Inexact Newton regularization methods in Hilbert scales

Qinian Jin · Ulrich Tautenhahn

Abstract We consider a class of inexact Newton regularization methods for solving nonlinear inverse problems in Hilbert scales. Under certain conditions we obtain the order optimal convergence rate result.

1 Introduction

In this paper we consider the nonlinear inverse problems

\[ F(x) = y, \]

where \( F : D(F) \subset X \mapsto Y \) is a nonlinear Fréchet differentiable operator between two Hilbert spaces \( X \) and \( Y \) whose norms and inner products are denoted as \( \| \cdot \| \) and \( (\cdot, \cdot) \) respectively. We assume that \( (1.1) \) has a solution \( x^\dagger \) in the domain \( D(F) \) of \( F \), i.e. \( F(x^\dagger) = y \). We use \( F'(x) \) to denote the Fréchet derivative of \( F \) at \( x \in D(F) \) and \( F'(x)^* \) the adjoint of \( F'(x) \). A characteristic property of such problems is their ill-posedness in the sense that their solutions do not depend continuously on the data. Let \( y^\delta \) be the only available approximation of \( y \) satisfying

\[ \| y^\delta - y \| \leq \delta \]

with a given small noise level \( \delta > 0 \). Due to the ill-posedness, the regularization techniques should be employed to produce from \( y^\delta \) a stable approximate solution of \( (1.1) \).

Many regularization methods have been considered in the last two decades. In particular, the nonlinear Landweber iteration \([6]\), the Levenberg-Marquardt method \([4,9]\), and the exponential Euler iteration \([7]\) have been applied to solve nonlinear inverse problems. These methods take the form

\[ x_{n+1} = x_n - g_{\alpha_n} \left( F'(x_n)^* F'(x_n) \right) F'(x_n)^* \left( F(x_n) - y^\delta \right), \]

Qinian Jin
Department of Mathematics, Virginia Tech, Blacksburg, VA 24060, USA
E-mail: qnjin@math.vt.edu

Ulrich Tautenhahn
Department of Mathematics, University of Applied Sciences Zittau/Görlitz, PO Box 1454, 02754 Zittau, Germany
E-mail: u.tautenhahn@hs-zigr.de
where $x_0$ is an initial guess of $x^\dagger$, $\{\alpha_n\}$ is a sequence of positive numbers, and $\{g_\alpha\}$ is a family of spectral filter functions. The scheme (1.3) can be derived by applying the linear regularization method defined by $\{g_\alpha\}$ to the equation

$$F'(x_n)(x - x_n) = y^\delta - F(x_n).$$

(1.4)

which follows from (1.1) by replacing $y$ by $y^\delta$ and $F(x)$ by its linearization $F(x_n) + F'(x_n)(x - x_n)$ at $x_n$. It is easy to see that

$$F(x_n) - y^\delta + F'(x_n)(x_{n+1} - x_n) = r_\alpha_n(F'(x_n)F'(x_n)^*)(F(x_n) - y^\delta),$$

where

$$r_\alpha(\lambda) = 1 - \lambda g_\alpha(\lambda)$$

(1.5)

which is called the residual function associated with $g_\alpha$. For well-posed problems where $F'(x_n)$ is invertible, usually one has $\|r_\alpha_n(F'(x_n)F'(x_n)^*)\| \leq \mu_n < 1$ and consequently

$$\|F(x_n) - y^\delta + F'(x_n)(x_{n+1} - x_n)\| \leq \mu_n\|F(x_n) - y^\delta\|.$$  

(1.6)

Thus the methods belong to the class of inexact Newton methods [2]. For ill-posed problems, however, there only holds $\|r_\alpha_n(F'(x_n)F'(x_n)^*)\| \leq 1$ in general. In [4] the Levenberg-Marquardt scheme was considered with $\{\alpha_n\}$ chosen adaptively so that (1.0) holds and the discrepancy principle was used to terminate the iteration. The order optimal convergence rates were derived recently in [5]. The general methods (1.3) with $\{\alpha_n\}$ chosen adaptively to satisfy (1.0) were considered later in [14] and the exponential Euler method in [1] for instance.

In this paper we will consider the inexact Newton methods in Hilbert scales which are more general than (1.3). Let $L$ be a densely defined self-adjoint strictly positive linear operator in $X$. For each $r \in \mathbb{R}$, we define $X_r$ to be the completion of $\cap_{k=0}^{\infty} D(L^k)$ with respect to the Hilbert space norm

$$\|x\|_r := \|L^r x\|.$$  

This family of Hilbert spaces $(X_r)_{r \in \mathbb{R}}$ is called the Hilbert scales generated by $L$. Let $x_0 \in D(F)$ be an initial guess of $x^\dagger$. The inexact Newton method in Hilbert scales defines the iterates $\{x_n\}$ by

$$x_{n+1} = x_n - g_{\alpha_n}(L^{-2s}F'(x_n)^*F'(x_n)) L^{-2s}F'(x_n)^*(F(x_n) - y^\delta),$$  

(1.7)

where $s \in \mathbb{R}$ is a given number to be specified later, and $\{\alpha_n\}$ is an a priori given sequence of positive numbers with suitable properties. We will terminate the iteration by the discrepancy principle

$$\|F(x_n) - y^\delta\| \leq \tau \delta < \|F(x_n) - y^\delta\|, \quad 0 \leq n < n_\delta$$  

(1.8)

with a given number $\tau > 1$ and consider the approximation property of $x_{n_\delta}$ to $x^\dagger$ as $\delta \to 0$. We will establish for a large class of spectral filter functions $\{g_\alpha\}$ the order optimal convergence rates for the method defined by (1.7) and (1.8).

Regularization in Hilbert scales has been introduced in [12] for the linear Tikhonov regularization with the major aim to prevent the saturation effect. Such technique has been extended in various ways, in particular, a general class of regularization methods in Hilbert scales has been considered in [15] with the regularization parameter chosen by the Morozov’s discrepancy principle. Regularization in Hilbert scales have
also been applied for solving nonlinear ill-posed problems. The nonlinear Tikhonov regularization in Hilbert scales has been considered in [10,3], a general continuous regularization scheme for nonlinear problems in Hilbert scales has been considered in [17], the general iteratively regularized Gauss-Newton methods in Hilbert scales has been considered in [8], and the nonlinear Landweber iteration in Hilbert scales has been considered in [13].

