for the stable solution of the linearized equations

$$F'(x_n)(x_{n+1} - x_n) = y - F(x_n) \tag{11}$$

in each Newton step. We state and prove our results only for a class of accelerated Newton-Landweber methods, i.e., iterations (10), where g_{k_n} repectively $r_k(\lambda) = 1 - \lambda g_k(\lambda)$ belongs to a class of semiiterative regularization methods,

$$r_k(\lambda) = r_{k-1}(\lambda) + m_k(r_{k-1}(\lambda) - r_{k-2}(\lambda)) + w_k \lambda r_{k-1}(\lambda), \qquad k \geq 2 \tag{12}$$

with *optimal speed of convergence*, i.e.,

$$\omega_\mu(k) := \|\lambda^\mu r_k(\lambda)\|_{\mathcal{C}[0,1]} \leq c_\mu k^{-2\mu}, \quad \text{for} \quad 0 \leq \mu \leq \mu_0. \tag{13}$$

Nevertheless, the results extend natuarally to more general Newton-type methods, i.e., if g_{k_n} in (10) is replaced by an appropriate regularization method g_{α_n}, which will be pointed out at the end of Section 3.

The first result is concerned with convergence rates of accelerated Newton-Landweber iterations under *a priori* stopping, extending the results of [9] to the class of accelerated Newton-Landweber methods:

Theorem 1.1 *Let $x^\dagger$ denote a solution of* (3)*, x_0 such that*

$$x_0 - x^\dagger = (A^*A)^\mu w, \quad \textit{for some} \quad \mu > 0, \qquad A := F'(x^\dagger), \tag{14}$$

and $\|w\| \leq \omega$ with ω sufficiently small, in particular, $x_0 \in \mathcal{B}_\rho(x^\dagger)$ for some sufficiently small $\rho > 0$. Additionally, let y^δ be such that (2) *holds with δ sufficiently small, and x_n denote the iterates defined by* (10) *for a semiiterative method $\{r_k\}_{k\in\mathcal{N}}$ with* (13) *for some $\mu_0 \geq 1$, and let k_n satisfy*

$$1 \leq \frac{k_n}{k_{n-1}} \leq \beta \quad \textit{for} \quad n \in \mathcal{N}, \quad \textit{and} \quad \lim_{n\to\infty} k_n = \infty. \tag{15}$$

for $n \in \mathbb{N}$. If

(a) $\mu \leq 1/2$ and (3)*-*(5) *holds, with $Q = 0$ for $\mu < 1/2$, and $4\beta^{2\mu+1}C_R \leq 1$,*

or

(b) $\mu > 1/2$ *and* (8) *holds,*

and if

$$\underline{c}\left(\frac{\omega}{\delta}\right)^{\frac{1}{2\mu+1}} \leq k_{N(\delta)} \leq \overline{c}\left(\frac{\omega}{\delta}\right)^{\frac{1}{2\mu+1}},$$

for some $0 < \underline{c} \leq \overline{c} \leq c_\mu^{\frac{1}{2\mu+1}}$*, then*

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

The assertion of this Theorem follow immediately from Proposition 3.3.

The second main result of this paper is concerned with convergence rates under *a posteriori* stopping. In [10], optimal convergence rates

$$\|x_k^\delta - x^\dagger\| = O(\delta^{\frac{2\mu}{2\mu+1}}),$$

for Newton-type iterations have been shown for the range $0 < \mu \leq 1/2$, yielding a best possible rate of $O(\delta^{1/2})$. Including an additional lower bound on the number of iterations in a discrepancy like stopping rule, we are able to show that better rates than $O(\delta^{1/2})$ can be obtained for $\mu > 1/2$. Consider the following

Stopping Rule 1 *For given* $\tau > 1$*,* $\mu_{min} > 1/2$ *and* $\sigma > 0$*, let* $n_* = N(\delta, y^\delta)$ *be the smallest integer, such that*

$$k_{n_*}^{-(2\mu_{min}+1)} \leq \sigma\delta \tag{16}$$

and

$$\max\{\|y^\delta - F(x_{n_*})\|, \|y^\delta - F(x_{n_*-1})\|\} \leq \tau\delta. \tag{17}$$

Thus, the (outer) Newton iteration is stopped, when the first time two consecutive residuals are less than $\tau\delta$ and a minimal number of inner iterations has been performed. A similar criterion, but without (16) has been used in [10], to prove convergence rates for $\mu \leq 1/2$. For accelerated Newton-Landweber iterations (10), this stopping rule yields the following

Theorem 1.2 *Let the assumptions of Theorem 1.1 be satisfied, and the iteration* (10) *be stopped after* $n_* = N(\delta, y^\delta)$ *steps according to the Stopping Rule 1 with some* $\tau > 1$ *sufficiently large,* $\sigma > 0$ *sufficiently small, and some* $\mu_{min} \geq 1/2$*. Additionally, let* F *satisfy* (3)-(5) *with* $Q = 0$ *for* $0 < \mu < 1/2$*. Then*

$$k_{n_*} = O(\delta^{-\frac{1}{2\bar{\mu}+1}}) \tag{18}$$

with $\bar{\mu} = \min(\mu, \mu_{min})$ *and*

$$\|x_{N(\delta)^\delta} - x^\dagger\| = O(\delta^{f(\mu)}), \tag{19}$$

where $f(\mu)$ *is defined by*

$$f(\mu) = \begin{cases} \frac{2\mu}{2\mu+1}, & 0 < \mu \leq 1/2, \\ \frac{2\mu}{2\mu_{min}+1}, & 1/2 < \mu < 1, \\ \min(\frac{2\bar{\mu}}{2\mu_{min}+1}, \frac{2\mu_0-1}{2\mu_0}), & \mu \geq 1. \end{cases} \tag{20}$$

In the range $0 < \mu \leq 1/2$, these rates are optimal. The proof of this result and some remarks are given in Section 3. The effect of improved convergence rates for $\mu > 1/2$ is illustrated in numerical tests in Section 4.

The outline of the paper is as follows: In Section 2, we formulate in more detail the class of acclererated Newton-Landweber respectively accelerated Newton-Landweber methods, we have in mind, and recall some convergence results for semiiterative regularization methods for linear problems. Section 3 then presents the convergence rate analysis for the proposed Newton-type methods under the *a priori* and *a posteriori* stopping rules of Theorem 1.1 and 1.2. In Section 4, we verify our assumptions for several test examples, and present numerical tests confirming the theoretical results.