This paper is organized as follows. In Section 2 we first briefly review the relevant properties of Hilbert scales, and then formulate the necessary condition on \( \{a_n\} \), \( \{g_{\alpha}\} \) and \( F \) together with some crucial consequences. In Section 3 we obtain the main result concerning the order optimal convergence property of the method given by (1.7) and (1.8). Finally we present in Section 4 several examples of the method (1.2) for which \( \{g_{\alpha}\} \) satisfies the technical conditions in Section 2.

2 Assumptions

We first briefly review the relevant properties of the Hilbert scales \( (X_r)_{r \in \mathbb{R}} \) generated by a densely defined self-adjoint strictly positive linear operator \( L \) in \( X \), see [3]. It is well known that \( X_r \) is densely and continuously embedded into \( X_q \) for any \( -\infty < q < r < \infty \), i.e.

\[
\|x\|_q \leq \theta^{r-q}\|x\|_r, \quad x \in X_r,
\]

where \( \theta > 0 \) is a constant such that

\[
\|x\|^2 \leq \theta(Lx, x), \quad x \in D(L).
\]

Moreover there holds the important interpolation inequality, i.e. for any \( -\infty < p < q < r < \infty \) there holds for any \( x \in X_r \) that

\[
\|x\|_q \leq \|x\|^{\frac{p-q}{p}}\|x\|^{\frac{r-q}{r}}.
\]

Let \( T : X \mapsto Y \) be a bounded linear operator satisfying

\[
m\|h\|_{-a} \leq \|Th\| \leq M\|h\|_{-a}, \quad h \in X
\]

for some constants \( M \geq m > 0 \) and \( a \geq 0 \). Then the operator \( A := TL^{-s} : X \mapsto Y \) is bounded for \( s \geq a \) and the adjoint of \( A \) is given by \( A^* = L^{-s}T^* \), where \( T^* : Y \mapsto X \) is the adjoint of \( T \). Moreover, for any \( |\nu| \leq 1 \) there hold

\[
R((A^*A)^{\nu/2}) = X_{\nu(a+s)}
\]

and

\[
\mathcal{L}(\nu)\|h\|_{-\nu(a+s)} \leq \|(A^*A)^{\nu/2}h\| \leq \mathcal{C}(\nu)\|h\|_{-\nu(a+s)}
\]

on \( D((A^*A)^{\nu/2}) \), where

\[
\mathcal{L}(\nu) := \min\{m^\nu, M^\nu\} \quad \text{and} \quad \mathcal{C}(\nu) = \max\{m^\nu, M^\nu\}.
\]

If \( g : [0, \|A\|^2] \mapsto \mathbb{R} \) is a continuous function, then

\[
g(A^*A)L^s = L^s g(L^{-2s}T^*T).
\]

In order to carry out the convergence analysis on the method defined by (1.7) and (1.8), we need to impose some suitable conditions on \( \{a_n\} \), \( \{g_{\alpha}\} \) and \( F \). For the sequence \( \{a_n\} \) of positive numbers, we set

\[
s_{-1} = 0, \quad s_n := \sum_{j=0}^{n} \frac{1}{\alpha_j}, \quad n = 0, 1, \ldots.
\]
We will assume that there are constants $c_0 > 1$ and $c_1 > 0$ such that
\[
\lim_{n \to \infty} s_n = \infty, \quad s_{n+1} \leq c_0 s_n \quad \text{and} \quad 0 < \alpha_n \leq c_1, \quad n = 0, 1, \ldots . \tag{2.8}
\]

We will also assume that, for each $\alpha > 0$, the function $g_\alpha$ is defined on $[0, 1]$ and satisfies the following structure condition, where $\mathbb{C}$ denotes the complex plane.

**Assumption 1** For each $\alpha > 0$, the function
\[
\varphi_\alpha(\lambda) := g_\alpha(\lambda) - \frac{1}{\alpha + \lambda}
\]
extends to a complex analytic function defined on a domain $D_\alpha \subset \mathbb{C}$ such that $[0, 1] \subset D_\alpha$, and there is a contour $\Gamma_\alpha \subset D_\alpha$ enclosing $[0, 1]$ such that
\[
|z| \geq \frac{1}{2} \alpha \quad \text{and} \quad |z + \lambda| \leq b_0, \quad \forall z \in \Gamma_\alpha, \alpha > 0 \quad \text{and} \quad \lambda \in [0, 1], \tag{2.9}
\]
where $b_0$ is a constant independent of $\alpha > 0$. Moreover, there is a constant $b_1$ such that
\[
\int_{\Gamma_\alpha} |\varphi_\alpha(z)| |dz| \leq b_1 \tag{2.10}
\]
for all $0 < \alpha \leq c_1$.

By using the spectral integrals for self-adjoint operators, it follows easily from (2.9) in Assumption 1 that for any bounded linear operator $A$ with $\|A\| \leq 1$ there holds
\[
\|(zI - A^* A)^{-1}(A^* A)\| \leq \frac{b_0}{|z|^{1-\nu}} \tag{2.11}
\]
for $z \in \Gamma_\alpha$ and $0 \leq \nu \leq 1$.

Moreover, since Assumption 1 implies $\varphi_\alpha(z)$ is analytic in $D_\alpha$ for each $\alpha > 0$, there holds the Riesz-Dunford formula (see 1)
\[
\varphi_\alpha(A^* A) = \frac{1}{2\pi i} \int_{\Gamma_\alpha} \varphi_\alpha(z)(zI - A^* A)^{-1} dz
\]
for any linear operator $A$ satisfying $\|A\| \leq 1$.

**Assumption 2** Let $\{\alpha_n\}$ be a sequence of positive numbers, let $\{s_n\}$ be defined by (2.7). There is a constant $b_2 > 0$ such that
\[
0 \leq \lambda^\nu \prod_{k=j}^n r_{\alpha_k}(\lambda) \leq (s_n - s_{j-1})^{-\nu}, \tag{2.12}
\]
and
\[
0 \leq \lambda^\nu g_{\alpha_j}(\lambda) \prod_{k=j+1}^n r_{\alpha_k}(\lambda) \leq b_2 \frac{1}{\alpha_j} (s_n - s_{j-1})^{-\nu} \tag{2.13}
\]
for $0 \leq \nu \leq 1$, $0 \leq \lambda \leq 1$ and $j = 0, 1, \ldots , n$, where $r_{\alpha}(\lambda)$ is defined by (1.5).