2 A class of regularized Newton methods

Newton's method for the solution of nonlinear problems (1) relies on the solution of linearized equations (11), which, are usually ill-posed, if (3) is, and thus some kind of regularization has to be used for their stable solution. Application of a Tikhonov-type regularization with an appropriate regularization parameter α_n, for instance, yields

$$[F'(x_n)^* F'(x_n) + \alpha_n I](x_{n+1} - x_n) = F'(x_n)^*(y - F(x_n)) + \alpha_n(x_0 - \hat{x}_n), \qquad (21)$$

which corresponds to the iteratively regularized Gauß-Newton ($\hat{x}_n = x_n$), respectively the Levenberg-Marquardt ($\hat{x}_n = x_0$) method (cf. [1, 6]).

Alternatively, (11) can be solved by an iterative regularization method (*inner iteration*), in which case, the Newton iteration takes the form (10). In this case, the regularization method for the linerized problems (11) is specified by the *iteration polynomials* g_{k_n}, and k_n is an appropriate stopping index for the n^{th} inner iteration (10). For

$$g_{k+1}(\lambda) = \sum_{j=0}^{k} (1 - \lambda)^j,$$

which are the iteration polynomials of Landweber iteration, (10) amounts to the Newton-Landweber iteration (cf. [9]). Since Landweber iteration is known to converge rather slow, it seems advantageaous to use faster semiiterative regularization methods for the solution of (11) (accelerated Landweber iterations (cf. [3, 5]). We will call this class of Newton-type iterations *accelerated Newton-Landweber* and investigate their stability and convergence behaviour in detail below.

2.1 Semiiterative regularization methods

Before we formulate and discuss the regularizing properties of the iterative algorithm (10), we recall some results of the convergence theory for semiiterative regularization method (12) (cf. [3, 5]), which will be needed later on. For this purpose, consider the linear equation

$$Ax = y^\delta,$$

with $A : \mathcal{X} \to \mathcal{Y}$. Let $g_k(\lambda)$ define a semiiterative method and $r_k(\lambda) := 1 - \lambda g_k(\lambda)$ be the corresponding residual polynomials. Then the *approximation error* $x_k - x^\dagger$ and the *propagated data error* $x_k - x_k^\delta$ have the form

$$x_k - x^\dagger = r_k(A^*A)(x_0 - x^\dagger) \qquad \text{and} \qquad x_k^\delta - x_k = g_k(A^*A)A^*(y^\delta - y), \tag{22}$$

where x_k, x_k^δ denote the iterates obtained for data y, y^δ, respectively. The important property, which lets a sequence of residual polynomials $\{r_k\}_{k\in\mathcal{N}}$ define a regularization method is

$$\omega_\mu(k) \le c_\mu k^{-\mu}, \qquad 0 \le \mu \le \mu_0, k \in \mathbb{N}. \tag{23}$$

A largest value μ_0 for which (23) holds is often called *qualification* of the regularization method under consideration. Landweber iteration, for instance, satisfies (23) for any $\mu > 0$, and thus has qualification $\mu_0 = +\infty$.

Especially attractive from a numerical point of view are algorithms, that satisfy the stronger estimate (13), which yield optimal rates of convergence with approximately the square root of iterations compared to, e.g., Landweber iteration. As shown in [5], such an estimate is the best possible (in terms of powers of k), which motivates the notion of *optimal speed of convergence* for such methods. A prominent example of semiiterative regularization methods with optimal speed of convergence are the $\nu-$methods by Brakhage (cf. [2, 5]), which are defined by (12) with $m_1 = 0$, $w_1 = (4\nu + 2)/(4\nu + 1)$, and

$$\begin{aligned} m_k &= \tfrac{(k-1)(2k-3)(2k+2\nu-1)}{(k+2\nu-1)(2k+4\nu-1)(2k+2\nu-3)}, \\ w_k &= 4\tfrac{(2k+2\nu-1)(k+\nu-1)}{(k+2\nu-1)(2k+2\nu-1)}, \qquad k > 1, \end{aligned} \tag{24}$$

The $\nu-$methods satisfy (13) for $\mu_0 = \nu$ and will also be used in our numerical tests in Section 4.

The following theorem summarizes the main convergence results for semiiterative regularization methods:

Theorem 2.1 (Theorem 6.11 in [3]) *Let $y \in \mathcal{R}(A)$, and let the residual polynomials r_k satisfy* (13) *for some $\mu_0 > 0$. Then the semiiterative method $\{r_k\}_{k\in\mathcal{N}}$ is a regularization method of optimal order for $T^\dagger y \in \mathcal{R}((T^*T)^\mu)$ with $0 < \mu \le \mu_0 - 1/2$ provided the iteration is stopped with $k_* = k_*(\delta, y^\delta)$ according to the discrepancy principle* (9) *with fixed $\tau > \sup_{k\in\mathcal{N}} \|r_k\|_{\mathcal{C}[0,1]}$. In this case we have $k_* = O(\delta^{-\frac{1}{2\mu+1}})$ and $\|x_k^\delta - x^\dagger\| = O(\delta^{\frac{2\mu}{2\mu+1}})$. The same rate holds for $0 < \mu \le \mu_0$, if the iteration is stopped according to the* a priori *rule $k_* = O(\delta^{-\frac{1}{2\mu+1}})$.*

An analogous result holds for iterative methods satisfying only (23), in which case one has $k_* = O(\delta^{-\frac{2}{2\mu+1}})$.

3 Convergence rate analysis

In [9], the convergence of some Newton-type methods under *a priori* stopping rules has been investigated. The analysis there explicitly relies on certain properties of the regularization methods used for the solution of the linearized equations (11), e.g.,

$$\|r_k(A^*A) - r_k(A_n^*A_n)\| \le C\|A - A_n\|,$$

for bounded linear operators A, A_n, which are not verified for the general class of semiiterative methods under consideration. Thus, in a first step, we will prove the corresponding (*a priori*) convergence rate results also for our class of accelerated Newton-Landweber methods.

3.1 A priori stopping

Our convergence analysis below is based on an interpretation of the source condition $x^\dagger - x_0 \in \mathcal{R}((A^*A)^\mu)$ in terms of $\mathcal{R}((A_n^*A_n)^\mu)$:

Lemma 3.1 *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 $0 < \mu \le 1/2$ and $w \in \mathcal{X}$ there exist positive constants $\underline{c}, \overline{c}$ and an element $v \in \mathcal{X}$ such that*

$$(A^*A)^\mu w = (B^*B)^\mu v, \tag{25}$$

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

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., e.g., [10] for details). ■

The following lemma is based on an integral representation of fractional powers of selfadjoint operators due to Krasnoselskii, see, e.g., [13]:

Lemma 3.2 *Let A, B be linear bounded operators between Hilbert spaces. Then for $\mu \ge 0$ we have*

$$\begin{aligned} &\|(A^*A)^\mu - (B^*B)^\mu\| \\ &\le c(\mu) \begin{cases} \|A - B\|^{2\mu}, & \mu < 1/2, \\ \|A - B\|\,[1 + \|A\| + \|B\| + |\ln(\|A - B\|)|]\,, & \mu = 1/2, \\ \|A - B\|\,(\|A\| + \|B\|)^\mu\,, & \mu > 1/2. \end{cases} \end{aligned} \tag{26}$$

Proof. First note that by scaling, we can always assume $\|A\|, \|B\| \le 1$. Additionally, for $A = B$, the assertion is always true, so we may assume $A \ne B$. Then observe that

$$\begin{aligned} &(A^*A + tI)^{-1} - (B^*B + tI)^{-1} \\ &\quad = -(A^*A + tI)^{-1}(A^*A - B^*B)(B^*B + tI)^{-1} \\ &\quad = -(A^*A + tI)^{-1}[A^*(A - B) + (A^* - B^*)B](B^*B + tI)^{-1}, \end{aligned}$$

which yields the following estimate

$$\begin{aligned} &\|(A^*A + tI)^{-1} - (B^*B + tI)^{-1}\| \\ &\quad \le \min\left(\frac{2\|A - B\|}{t^{3/2}}, \frac{\|A - B\|(\|A\| + \|B\|)}{t^2}, \frac{1}{t}\right). \end{aligned}$$

Next, for $\mu \in (0,1)$, we use the following identity (cf. [13]),

$$(A^*A)^\mu - (B^*B)^\mu = \frac{\sin \mu\pi}{\pi} \int_0^\infty t^\mu [A^*A + tI)^{-1} - (B^*B + tI)^{-1}] dt,$$

which yields

$$\begin{aligned} &\|(A^*A)^\mu - (B^*B)^\mu\| \\ &\leq \frac{\sin \mu\pi}{\pi} \left[\int_0^{t_1} t^{\mu-1} dt + 2\|A - B\| \int_{t_1}^{t_2} t^{\mu-3/2} dt + \|A - B\| (\|A\| + \|B\|) \int_{t_2}^\infty t^{\mu-2} dt \right] \end{aligned}$$

Now, the estimates for $0 < \mu < 1$ follows with the following setting: $t_1 \sim \|A - B\|^2$, $t_2 = \infty$ for $0 < \mu < 1/2$; $t_1 \sim \|A - B\|^2$, $t_2 \sim \|A - B\|$ for $1/2 \leq \mu < 1$. In the case $\mu = 1/2$, the logarithmic term in (26) is due to the second integral.

For $\mu = 1$, we have

$$\|A^*A - B^*B\| = \|A^*(A - B) + (A^* - B^*)B\| \leq \|A - B\| (\|A\| + \|B\|).$$

Similarly, for $1 < \mu \leq 2$, we have

$$(A^*A)^\mu - (B^*B)^\mu = (A^*A)^{\mu/2}[(A^*A)^{\mu/2} - (B^*B)^{\mu/2}] + [(A^*A)^{\mu/2} - (B^*B)^{\mu/2}](B^*B)^{\mu/2}$$

and thus with the above estimate for $\mu/2$,

$$\begin{aligned} &\|(A^*A)^\mu - (B^*B)^\mu\| \\ &\quad \leq c(\mu/2) \|A - B\| (\|(A^*A)^{\mu/2}\| + \|(B^*B)^{\mu/2}\|)(\|A\| + \|B\|)^{\mu/2}. \end{aligned}$$

The estimate for $\mu > 0$ then follows in the same manner. ■

We are now in the position to derive an estimate of the iteration error in terms of the number of inner iterations k_n, which we assume to grow not too rapidly, i.e., (15) holds.

Proposition 3.3 *Let $x^\dagger$ denote a solution of* (3), *with*

$$x_0 - x^\dagger = (A^*A)^\mu w, \quad \text{for some} \quad \mu > 0, \qquad A := F'(x^\dagger)$$

and $\|w\| \leq \omega$ with ω sufficiently small, in particular $x_0 \in \mathcal{B}_\rho(x^\dagger)$ for some sufficiently small $\rho > 0$. Additionally, let y^δ be such that (2) *holds with δ sufficiently small, and x_n denote the iterates defined by* (10) *for a semiiterative method $\{r_k\}_{k \in \mathcal{N}}$ with* (13) *for some $\mu_0 \geq 1$, and let k_n satisfy* (15) *for $n \in \mathbb{N}$. If*

(a) $\mu \leq 1/2$ and (3)-(5) *holds, with $Q = 0$ for $\mu < 1/2$, and $4\beta^{2\mu+1} C_R \leq 1$,*

or if

(b) $\mu > 1/2$ and (8) *holds,*

then there exists a constant C_μ independent of n and δ, such that

$$\|x_{n+1}^\delta - x^\dagger\| \leq C_\mu k_n^{-2\mu} \omega, \qquad 0 \leq n < N(\delta) \tag{27}$$

for $\mu \leq \mu_0$ and, with $A_{n+1} = F'(x_{n+1}^\delta)$,

$$\|A_{n+1}(x_{n+1}^\delta - x^\dagger)\| \leq C_\mu k_n^{-2\mu-1} \omega, \qquad 0 \leq n < N(\delta) \tag{28}$$

for $\mu \le \mu_0 - 1/2$, where $N(\delta)$ denotes the largest integer such that

$$k_n \le \left(c_\mu \frac{\omega}{\delta}\right)^{\frac{1}{2\mu+1}}, \tag{29}$$

for all $0 \le n \le N(\delta)$. Additionally, $x_n \in \mathcal{B}_\rho(x^\dagger)$ for $n \le N(\delta)$.