In Section 4 we will give several important examples of $\{g_\alpha\}$ satisfying Assumptions 1 and 2. These examples of $\{g_\alpha\}$ include the ones arising from (iterated) Tikhonov regularization, asymptotical regularization, Landweber iteration and Lardy method.

**Lemma 1** The inequality (2.13) implies for $0 \leq \nu \leq 1$ and $\alpha > 0$ that
\[
0 \leq \lambda^\nu (\alpha + \lambda)^{-1} \prod_{k=j+1}^n r_{\alpha_k}(\lambda) \leq 2\alpha^{\nu-1} (1 + \alpha(s_n - s_j))^{-\nu} \tag{2.14}
\]
for all $0 \leq \lambda \leq 1$ and $j = 0, 1, \ldots , n$. 


Proof For $0 \leq \nu \leq 1$ and $\alpha > 0$ it follows from (2.12) that

$$0 \leq \lambda^\nu (\alpha + \lambda)^{-1} \prod_{k=j+1}^n r_{a_k}(\lambda) \leq \min \{ \alpha^{-1}, \alpha^{-\nu}(s_n - s_j)^{-\nu} \}$$

$$= \alpha^{-\nu} \min \{ 1, \alpha^{-\nu}(s_n - s_j)^{-\nu} \}$$

$$\leq 2^\nu \alpha^{-\nu} (1 + \alpha(s_n - s_j))^{-\nu}$$

for all $0 \leq \lambda \leq 1$ and $j = 0, 1, \cdots, n$. \hfill \Box

**Assumption 3**

(a) There exist constants $a \geq 0$ and $0 < m \leq M < \infty$ such that

$$m\|h\|_{-a} \leq \|F'(x)h\| \leq M\|h\|_{-a}, \quad h \in X$$

for all $x \in B_p(x^\dagger)$.

(b) $F$ is properly scaled so that $\|F'(x)L^{-s}\|_{X \to Y} \leq \min \{ 1, \sqrt{a_0} \}$ for all $x \in B_p(x^\dagger)$, where $s \geq -a$.

(c) There exist $0 < \beta \leq 1$, $0 \leq b \leq a$ and $K_0 \geq 0$ such that

$$\|F'(x)^* - F'(x^\dagger)^*\|_{Y \to X_b} \leq K_0\|x - x^\dagger\|^\beta$$

(2.15)

for all $x \in B_p(x^\dagger)$.

The number $a$ in condition (a) can be interpreted as the degree of ill-posedness of $F'(x)$ for $x \in B_p(x^\dagger)$. When $F$ satisfies the condition

$$F'(x) = R_x F'(x^\dagger) \quad \text{and} \quad \|I - R_x\| \leq K_0\|x - x^\dagger\|,$$

(2.16)

which has been verified in [6] for several nonlinear inverse problems, condition (a) is equivalent to

$$m\|h\|_{-a} \leq \|F'(x^\dagger)h\| \leq M\|h\|_{-a}, \quad h \in X$$

From (a) and (2.1) it follows for $s \geq -a$ that $\|F'(x)L^{-s}\|_{X \to Y} \leq M\theta^{s+a}$ for all $x \in B_p(x^\dagger)$. Thus $\|F'(x)L^{-s}\|_{X \to Y}$ is uniformly bounded over $B_p(x^\dagger)$. By multiplying (1.13) by a sufficiently small number, we may assume that $F$ is properly scaled so that condition (b) is satisfied. Furthermore, condition (a) implies that $F'(x)^*$ maps $Y$ into $X_b$ for $b \leq a$ and $\|F'(x)^*\|_{Y \to X_b} \leq M\theta^{a-b}$ for all $x \in B_p(x^\dagger)$. Condition (c) says that $F'(x)^*$ is locally Hölder continuous around $x^\dagger$ with exponent $0 < \beta \leq 1$ when considered as operators from $Y$ to $X_b$. It is equivalent to

$$\|L^h[F'(x)^* - F'(x^\dagger)^*]\|_{Y \to X} \leq K_0\|x - x^\dagger\|^\beta, \quad x \in B_p(x^\dagger)$$

or

$$\|[F'(x) - F'(x^\dagger)]L^h\|_{X \to Y} \leq K_0\|x - x^\dagger\|^\beta, \quad x \in B_p(x^\dagger).$$

Condition (c) was used first in [13] for the convergence analysis of Landweber iteration in Hilbert scales. It is easy to see that when $b = 0$ and $\beta = 1$, this is exactly the Lipschitz condition on $F'(x)$. When $F$ satisfies (2.16), (c) holds with $b = a$ and $\beta = 1$. In [13] it has been shown that (c) implies

$$\|F(x) - y - F'(x^\dagger)(x - x^\dagger)\| \leq K_0\|x - x^\dagger\|^\beta \|x - x^\dagger\|_{-b}$$

(2.17)

which follows easily from the identity

$$F(x) - y - F'(x^\dagger)(x - x^\dagger) = \int_0^1 [F'(x^\dagger + t(x - x^\dagger)) - F'(x^\dagger)] L^hL^{-b}(x - x^\dagger)dt.$$
In this paper we will derive, under the above assumptions on \( \{\alpha_n\} \), \( \{g_\alpha\} \) and \( F \), the rate of convergence of \( x_{n_\delta} \) to \( x^\dagger \) as \( \delta \to 0 \) when \( \varepsilon_0 := x_0 - x^\dagger \) satisfies the smoothness condition

\[
x_0 - x^\dagger \in X_\mu \quad \text{with} \quad \frac{a-b}{\beta} < \mu \leq b+2s,
\]

where \( n_\delta \) is the integer determined by the discrepancy principle \( 1.8 \) with \( \tau > 1 \).

The following consequence of the above assumptions on \( F \) and \( \{g_\alpha\} \) plays a crucial role in the convergence analysis.

**Lemma 2** Let \( \{g_\alpha\} \) satisfy Assumptions 1 and 2, let \( F \) satisfy Assumption 3, and let \( \{\alpha_n\} \) be a sequence of positive numbers. Let \( A = F'(x^\dagger)L^{-s} \) and for any \( x \in B_p(x^\dagger) \) let \( A_x = F'(x)L^{-s} \). Then for \( -\frac{b+a}{a+s+\delta} \leq \nu \leq 1/2 \) there holds \( 4 \)

\[
\left\| (A^*A)^\nu \prod_{k=j+1}^n r_{\alpha_k}(A^*A) \left[ g_{\alpha_j}(A^*A)A^* - g_{\alpha_j}(A^*_xA_x)A^*_x \right] \right\| \lesssim \frac{1}{\alpha_j} (s_n - s_{j-1})^{-\nu - \frac{b+a}{a+s+\delta}} K_0 \|x - x^\dagger\|^2
\]

for \( j = 0, 1, \ldots, n \).

**Proof** Let \( \eta_\alpha(\lambda) = (\alpha + \lambda)^{-1} \) and \( \varphi_\alpha(\lambda) = g_\alpha(\lambda) - (\alpha + \lambda)^{-1} \). We can write

\[
(A^*A)^\nu \prod_{k=j+1}^n r_{\alpha_k}(A^*A) \left[ g_{\alpha_j}(A^*A)A^* - g_{\alpha_j}(A^*_xA_x)A^*_x \right] = J_1 + J_2 + J_3,
\]

where

\[
J_1 := (A^*A)^\nu \prod_{k=j+1}^n r_{\alpha_k}(A^*A) g_{\alpha_j}(A^*A) [A^* - A^*_x],
\]

\[
J_2 := (A^*A)^\nu \prod_{k=j+1}^n r_{\alpha_k}(A^*A) \left[ \eta_{\alpha_j}(A^*A) - \eta_{\alpha_j}(A^*_xA_x) \right] A^*_x,
\]

\[
J_3 := (A^*A)^\nu \prod_{k=j+1}^n r_{\alpha_k}(A^*A) \left[ \varphi_{\alpha_j}(A^*A) - \varphi_{\alpha_j}(A^*_xA_x) \right] A^*_x.
\]

It suffices to show that the desired estimates hold for the norms of \( J_1, J_2 \) and \( J_3 \).

From 2.5, 2.13 in Assumption 2 and Assumption 3 it follows that

\[
\left\| J_1 \right\| \lesssim \left\| (A^*A)^\nu \prod_{k=j+1}^n r_{\alpha_k}(A^*A) g_{\alpha_j}(A^*A)(A^*_x - A^*) \right\| \times \left\| (A^*A)^{-\frac{b+a}{a+s+\delta}} [A^*_x - A^*] \right\|
\]

\[
\lesssim \sup_{0 \leq \lambda \leq 1} \left( \lambda^{b+a/s} \prod_{k=j+1}^n \left\| r_{\alpha_k}(\lambda) \right\| \|L^b(F'(x))^- - F'(x^\dagger)^-\|_Y \to X \right)
\]

\[
\lesssim \frac{1}{\alpha_j} (s_n - s_{j-1})^{-\nu - \frac{b+a}{a+s+\delta}} K_0 \|x - x^\dagger\|^2
\]

\[1\] Throughout this paper we will always use \( C \) to denote a generic constant independent of \( \delta \) and \( n \). We will also use the convention \( \Phi \lesssim \Psi \) to mean that \( \Phi \leq C\Psi \) for some generic constant \( C \).
which is the desired estimate.
In order to estimate \( \| J_2 \| \), we note that
\[
\eta_{\alpha_j}(A^*A) - \eta_{\alpha_j}(A^*_w A_w) = (\alpha_j I + A^*A)^{-1} \lambda^* (A_w - A)(\alpha_j I + A^*_w A_w)^{-1}
+ (\alpha_j I + A^*A)^{-1} \lambda (A_w - A^*)A_w (\alpha_j I + A^*_w A_w)^{-1}.
\]
Therefore \( J_2 = J_2^{(1)} + J_2^{(2)} \), where
\[
J_2^{(1)} = (A^*A)^{\nu} \prod_{k=j+1}^n r_{\alpha_k} (A^*A) (\alpha_j I + A^*A)^{-1} \lambda^*(A_w - A)(\alpha_j I + A^*_w A_w)^{-1},
\]
\[
J_2^{(2)} = (A^*A)^{\nu} \prod_{k=j+1}^n r_{\alpha_k} (A^*A) (\alpha_j I + A^*A)^{-1} \lambda (A_w - A^*)A_w (\alpha_j I + A^*_w A_w)^{-1}.
\]
With the help of Assumption \[3\] and (2.35) we have for any \( w \in Y \) that
\[
\| (A_w - A)(\alpha_j I + A^*_w A_w)^{-1} A^*_w w \|
= \| [F'(x) - F'(x^\dagger)]L^b L^{-(b+s)} (A_j I + A^*_w A_w)^{-1} A^*_w w \|
\leq K_0 \| x - x^\dagger \|^{\beta} \| (\alpha_j I + A^*_w A_w)^{-1} A^*_w w \|^{1-(b+s)}
\leq K_0 \| x - x^\dagger \|^{\beta} \| (A^*_w A_w)^\frac{b+s}{b+s} (\alpha_j I + A^*_w A_w)^{-1} A^*_w w \|
\leq K_0 \| x - x^\dagger \|^{\beta} \| a_j \|^{-\frac{1}{2}+\frac{b+s}{2(b+s)}} \| w \|.
\]
This implies
\[
\| (A_w - A)(\alpha_j I + A^*_w A_w)^{-1} A^*_w w \| \leq K_0 \| x - x^\dagger \|^{\beta} a_j \|^{-\frac{1}{2}+\frac{b+s}{2(b+s)}}.
\] (2.19)
Thus, by using Lemma \[11\] we derive
\[
\| J_2^{(1)} \| \leq \sup_{0 \leq \lambda \leq 1} \left( \lambda^{\nu+\frac{1}{2}} (\alpha_j + \lambda)^{-1} \prod_{k=j+1}^n r_{\alpha_k}(\lambda) \right) \| (A_w - A)(\alpha_j I + A^*_w A_w)^{-1} A^*_w w \|
\leq a_j \|^{-\nu+\frac{b+s}{2(b+s)}} (1 + \alpha_j (s_n - s_j))^{-\nu-\frac{1}{2}} K_0 \| x - x^\dagger \|^{\beta}.
\]
By using Assumption \[3\], Lemma \[11\] and a similar argument in estimating \( J_1 \) we can derive
\[
\| J_2^{(2)} \| \leq \sup_{0 \leq \lambda \leq 1} \left( \lambda^{\nu+\frac{b+s}{2(b+s)}} (\alpha_j + \lambda)^{-1} \prod_{k=j+1}^n r_{\alpha_k}(\lambda) \right) \| L^b F'(x)^* - F'(x^\dagger)^* \|_{Y \to X}
\leq a_j \|^{-\nu+\frac{b+s}{2(b+s)}} (1 + \alpha_j (s_n - s_j))^{-\nu+\frac{b+s}{2(b+s)}} K_0 \| x - x^\dagger \|^{\beta}.
\]
Combining the above estimates on \( J_2^{(1)} \) and \( J_2^{(2)} \) and noting \( \frac{b+s}{2(b+s)} \leq \frac{1}{2} \), it follows that
\[
\| J_2 \| \leq a_j \|^{-\nu+\frac{b+s}{2(b+s)}} (1 + \alpha_j (s_n - s_j))^{-\nu+\frac{b+s}{2(b+s)}} K_0 \| x - x^\dagger \|^{\beta}
= \frac{1}{\alpha_j} (s_n - s_j)^{-\nu+\frac{b+s}{2(b+s)}} K_0 \| x - x^\dagger \|^{\beta}.
\]
It remains to estimate \( J_3 \). Since Assumption \[11\] implies that \( \varphi_{\alpha_j}(z) \) is analytic in \( D_{\alpha_j} \), we have from the Riesz-Dunford formula that
\[
J_3 = \frac{1}{2\pi i} \int_{\Gamma_{\alpha_j}} \varphi_{\alpha_j}(z) T_j(z) dz,
\] (2.20)
where
\[ T_j(z) := (A^*A)^\nu \prod_{k=j+1}^n r_{\alpha_k}(A^*A) \left[ (zI - A^*A)^{-1} - (zI - A^*_x A_x)^{-1} \right] A_x^*. \]