Proof. We start with the case $\mu < 1/2$ and prove the assertion by induction: for $n = 0$ the result follows if δ and ω are small enough, since $g_{k_0}(A_0^* A_0)$ is a continuous operator. Now let (27)-(29) hold for some $n > 0$, $C_\mu > 0$, and $x_n \in \mathcal{B}_\rho(x^\dagger)$. Then with the notation $e_n^\delta = x_n^\delta - x^\dagger$ and (10), we get the closed form representation

$$e_{n+1}^\delta = r_{k_n}(A_n^* A_n)(x_0 - x^\dagger) + g_{k_n}(A_n^* A_n) A_n^* (y^\delta - y + l_n), \tag{30}$$

with $l_n = \int_0^1 (F'(x^\dagger + t e_n^\delta) - F'(x_n^\delta)) e_n^\delta dt$. Together with w_n such that $(A^* A)^\mu w = (A_n^* A_n)^\mu w_n$, which exists by Lemma 3.1, the nonlinearity conditions (3), (4) yield

$$\begin{aligned}
\|e_{n+1}\| &\le \|r_{k_n}(A_n^* A_n)(x_0 - x^\dagger)\| + \|g_{k_n}(A_n^* A_n) A_n^* (y^\delta - y + l_n)\| \\
&\le \|r_{k_n}(A_n^* A_n)(A_n^* A_n)^\mu w_n\| + 2k_n\delta \\
&\qquad + \|g_k(A_n^* A_n) A_n^* \int_0^1 (R(x^\dagger + t e_n^\delta, x_n^\delta) - I) dt\| \, \|A_n e_n^\delta\| \\
&\le k_n^{-2\mu}\omega \left[c_\mu + 2C_R C_\mu \beta^{2\mu+1} + 2k_n^{1+2\mu}\delta/\omega\right].
\end{aligned}$$

Now, by assumption $4\beta^{2\mu+1} C_R \le 1$, and it suffices to require

$$c_\mu + 2k_n^{1+2\mu}\delta/\omega \le \frac{C_\mu}{2},$$

which holds for $C_\mu \ge 6c_\mu$ and $k_n \le \left(c_\mu \frac{\omega}{\delta}\right)^{\frac{1}{2\mu+1}}$. Furthermore, if ω is sufficiently small, $x_{n+1}^\delta \in \mathcal{B}_\rho(x^\dagger)$ follows by (27). In the same manner, one derives the estimate

$$\|A_n e_{n+1}\| \le \frac{C_\mu}{1 + C_R} k_n^{-2\mu-1}\omega,$$

under the additional condition

$$C_\mu \ge 4c_{\mu+1/2} + 8\frac{\delta}{\omega},$$

where we used $C_R \le \frac{1}{4}$. Finally, by (3) and (4), we have

$$\|A_{n+1} e_{n+1}^\delta\| \le (1 + C_R)\|A_n e_{n+1}^\delta\| \le C_\mu k_n^{-2\mu-1}\omega.$$

This yields (27), (28) for $0 < \mu < 1/2$.

The case $\mu = 1/2$ is treated in a similar way, using

$$(A^* A)^{1/2} w = A^* \tilde{w} = (A_n^* - Q(x^\dagger, x_n)^*)[R(x^\dagger, x_n)^*]^{-1}\tilde{w},$$

cf. (3)-(5).

We now turn to the case $\mu > 1/2$. Similarily as above, we get with (30), (8) and Lemma 3.2,

$$\begin{aligned}\|e_{n+1}^\delta\| &\leq \|r_{k_n}(A_n^*A_n)(A_n^*A_n)^\mu w\| + \|r_{k_n}(A_n^*A_n)[(A^*A)^\mu - (A_n^*A_n)^\mu]w\| \\ &\quad + \|g_{k_n}(A_n^*A_n)A_n^*(y^\delta - y + l_n)\| \\ &\leq c_\mu k_n^{-2\mu}\omega + 2C(\mu)L\|e_n\|\omega + 2k_n[\delta + L\|e_n\|^2] \\ &\leq k_n^{-2\mu}\omega\left[c_\mu + 2C(\mu)\beta^{2\mu}C_\mu L\omega + 2\beta^{4\mu}C_\mu^2 L k_n^{1-2\mu}\omega + 2k_n^{1+2\mu}\delta/\omega\right].\end{aligned}$$

For sufficiently large $C_\mu \geq 6c_\mu$ and sufficiently small ω, such that

$$4[C(\mu)\beta^{2\mu} + \beta^{4\mu}C_\mu k_n^{1-2\mu}]L\omega \leq 1,$$

this yields (27), and $x_n^\delta \in \mathcal{B}_\rho(x^\dagger)$ as in the case $\mu < 1/2$. Note, that $k_0 \leq \left(c_\mu \frac{\omega}{\delta}\right)^{\frac{1}{2\mu+1}}$ for δ sufficiently small. The estimate for $\|F'(x_{n+1})e_{n+1}^\delta\|$ is derived in the same way as above where we require (13) to hold for $\mu_0 \geq \mu + 1/2$ only for (28). ■

Remark 3.4 Proposition 3.3 immediately implies the convergence rate result of Theorem 1.1. An estimate for $\mu \geq 1/2$ under the nonlinearity condition (3) with $Q \neq 0$ can be proven under additional, restrictive assumptions on the regularization method $\{r_k\}_{k\in\mathcal{N}}$ for the linearized problem (cf. [10, Lemma 2.1]), which have not been verified for general semiiterative methods, and thus the results there are not applicable in our case.

Remark 3.5 Theorem 1.1 states optimal rates of convergence for $0 < \mu \leq \mu_0 - 1/2$, extending the corresponding results for the iteratively regularized Gauß-Newton method and the Newton-Landweber iteration in [9] to the class of accelerated Newton-Landweber methods considered in this paper. In [9], special properties of the applied regularization methods, e.g.,

$$\|[r_k(A_n^*A_n) - r_k(A^*A)](A^*A)^\mu\| \leq C\|A_n - A\|,$$

were used, which could not been verified for the general class of semiiterative algorithms under consideration. In principle, the above results also apply to more general regularization algorithms used for the solution of the linearized problems, in which case additional (standard) conditions on g_α have to be satisfied, (cf. [3]), which are automatically fulfilled for iterative methods by Markov's inequality. In particular, the above results hold with obvious modification also for iterative regularization methods satisfying only the weaker condition (23) instead of (13). Convergence for $\mu = 0$ under the weaker nonlinearity condition

$$\|F(x) - F(x^\dagger) - F'(x^\dagger)(x - x^\dagger)\| \leq \eta\|F(x) - F(x^\dagger)\|, \quad \eta < 1/2, \tag{31}$$

has been proven in [6]. If the increasing sequence of inner iterations k_n satisfies

$$k_n \sim \beta^n, \qquad n \geq 0, \tag{32}$$

with some $\beta > 1$, then the overall number of iterations is bounded by

$$k_* = \sum_{n=0}^{N(\delta)} k_n = O(\delta^{-\frac{1}{2\mu+1}}),$$

whereas for the Newton-Landweber iteration, one only has $k_* = O(\delta^{-\frac{2}{2\mu+1}})$, cf. [9].