We can write \( T_j(z) = T_j^{(1)}(z) + T_j^{(2)}(z) \), where
\[ T_j^{(1)}(z) := (A^*A)^\nu \prod_{k=j+1}^n r_{\alpha_k}(A^*A)(zI - A^*A)^{-1} A^*(A - A_x)(zI - A_x^* A_x)^{-1} A_x^*, \]
\[ T_j^{(2)}(z) := (A^*A)^\nu \prod_{k=j+1}^n r_{\alpha_k}(A^*A)(zI - A^*A)^{-1}(A^* - A_x^*) A_x A_x^*(zI - A_x A_x^*)^{-1}. \]

We will estimate the norms of \( T_j^{(1)}(z) \) and \( T_j^{(2)}(z) \) for \( z \in \Gamma_{\alpha_j} \). With the help of Assumption 3 (2.5) and (2.11), similar to the derivation of (2.19) we have
\[ \|(A - A_x)(zI - A_x^* A_x)^{-1} A_x^*\| \lesssim K_0 \|x - x^\dagger\|^\beta \frac{1}{\sqrt{\alpha + s}}. \]

Since \( |z| \geq \alpha_j/2 \) and \( |z| - \lambda \leq b_0(|z| + \lambda)^{-1} \) for \( z \in \Gamma_{\alpha_j} \), we have from (2.14) in Lemma 1 that
\[ \|T_j^{(1)}(z)\| \lesssim K_0 \|x - x^\dagger\|^\beta \left| \frac{1}{\sqrt{\alpha + s}} \right| \sup_{0 \leq \lambda \leq 1} \left( \lambda^{\nu + \frac{1}{2}} |z - \lambda|^{-1} \prod_{k=j+1}^n r_{\alpha_k}(\lambda) \right) \]
\[ \lesssim K_0 \|x - x^\dagger\|^\beta \left| \frac{1}{\sqrt{\alpha + s}} \right| \sup_{0 \leq \lambda \leq 1} \left( \lambda^{\nu + \frac{1}{2}} (|z| + \lambda)^{-1} \prod_{k=j+1}^n r_{\alpha_k}(\lambda) \right) \]
\[ \lesssim K_0 \|x - x^\dagger\|^\beta |z|^{\nu - 1 + \frac{b_0}{2(\alpha + s)}} (1 + (s_n - s_j)|z|)^{-\nu - 1/2} \]
\[ \lesssim K_0 \|x - x^\dagger\|^\beta \frac{1}{\alpha_j} |z|^{\nu - 1 + \frac{b_0}{2(\alpha + s)}} (1 + (s_n - s_j)\alpha_j)^{-\nu - 1/2}. \]

Next, by using (2.14) in Lemma 1, 2.5, Assumption 3a) and (2.11), we have for \( z \in \Gamma_{\alpha_j} \) that
\[ \|T_j^{(2)}(z)\| \lesssim \left\| (A^*A)^\nu \prod_{k=j+1}^n r_{\alpha_k}(A^*A)(zI - A^*A)^{-1} (A^*A)^{\frac{b_0}{2(\alpha + s)}} \right\| \]
\[ \times \left\| (A^*A)^{-\frac{b_0}{2(\alpha + s)}} (A^* - A_x^*) A_x A_x^* (zI - A_x A_x^*)^{-1} \right\| \]
\[ \lesssim \sup_{0 \leq \lambda \leq 1} \left( \lambda^{\nu + \frac{b_0}{2(\alpha + s)}} |z - \lambda|^{-1} \prod_{k=j+1}^n r_{\alpha_k}(\lambda) \right) \left\| L^b(F(x^\dagger)^* - F'(x)^*) \right\| \]
\[ \lesssim K_0 \|x - x^\dagger\|^\beta \sup_{0 \leq \lambda \leq 1} \left( \lambda^{\nu + \frac{b_0}{2(\alpha + s)}} (|z| + \lambda)^{-1} \prod_{k=j+1}^n r_{\alpha_k}(\lambda) \right) \]
\[ \lesssim K_0 \|x - x^\dagger\|^\beta |z|^{\nu - 1 + \frac{b_0}{2(\alpha + s)}} (1 + (s_n - s_j)|z|)^{-\nu - \frac{b_0}{2(\alpha + s)}} \]
\[ \lesssim K_0 \|x - x^\dagger\|^\beta \frac{1}{\alpha_j} |z|^{\nu - 1 + \frac{b_0}{2(\alpha + s)}} (1 + (s_n - s_j)\alpha_j)^{-\nu - \frac{b_0}{2(\alpha + s)}}. \]