3.2 A posteriori stopping

In [10], convergence rates under an appropriate *a posteriori* stopping rule are proven in case (14) holds for some $0 < \mu \leq 1/2$. The aim of this section is to show that better convergence rates than $O(\delta^{1/2})$ can be obtained for $\mu > 1/2$, if a lower bound (16) on the number of iterations is included in the stopping rule. The following Lemma guarantees stability of our class of Newton-type methods (10), when stopped according to the stopping rule 1:

Lemma 3.6 *Let the assumption of Proposition 3.3 be valid, and the iteration* (10) *(with r_k satisfying* (13) *for $\mu \leq \mu_0$) be stopped according to the stopping rule 1 with $1/2 < \mu_{min} \leq \mu_0 - 1/2$, $\tau > 1$ sufficiently large, and $\sigma > 0$ sufficiently small. Then, for $\mu \leq \mu_{min}$,* (29) *holds, i.e.,*

$$k_{n_*} \leq \left(c_\mu \frac{\omega}{\delta}\right)^{\frac{1}{2\mu+1}}. \tag{33}$$

Proof. The assertion follows from Proposition 3.3. ∎

In the convergence rate proof below, we utilize the following Lemma:

Lemma 3.7 *Let A, B, Q be bounded, linear operators between Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$ with $\|A\|, \|B\| \leq 1$, and R be a linear operator on $\mathcal{Y}$, such that*

$$B = RA + Q \qquad \text{and} \quad \|I - R\| < 1. \tag{34}$$

Then, for $\mu > 1/2$,

$$(A^*A)^\mu w = (B^*B)^\mu w_B + B^* P w_P + Q^* w_Q, \tag{35}$$

with

$$\|P\| \leq c\,(\|I - R\|^{\min(2\mu-1,1)} + \|Q\|),$$

and $\|w_B\|$, $\|w_P\|$, $\|w_Q\| \leq c\|w\|$.

Proof. We start with the estimate for $1/2 < \mu < 1$: Since $\mathcal{R}(A^*) = \mathcal{R}((A^*A)^{1/2})$, there exists a $\tilde{w}$ such that

$$(A^*A)^\mu w = (A^*A)^\nu A^* \tilde{w} = A^*(AA^*)^\nu \tilde{w}$$

with $\nu = \mu - 1/2$. Now rewrite

$$\begin{aligned} A^*(AA^*)^\nu &= B^*(R^{-1})^*(AA^*)^\nu - Q^*(R^{-1})^*(AA^*)^\nu \\ &= B^*(BB^*)^\nu + B^*((R^{-1})^* - I)(BB^*)^\nu \\ &\qquad + B^*(R^{-1})^*[(AA^*)^\nu - (BB^*)^\nu] - Q^*(R^{-1})^*(AA^*)^\nu \end{aligned}$$

The first assertion now follows with $P = ((R^{-1})^* - I)(BB^*)^\nu + (R^{-1})^*[(AA^*)^\nu - (BB^*)^\nu]$ and Lemma 3.2.

Next consider the case $\mu = 1$, where we have

$$\begin{aligned} A^*A &= B^*B + B^*[((R^{-1})^*R^{-1} - I)B - (R^{-1})^*R^{-1}Q] \\ &\qquad - Q^*((R^{-1})^*R^{-1}B - (R^{-1})^*R^{-1}Q), \end{aligned}$$

which yields the assertion with (3). The case $\mu \in \mathbb{N}$ can be treated similarily, using the expansion

$$(B^*B)^\mu - (A^*A)^\mu = \sum_{j=1}^{\mu} (B^*B)^{\mu-j}(B^*B - A^*A)(A^*A)^{j-1}.$$

Finally, for $n < \mu < n+1$, we use the decomposition

$$(A^*A)^\mu = (A^*A)^n (A^*A)^{\mu-n},$$

and proceed as in the case $1/2 < \mu < 1$. ■

We are now in the position to prove Theorem 1.2:

Proof of Theorem 1.2. We start with the case $0 < \mu \leq 1/2$ and $Q = 0$: Observe, that for $n = n_*$ and $n = n_* - 1$, by (2) and (3),(4),

$$\begin{aligned} \|A_n e_n^\delta\| &= \|y - F(x_n^\delta) - \int_0^1 [F'(x_n^\delta + te_n^\delta) - A_n] e_n^\delta dt\| \\ &\leq \delta + \|F(x_n^\delta) - y^\delta\| + C_R \|A_n e_n^\delta\| \end{aligned}$$

holds, and hence, with $C_R < 1$ and (17),

$$\|A_n e_n^\delta\| \leq C\delta, \qquad \text{for} \quad n \in \{n_* - 1, n_*\}. \tag{36}$$

Next, by (10), and denoting $n = n_* - 1$, we have

$$\begin{aligned} A_n e_{n^*}^\delta &= A_n(x_0 - x^\dagger) + A_n g_{k_n}(A_n^* A_n) A_n^* [y^\delta - F(x_n^\delta) + A_n(x_n^\delta - x_0)] \\ &= A_n r_{k_n}(A_n^* A_n)(x_0 - x^\dagger) + A_n A_n^* g_{k_n}(A_n A_n^*)[y^\delta - F(x_n^\delta) + A_n(x_n^\delta - x^\dagger)] \end{aligned}$$

and thus with (3),(4), (36) and (17),

$$\begin{aligned} \|A_n r_{k_n}(A_n^* A_n)(x_0 - x^\dagger)\| &= \|A_n e_{n_*}^\delta - A_n g_{k_n}(A_n^* A_n) A_n^* [y^\delta - F(x_n^\delta) + A_n(x_n^\delta - x^\dagger)]\| \\ &\leq (1 + C_R)\|A_n e_{n_*}^\delta\| + \|y^\delta - F(x_n^\delta)\| + \|A_n e_n^\delta\| \\ &\leq C\delta. \end{aligned}$$

Finally, the error can be estimated as follows:

$$\begin{aligned} \|x_{n_*}^\delta - x^\dagger\| &\leq \|r_{k_n}(A_n^* A_n)(A^*A)^\mu w\| + \|g_{k_n}(A_n^* A_n) A_n^* (y^\delta - y - l_n)\| \\ &\leq \|r_{k_n}(A_n^* A_n)(A_n^* A_n)^\mu w_n\| + 2k_n(\delta + cC_R\|A_n e_n^\delta\|) \\ &\leq \|r_{k_n}(A_n^* A_n)(A_n^* A_n)^\mu w_n\| + ck_n\delta \end{aligned}$$

Now, the interpolation inequality and (20) yield

$$\begin{aligned} \|r_{k_n}(A_n^* A_n)(A_n^* A_n)^\mu w_n\| &\leq 2\|A_n r_{k_n}(A_n^* A_n)(A_n^* A_n)^\mu w_n\|^{\frac{2\mu}{2\mu+1}} \|w_n\|^{\frac{1}{2\mu+1}} \\ &\leq c\delta^{\frac{2\mu}{2\mu+1}} \omega^{\frac{1}{2\mu+1}}, \end{aligned}$$

which completes the proof for the case $\mu \leq 1/2$, since $k_{n^*} = O(\delta^{-\frac{1}{2\mu+1}})$ (cf. Lemma 3.6).

The case $\mu = 1/2$ and $Q \neq 0$, follows by with $\|Q(x_n^\delta, x^\dagger)\| \leq C_Q \|A_n e_n^\delta\|$ and minor modifications.

For $1/2 < \mu < 1$, Lemma 3.7 with B replaced by A_n yields together with $\|x_k^\delta - x^\dagger\| \leq C\delta^{1/2}$ (see above)

$$\begin{aligned} \|r_{k_n}(A_n^* A_n)(A^* A)^\mu w\| &\leq c\|r_{k_n}(A_n^* A_n)(A_n^* A_n)^\mu\| \, \|w\| \\ &\quad + (\|r_{k_n}(A_n^* A_n) A_n^* P\| + \|Q\|)\, \|w\| \\ &\leq C[k_n^{-2\mu} + k_n^{-1}\delta^{\frac{2\mu-1}{2}} + \delta] \, \|w\| . \end{aligned} \tag{37}$$

By Lemma 3.6, and the lower bound on the number of iterations, we have

$$c\delta^{-\frac{1}{2\mu_{min}+1}} \leq k_n \leq C\delta^{-\frac{1}{2\mu+1}},$$

and hence

$$\|x_k^\delta - x^\dagger\| \leq C\,\|w\|\,(k_n^{-2\mu} + k_n\delta) \leq \delta^{\min(\frac{2\mu}{2\mu_{min}+1}, \frac{2\mu}{2\mu+1})}.$$

For $\mu \geq 1$, the last estimate in (37) is replaced by

$$\leq \quad C[k_n^{-2\mu} + k_n^{-1}\delta^{1/2} + \delta] \, \|w\| .$$

The rest follows similarly as above. Note that Proposition 3.3 can only be applied for $\mu \leq \mu_0 - 1/2$, since we implicitly used the estimate (28) to bound the number of iterations. ■

Remark 3.8 The number μ_{min} in (16) determines the minimal number of inner iterations, which have to be performed before the iteration may be stopped by the discrepancy principle. The larger μ_{min} is, the lower the corresponding bound in (16) on the minimal number of inner iterations. Depending on the specific choice of μ_{min}, a second range of values of μ exists, for which improved convergence can be guaranteed, e.g., with $\mu_{min} = 1$ and $\mu_0 = 3/2$, Theorem 1.2 guarantees optimal convergence rates for $\mu \in (0, 1/2] \cap \{1\}$ and improved convergence for $\mu \in (3/4, 1)$ (cf. Figure 1).

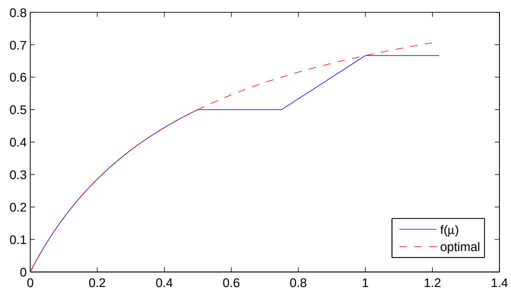

Figure 1: Convergence behaviour guaranteed by Theorem 1.2 for $\mu_{min} = 1$ and $\mu_0 \geq 3/2$.

For $\mu_{min} = 7/4$, the lower bound of iterations, which have to be performed before stopping may occur is $k_{N(\delta)} \sim \delta^{-\frac{2}{9}}$, which is less than the number of iterations one would expect by Theorem 1.1. Note, that the nonlinearity conditions in Theorem 1.2 are stronger than the ones for the *a priori* result of Theorem 1.1 in casse $\mu > 1/2$.

Remark 3.9 Like the results above, Theorem 1.2 with k_n replaced by α_n generalizes to arbitrary regularization methods $\{g_\alpha\}_{\alpha>0}$, as long as the usual conditions (cf. [3]) are satisfied. Note, that the iteratively regularized Gauß-Newton method exhibits saturation at $\mu = 1/2$, which is due to the finite qualification of Tikhonov regularization used for the solution of the linearized problems. Hence, in view of Theorem 1.2, semiiterative methods for the solution of (11) with sufficiently high qualification will show improved convergence in comparison with the iteratively regularized Gausß-Newton method for $\mu > 1/2$ (cf. the numerical results in Section 4).

The rates of Theorem 1.2 can also be derived for iterations $\{r_n\}$ only satisfying the weaker estimate (23) instead of (13). There, however, the estimate (18) for the number of iterations has to be replaced by $k_{N(\delta)} = O(\delta^{-\frac{2}{2\mu+1}})$, which shows that, as for linear problems, accelerated Newton-Landweber methods should yield a significant speed-up in comparison to, e.g., the Newton-Landweber method.

4 Some examples and numerical tests

In this section, we give some examples and verify the assumptions made in the convergence rate results of the previous section. Additionally, we present numerical tests confirming the theoretical results including a comparison of an accelerated Newton-Landweber iteration (the linearized equations are solved by a $\nu-$method with $\nu = 2$, with the iteratively regularized Gauß-Newton method and the Newton-Landweber iteration (cf. [9]).

In the numerical tests we use the following sequence of iteration numbers respectively regularization parameters

$$k_n = k_0\beta^n, \quad \text{and} \quad \alpha_n = \alpha_0/\beta^n. \tag{38}$$

The equations are discretized by piecewise linear finite elements (both, the parameter and state variables). In order to avoid inverse crimes, the data are calculated on a finer grid and random noise is added.