Combining the above estimates on \( T_j^{(1)}(z) \) and \( T_j^{(2)}(z) \) and noting \( \frac{b_0}{2(\alpha + s)} \leq \frac{1}{2} \), it follows for \( z \in \Gamma_{\alpha_j} \) that
\[ \|T_j(z)\| \lesssim K_0 \|x - x^\dagger\|^\beta \frac{1}{\alpha_j} \frac{1}{\alpha_j} |z|^{\nu - 1 + \frac{b_0}{2(\alpha + s)}} (1 + (s_n - s_j)\alpha_j)^{-\nu - \frac{b_0}{2(\alpha + s)}} \]
\[ = \frac{1}{\alpha_j} (s_n - s_{j-1})^{-\nu - \frac{b_0}{2(\alpha + s)}} K_0 \|x - x^\dagger\|^\beta. \]
Therefore, it follows from (2.20) and Assumption \[\| \psi \| \leq C \] that
\[
\| J_3 \| \lesssim \frac{1}{\alpha_j} (s_n - s_{j-1})^{-\nu - \frac{\nu+\tau}{s_n}} K_0 \| x - x' \|^2 \int_{\gamma_j} | \psi_{\alpha_j}(z) | dz
\]
\[
\lesssim \frac{1}{\alpha_j} (s_n - s_{j-1})^{-\nu - \frac{\nu+\tau}{s_n}} K_0 \| x - x' \|^2.
\]
The proof is therefore complete. \[\square\]

3 Convergence analysis

We begin with the following lemma.

**Lemma 3** Let \( \{ \alpha_n \} \) be a sequence of positive numbers satisfying \( \alpha_n \leq c_1 \), and let \( s_n \) be defined by (2.7). Let \( p \geq 0 \) and \( q \geq 0 \) be two numbers. Then we have
\[
\sum_{j=0}^{n} \frac{1}{\alpha_j} (s_n - s_{j-1})^{-p} s_j^{-q} \leq C_0 s_n^{-p+q} \left\{ \begin{array}{ll}
1, & \max\{p, q\} < 1, \\
\log(1 + s_n), & \max\{p, q\} = 1,
\end{array} \right.
\]
where \( C_0 \) is a constant depending only on \( c_1, p \) and \( q \).

**Proof** This result is essentially contained in [5] Lemma 4.3 and its proof. For completeness, we include here the proof with a simplified argument. We first rewrite
\[
\sum_{j=0}^{n} \frac{1}{\alpha_j} (s_n - s_{j-1})^{-p} s_j^{-q} = s_n^{1-p-q} \sum_{j=0}^{n} \frac{1}{\alpha_j s_n} \left( 1 - \frac{s_{j-1}}{s_n} \right)^{-p} \left( \frac{s_j}{s_n} \right)^{-q}.
\]
Observe that when \( 0 \leq s_{j-1}/s_n \leq 1/2 \) we have
\[
\left( 1 - \frac{s_{j-1}}{s_n} \right)^{-p} \left( \frac{s_j}{s_n} \right)^{-q} \leq 2^p \left( \frac{s_j}{s_n} \right)^{-q}
\]
while when \( s_{j-1}/s_n \geq 1/2 \) we have
\[
\left( 1 - \frac{s_{j-1}}{s_n} \right)^{-p} \left( \frac{s_j}{s_n} \right)^{-q} \leq 2^q \left( 1 - \frac{s_{j-1}}{s_n} \right)^{-p}.
\]
Consequently there holds with \( C_{p,q} = \max\{2^p, 2^q\} \)
\[
\sum_{j=0}^{n} \frac{1}{\alpha_j} (s_n - s_{j-1})^{-p} s_j^{-q} \leq C_{p,q} s_n^{1-p-q} \left( \sum_{j=0}^{n} \frac{1}{\alpha_j s_n} \left( \frac{s_j}{s_n} \right)^{-q} + \sum_{j=0}^{n} \frac{1}{\alpha_j s_n} \left( 1 - \frac{s_{j-1}}{s_n} \right)^{-p} \right).
\]
(3.1)

Note that \( s_j - s_{j-1} = 1/\alpha_j \), we have with \( h = \frac{1}{2\alpha_0 s_n} \)
\[
\int_{s_n/s_n - h}^{1} t^{-q} dt \geq \sum_{j=1}^{n} \int_{s_j/s_n}^{s_j/s_n - h} t^{-q} dt + \int_{s_n/s_n - h}^{s_n/s_n} t^{-q} dt
\]
\[
\geq \sum_{j=1}^{n} \frac{s_j}{s_n}^{-q} \frac{s_j - s_{j-1}}{s_n} + \frac{1}{2\alpha_0 s_n} \frac{s_0}{s_n}^{-q}
\]
\[
\geq \frac{1}{2} \sum_{j=0}^{n} \frac{1}{\alpha_j s_n} \frac{s_j}{s_n}^{-q}.
\]
Therefore
\[
\sum_{j=0}^{n} \frac{1}{\alpha_j s_n} \left( \frac{s_j}{s_n} \right)^{-q} \leq 2 \int_{s_n/s_n-h}^{1} t^{-q} dt \leq \begin{cases} \frac{1}{q} \log(2s_n), & q < 1, \\ \frac{2}{q-1}(2s_n)^{-1}, & q > 1. \end{cases} \tag{3.2}
\]

By a similar argument we have with \( h = \frac{1}{2s_n s_n} \)
\[
\sum_{j=0}^{n} \frac{1}{\alpha_j s_n} \left( 1 - \frac{s_{j+1}}{s_n} \right)^{-p} \leq 2 \int_{0}^{\frac{s_n-1}{s_n}+h} (1-t)^{-p} dt \leq \begin{cases} \frac{1}{1-p} \log(2s_n), & p < 1, \\ \frac{2}{p-1}(2s_n)^{-p}, & p > 1. \end{cases} \tag{3.3}
\]

Combining (3.1), (3.2) and (3.3) and using the condition \( \alpha_n \leq c_1 \), we obtain the desired inequalities.