We start with a nonlinear integral equation and then turn to the investigation of certain parameter identification problems:

Example 4.1 *A nonlinear Hammerstein integral equation, cf. [16].*
Let $F : H^1[0,1] \to L^2[0,1]$ be defined by

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

with Fréchet derivative given by

$$(F'(x)h)(s) = 2\int_0^s x(t)h(t)dt.$$

If $x^\dagger \geq \gamma > 0$, this yields

$$(F'(x)h)(s) = \int_0^s \frac{x(t)}{x^\dagger(t)}[F'(x^\dagger)(h)]'(t)dt. \tag{39}$$

Since $x \in H^1[0,1]$, we have $\bar{x} \geq \bar{\gamma} > 0$ for $\bar{x} \in \mathcal{B}_\rho(x^\dagger)$ with ρ sufficiently small, and thus $x^\dagger$ can be replaced by $\bar{x}$ in (39) yielding assumptions (3)-(5) with $Q = 0$ and

$$(R(\bar{x}, x)v)(s) = \frac{x(s)}{\bar{x}(s)} v(s) - \int_0^s [\frac{x(t)}{\bar{x}(t)}]' v(t) dt.$$

This in turn yields $\|R(\bar{x}, x) - I\| \leq c\|x - \bar{x}\|$.

In a first numerical test, we try to identify

$$x^\dagger = 3/2 - |\mathrm{erf}(4(t - 1/2))|, \tag{40}$$

from a starting guess $x_0 = 1/2$. We choose $\beta = 2$, $k_0 = 5$ and $\alpha_0 = 0.1$. The iterations are stopped according to (17) with $\tau = 1.2$. The results or the numerical test are listed in Table 1.

$\delta/\|y\|$	n^*_{accNLW}	err_{accNLW}	n^*_{IRGN}	err_{IRGN}	n^*_{NLW}	err_{NLW}
0.08	6	0.6025	7	0.6864	9	0.6616
0.04	7	0.4065	9	0.5186	11	0.4904
0.02	7	0.4429	10	0.4807	12	0.4557
0.01	8	0.3577	12	0.3761	13	0.4022
0.005	9	0.2901	13	0.3382	15	0.3289

Table 1: Iteration numbers n^* and error $\|x^\delta_{n^*} - x^\dagger\|$ for the accelerated Newton-Landweber (accNLW), the iteratively-regularized Gauß-Newton (IRGN) and the Newton-Landweber method (NLW) applied to Example 4.1 and (40).

As expected, the number of Newton iterations grows logarithmically with decreasing δ. The three iterations yield very similar results and a convergence rate of approximately $\delta^{0.25}$.

In a second test, we consider the reconstruction of a smooth solution

$$x^\dagger = t + 10^{-6}(196145 - 41286t^2 + 19775t^4 + 70t^6 + 436t^7), \tag{41}$$

from a starting value $x_0 = t$. It was shown in [16] in this case (7) holds with $\mu = 3/2$ and thus a rate of $\delta^{3/4}$ is optimal whereas only a rate of $O(\delta^{1/2})$ can be expected for the Gauß-Newton method. The results in Table 2 were obtained by stopping the iteration according to (17) without the additional lower bound on the iteration number. The corresponding convergence rates are approximately $O(\delta^{0.5})$ for all methods. For the numerical results listed in Table 3, an additional bound on the lower number of iterations (cf. (17)) has been used, i.e., $n_* \geq c_1 + c_2 * N$, where N denotes the index of the noise level, i.e., $\delta \sim 0.5^N$. Here we chose $c_1 = 2$ for all iterations and $c_2 = 1$ for the accelerated Newton-Landweber and the iteratively regularized Gauß-Newton method, and $c_2 = 2$ for the Newton-Landweber iteration. The results correspond to rates $\delta^{0.77}$, and $\delta^{0.78}$ for the accelerated Newton-Landweber and the Newton-Landweber method, and $\delta^{0.48}$ for the Gauß-Newton method.

$\delta/\|y\|$	n^*_{accNLW}	err_{accNLW}	n^*_{IRGN}	err_{IRGN}	n^*_{NLW}	err_{NLW}
0.02	2	0.10972	2	0.15824	2	0.11807
0.01	2	0.10829	4	0.10133	3	0.10956
0.005	4	0.03474	5	0.06884	5	0.06569
0.0025	4	0.03759	6	0.05006	6	0.04312
0.00125	4	0.03874	7	0.04019	7	0.03442

Table 2: Iteration numbers n^* and error $\|x^\delta_{n^*} - x^\dagger\|$ for the accelerated Newton-Landweber (accNLW), the iteratively-regularized Gauß-Newton (IRGN) and the Newton-Landweber method (NLW) applied to Example 4.1 and (41).

$\delta/\|y\|$	n^*_{accNLW}	err_{accNLW}	n^*_{IRGN}	err_{IRGN}	n^*_{NLW}	err_{NLW}
0.02	2	0.10522	2	0.15476	2	0.11507
0.01	3	0.07800	4	0.09954	4	0.09009
0.005	4	0.03781	5	0.06814	6	0.04243
0.0025	5	0.02050	6	0.04994	8	0.02376
0.00125	6	0.01314	7	0.04114	10	0.01484

Table 3: Iteration numbers n^* and error $\|x^\delta_{n^*} - x^\dagger\|$ for the accelerated Newton-Landweber (accNLW), the iteratively-regularized Gauß-Newton (IRGN) and the Newton-Landweber method (NLW) applied to Example 4.1bb.

Example 4.2 *Parameter identification 1, cf. [8]*
In this example we try to identify the parameter c in the elliptic equation

$$\begin{aligned} -\Delta u + cu &= f \quad \text{in} \quad \Omega, \\ u &= g \quad \text{in} \quad \partial\Omega, \end{aligned} \tag{42}$$

from distributed measurements of the state u.

We assume Ω to be an interval 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

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

where $u(c)$ denotes the solution of (42) and $\gamma > 0$ has to be sufficiently small. In 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 $A(c)u = -\Delta u + cu$. If $u(c^\dagger) \ge \kappa > 0$ a.e. in Ω, then for all c with $\|c - c^\dagger\| \le \rho \le \gamma$ (see [8] for details)

$$F'(c)^* = F'(c^\dagger) R_c(c^\dagger), \tag{43}$$

with

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

The estimate is again valid for $\bar{c} \in \mathcal{B}_\rho(c^\dagger)$, since by continuity of the parameter to solution map between spaces L^2 and $H^2 \cap H^1_0$ we have $u(\bar{c}) \ge \kappa_\rho$ for some $\kappa_\rho > 0$

as long as $\rho > 0$ is small enough.