In order to derive the necessary estimates on \( x_n - x^\dagger \), we need some useful identities. For simplicity of presentation, we set
\[
e_n := x_n - x^\dagger, \quad A := F'(x^\dagger)L^{-s} \quad \text{and} \quad A_n := F'(x_n)L^{-s}.
\]

It follows from (1.7) and (2.6) that
\[
e_{n+1} = e_n - L^{-s}g\alpha_n (A_n^* A_n) A_n^* (F(x_n) - y^\delta).
\]

Let
\[
u_n := F(x_n) - y^\delta - F'(x^\dagger)(x_n - x^\dagger).
\]

Then we can write
\[
e_{n+1} = e_n - L^{-s}g\alpha_n (A^* A) A^* (F(x_n) - y^\delta) - L^{-s}g\alpha_n (A^* A) (y^\delta + u_n) - L^{-s}g\alpha_n (A^* A) A^* (F(x_n) - y^\delta). \tag{3.4}
\]

By telescoping (3.3) we can obtain
\[
e_{n+1} = L^{-s} \prod_{j=0}^{n} r_{\alpha_j} (A^* A) L^s e_0 - L^{-s} \sum_{j=0}^{n} \prod_{k=j+1}^{n} r_{\alpha_k} (A^* A) g\alpha_j (A^* A) (y^\delta + u_j) - L^{-s} \sum_{j=0}^{n} \prod_{k=j+1}^{n} r_{\alpha_k} (A^* A) g\alpha_j (A^* A) A^* (F(x_j) - y^\delta). \tag{3.5}
\]

By multiplying (3.5) by \( T := F'(x^\dagger) \) and noting that \( A = TL^{-s} \) and
\[
I - \sum_{j=0}^{n} \prod_{k=j+1}^{n} r_{\alpha_k} (A^* A) g\alpha_j (A^* A) A^* = \prod_{j=0}^{n} r_{\alpha_j} (A^* A),
\]

and let
\[
\{\text{constant}\}
\]

We will show (3.9) by induction. By using (3.7) and
\[
T e_{n+1} - y^\delta + y
\]

Based on (3.5) and (3.6) we will derive the order optimal convergence rate of \(x_{n_k} \to x^\dagger\) when \(e_0 := x_0 - x^\dagger\) satisfies the smoothness condition (2.18). Under such condition we have \(L^s e_0 \in X_{\mu-s} \) and \(\|\frac{d}{dt}L_t\| \leq 1\). Thus, with the help of Assumption 3(a), it follows from (2.4) and (2.5) that there exists \(\omega \in X\) such that
\[
L^s e_0 = (A^* A)^{\frac{n}{2(\mu+s)}} \omega \quad \text{and} \quad c_2 \|\omega\| \geq \|e_0\| \leq c_3 \|\omega\|
\]

for some generic constants \(c_3 \geq c_2 > 0\). We will first derive the crucial estimates on \(\|e_n\|\) and \(\|T e_n\|\). To this end, introduce the integer \(\tilde{n}_\delta\) satisfying
\[
s_{\tilde{n}_\delta} \leq \frac{(r - 1)\delta}{2c_0\|\omega\|} < s_n \frac{n+\mu}{2(\mu+s)}, \quad 0 \leq n < \tilde{n}_\delta,
\]

where \(c_0 > 1\) is the constant appearing in (2.5). Such \(\tilde{n}_\delta\) is well-defined since \(s_n \to \infty\) as \(n \to \infty\).

**Proposition 1** Let \(F\) satisfy Assumptions 3 and let \(\{g_{\alpha}\}\) satisfy Assumptions 7 and 8 and let \(\{\alpha_n\}\) be a sequence of positive numbers satisfying (2.3). If \(e_0 \in X_{\mu}\) for some \((a-b)\beta < b + 2s\) and if \(K_0 ||\omega||^2\) is suitably small, then there exists a generic constant \(C_* > 0\) such that
\[
\|e_n\| \leq C_* ||\omega|| \quad \text{and} \quad \|T e_n\| \leq C_* s_n \frac{n+\mu}{2(\mu+s)} ||\omega||
\]

and
\[
\|T e_n - y^\delta + y\| \leq (c_0 + C_0 \|\omega\|^2) s_n \frac{n+\mu}{2(\mu+s)} ||\omega|| + \delta
\]

for all \(0 \leq n \leq \tilde{n}_\delta\).

**Proof** We will show (3.9) by induction. By using (3.7) and \(\|A\| \leq \sqrt{c_0}\) we have
\[
\|T e_0\| = \|AL^s e_0\| = \|\mu A^* A^{\frac{n}{2(\mu+s)}} \omega\| = \|\mu A^* A^{\frac{n+\mu}{2(\mu+s)}} \omega\| \leq \alpha_0 \frac{n+\mu}{2(\mu+s)} ||\omega||.
\]

This together with (3.7) shows (3.9) for \(n = 0\) if \(C_* \geq \max\{1, c_3\}\). Next we assume that (3.9) holds for all \(0 \leq n \leq l\) for some \(l < \tilde{n}_\delta\) and we are going to show (3.9) holds for \(n = l + 1\).

With the help of (2.5) and (3.7) we can derive from (3.5) that
\[
\|e_{l+1}\| \leq \sum_{j=0}^{l} \|\alpha_j (A^* A)\| + \sum_{j=0}^{l} \|A^* A\| \sum_{k=j+1}^{l} \|r_{\alpha_k} (A^* A) (y - y^\delta + u_j)\|
\]

\[
+ \left(\sum_{j=0}^{l} \|A^* A\| \sum_{k=j+1}^{l} \|r_{\alpha_k} (A^* A) \left[\alpha_j (A^* A_j) A_j - \alpha_j (A^* A) A^*\right] (F(x_j) - y^\delta)\|ight).
\]
Therefore, by using the fact

$$0 \leq \frac{a + 2s - \mu}{2(a + s)} < 1 \quad \text{and} \quad -\frac{b + s}{2(a + s)} \leq \frac{s - \mu}{2(a + s)} < \frac{1}{2}$$

Thus we may use Assumption 2 and Lemma 2 to conclude

$$\|e_{l+1}\|_\mu \lesssim \|\omega\| + \sum_{j=0}^{l} \frac{1}{\alpha_j} (s_l - s_{j-1})^{-\frac{a + 2s - \mu}{2(a + s)}} (\delta + \|u_j\|)$$

$$+ \sum_{j=0}^{l} \frac{1}{\alpha_j} (s_l - s_{j-1})^{-\frac{b + s}{2(a + s)}} K_0 \|e_j\|^{\beta} \|F(x_j) - y^\beta\|.$$  \hspace{1cm} (3.11)