For a numerical test, we consider the two dimensional case $\Omega = [0,1]^2$, set $g = 1$, $f = 1$, $\alpha_0 = 0.1$, $k_0 = 5$, and try to identify

$$c^\dagger = 1 + \text{sign}(x - 1/2)\text{sign}(y - 1/2)$$

from the initial guess $c_0 = 0$. By (43) we have $\mathcal{R}(F'(c^\dagger)^*) \subset \mathcal{H}_0^1(\Omega) \cap H^2(\Omega)$. Thus, $c^\dagger - c_0 \notin \mathcal{R}((F'(c^\dagger)^* F'(c^\dagger))^\mu)$ for any $\mu > 1/8$, and at most a convergence rate of $O(\delta^{1/5})$ can be expected. The results of the numerical reconstruction are listed in Table 4. The corresponding convergence rates lie between $\delta^{0.24}$ and $\delta^{0.22}$ for all

$\delta/\|u\|$	n^*_{accNLW}	err_{accNLW}	n^*_{IRGN}	err_{IRGN}	n^*_{NLW}	err_{NLW}
0.08	1	1.0934	2	0.9866	2	1.0745
0.04	3	0.7191	4	0.7111	5	0.7234
0.02	4	0.6107	5	0.6369	6	0.6440
0.01	4	0.6051	6	0.5857	7	0.5959
0.005	5	0.5166	8	0.4932	9	0.5036

Table 4: Iteration numbers n^* and error $\|x^\delta_{n^*} - x^\dagger\|$ for the accelerated Newton-Landweber (accNLW), the iteratively-regularized Gauß-Newton (IRGN) and the Newton-Landweber method (NLW) applied to Example 4.2.

methods.

Example 4.3 *Parameter identification 2.*
We study the identification of a diffusion coefficient a in

$$-\nabla \cdot (a \nabla u) = f, \qquad u|_{\partial\Omega} = 0, \tag{44}$$

from distributed measurements of the state u. The operator $F : \mathcal{K} \subset \mathcal{H}^1(\Omega) \to \mathcal{L}_2(\Omega)$ is defined by the parameter-to-solution mapping $F(a) := u(a)$, where $u(a)$ denotes the solution of (44). Let $A(a)$ with $\mathcal{D}(A(a)) = \mathcal{H}^2(\Omega) \cap \mathcal{H}^2(\Omega) \subset \mathcal{L}_2(\Omega)$ be defined by $A(a)u = -\nabla \cdot (a\nabla u)$, then $F(a) = A(a)^{-1} f$ and

$$\begin{aligned} F'(a)h &= -A(a)^{-1}A(h)F(a) \\ &= -A(a)^{-1}A(h)F(b) + A(a)^{-1}A(h)[F(b) - F(a)] \\ &= A(a)^{-1}A(b)F'(b)h + A(a)^{-1}A(h)[F(b) - F(a)] \\ &= R(a,b)F'(b)h + Q(a,b)h, \end{aligned}$$

which shows (3). Additionally, we have $\|R(a,b) - I\| \le c\|a - b\|$ and $\|Q\| \le \|F(a) - F(b)\|$. It was shown in [8], that $F(a)$ satisfies the nonlinearity condition (31), which yields

$$\|F(a) - F(b)\| \sim \|F'(a)(a-b)\| \sim \|F'(b)(a-b)\|,$$

and thus (5) holds.

As a numerical test, we try to reconstruct

$$q^\dagger = 1 + 0.5\sin(\pi x)\sin(2\pi y),$$

$\delta/\|u\|$	n^*_{accNLW}	err_{accNLW}	n^*_{IRGN}	err_{IRGN}	n^*_{NLW}	err_{NLW}
0.08	1	0.2328	1	0.2340	1	0.2328
0.04	1	0.2319	1	0.2333	1	0.2320
0.02	4	0.1312	5	0.1442	6	0.1395
0.01	5	0.0626	7	0.0847	8	0.0757
0.005	6	0.0307	8	0.0572	9	0.0506

Table 5: Iteration numbers n^* and error $\|x^\delta_{n^*} - x^\dagger\|$ for the accelerated Newton-Landweber (accNLW), the iteratively-regularized Gauß-Newton (IRGN) and the Newton-Landweber method (NLW) applied to Example 4.3.

from the initial guess $q_0 = 0.5$. As in the previous example, we choose $\beta = 2$ and $\tau = 1.5$. The results of the numerical reconstruction are listed in Table 5. The corresponding convergence rates are $O(\delta^{0.77})$ and $O(\delta^{0.60})$ for the accelerated Newton-Landweber and Newton-Landweber iteration, while the rate for the iteratively-regularized Gauß-Newton method is only $O(\delta^{0.55})$. Like in Example 4.1, the Newton-Iterative methods exhibit improved convergence when stopped according to the modified discrepancy principle (17).

5 Conclusions

A class of Newton-type iterations has been analyzed in the framework of regularization. The results of this paper contribute to the convergence analysis of Newton-type methods for the solution of ill-posed problems in two ways: first, the convergence theory of Newton-type iterations (cf., e.g., [9]) are extended to the proposed class of accelerated Newton-Landweber iterations. Secondly, improved convergence rates can be achieved when the iteration is stopped according to a modified discrepancy principle, where the minimal number of iterations is bounded from below. As shown in our numerical tests, the proposed accelerated Newton-Landweber method may yield faster convergence than the iteratively-regularized Gauß-Newton method, while needing much fewer inner iterations than the Newton-Landweber method.

Some open question in our convergence analysis are, if the convergence rates (cf. Theorem 1.2 and Remark 3.8) are optimal under the given assumptions, in particular, if the rates $O(\delta^{\frac{2\mu}{2\mu+1}})$ can also be established for the range $\mu \in (1/2, \mu_{min})$, and/or if the Lipschitz condition (8) suffices to obtain the convergence rates in case $\mu > 1/2$.

References

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

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

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

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

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

[6] M. Hanke. A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems. *Inverse problems*, 13:79–95, 1997.

[7] M. Hanke. Regularizing properties of a truncated Newton-CG algorithm for nonlinear inverse problems. *Numer. funct. anal. optim.*, 18:165–183, 1997.

[8] 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.

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

[10] 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.

[11] B. Kaltenbacher, A. Neubauer, and A.G. Ramm. Convergence rates of the continuous regularized Gauss-Newton method. *J. Inv. Ill-Posed Problems*, 10:261–280, 2002.

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

[13] M. A. Krasnoselskii, P. P. Zabreiko, E. I. Pustylnik, and P. E. Sbolevskii. *Integral Operators in spaces of Summable Functions*. Nordhoff International Publishing, Leyden, 1976.

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

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

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

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