Moreover, by using (3.7), Assumption 2 and Lemma 2 we have from (3.6) that

$$\|Te_{l+1} - y^\beta + y\| \leq s_l^{-\frac{b + s}{2(a + s)}} \|\omega\| + \delta + \sum_{j=0}^{l} \frac{1}{\alpha_j} (s_l - s_{j-1})^{-1} \|u_j\|$$

$$+ c_4 \sum_{j=0}^{l} \frac{1}{\alpha_j} (s_l - s_{j-1})^{-\frac{2(a + s - \mu)}{2(a + s)}} K_0 \|e_j\|^{\beta} \|F(x_j) - y^\beta\|,$$  \hspace{1cm} (3.12)

where $c_4 > 0$ is a generic constant.

By using the interpolation inequality (2.3), Assumption 3(a) and the induction hypotheses, it follows for all $0 \leq j \leq l$ that

$$\|e_j\| \leq \frac{n}{a - \alpha} \|e_j\|_{\mu + \beta} \lesssim \|Te_j\| \|e_j\|_{\mu + \beta} \lesssim \|\omega\| s_j^{-\frac{a + \beta}{2(a + s)}}.$$  \hspace{1cm} (3.13)

With the help of (2.17) and the interpolation inequality (2.3), we have

$$\|u_j\| \lesssim K_0 \|e_j\|^{\beta} \|e_j\|_{-\alpha} \lesssim K_0 \|e_j\|_{\mu + \beta} \lesssim K_0 \|\omega\|^{1+\beta} s_j^{-\frac{a + \beta}{2(a + s)}}.$$  \hspace{1cm} (3.15)

On the other hand, since (2.14) and the induction hypotheses implies

$$\|e_j\|_{-\alpha} \lesssim \|e_j\|_{\mu} \lesssim \|\omega\|, \quad 0 \leq j \leq l$$

and since $\mu > (a - b)/\beta$, we have from (3.14) and Assumption 3(a) that

$$\|u_j\| \lesssim K_0 \|e_j\|_{-\alpha} \|e_j\|_{\mu + \beta} \lesssim K_0 \|\omega\|^{\beta} \|Te_j\|.$$  \hspace{1cm} (3.16)

Therefore, by using the fact

$$\delta \lesssim \frac{2c_0}{\tau - 1} \|\omega\| s_j^{\frac{-\mu}{a + \beta}}, \quad 0 \leq j \leq l$$

and the induction hypotheses we have

$$\|F(x_j) - y^\beta\| \leq \delta + \|Te_j\| + \|u_j\| \lesssim \|\omega\| s_j^{\frac{-\mu}{2(a + s)}}.$$  \hspace{1cm} (3.18)

In view of the estimates (3.13), (3.15), (3.18) and the inequality

$$\sum_{j=0}^{l} \frac{1}{\alpha_j} (s_l - s_{j-1})^{-\frac{a + 2s - \mu}{2(a + s)}} \lesssim s_l^{-\frac{a + \beta}{2(a + s)}}$$
which follows from Lemma 3 we have from (3.11) and (3.12) that
\[
\|e_{t+1}\|_{\mu} \leq c_5 \|w\| + c_5 s_l \frac{a+b}{a+b \beta} \delta \\
+ CK_0 \|w\|^{1+\beta} \sum_{j=0}^{l} \frac{1}{a_j} (s_l - s_{j-1}) \frac{a+b}{a+b \beta} s_j \\
+ CK_0 \|w\|^{1+\beta} \sum_{j=0}^{l} \frac{1}{a_j} (s_l - s_{j-1}) \frac{b+b \beta}{b+b \beta} s_j 
\]
and
\[
\|Te_{t+1} - y^\delta + y\| \leq \|w\| s_l^{-\frac{a+b}{a+b \beta}} + \delta \\
+ CK_0 \|w\|^{1+\beta} \sum_{j=0}^{l} \frac{1}{a_j} (s_l - s_{j-1})^{-1} s_j^{-\frac{b+b \beta}{b+b \beta}} \\
+ CK_0 \|w\|^{1+\beta} \sum_{j=0}^{l} \frac{1}{a_j} (s_l - s_{j-1})^{-\frac{b+b \beta}{b+b \beta}} s_j^{-\frac{b+b \beta}{b+b \beta}},
\]
where \(c_5\) and \(C\) are two positive generic constants.

With the help of Lemma 3 \(\mu > (a-b)/\beta\), (3.17) and (2.8) we have
\[
\|e_{t+1}\|_{\mu} \leq \left( c_5 + \frac{2}{\tau - 1} c_0 c_5 + CK_0 \|w\|^\beta \right) \|w\|
\]
and
\[
\|Te_{t+1} - y^\delta + y\| \leq \delta + (1 + CK_0 \|w\|^\beta) \|w\| s_l^{-\frac{a+b}{a+b \beta}} \\
\leq \delta + c_0 \left( 1 + CK_0 \|w\|^\beta \right) \|w\| s_{t+1}^{-\frac{a+b}{a+b \beta}}. \tag{3.19}
\]
Consequently \(\|e_{t+1}\|_{\mu} \leq C_\ast \|w\|\) if \(C_\ast \geq 2c_5 + \frac{2}{\tau - 1} c_0 c_5\) and \(K_0 \|w\|^\beta\) is suitably small. Moreover, from (3.19), (3.17) and (2.8) we also have
\[
\|Te_{t+1}\| \leq 2\delta + c_0 \left( 1 + CK_0 \|w\|^\beta \right) \|w\| s_{t+1}^{-\frac{a+b}{a+b \beta}} \\
\leq \left( \frac{4c_0^2}{\tau - 1} + c_0 + CK_0 \|w\|^\beta \right) \|w\| s_{t+1}^{-\frac{a+b}{a+b \beta}} \\
\leq C_\ast \|w\| s_{t+1}^{-\frac{a+b}{a+b \beta}}
\]
if \(C_\ast \geq 2c_0 + \frac{4c_0^2}{\tau - 1}\) and \(K_0 \|w\|^\beta\) is suitably small. We therefore complete the proof of (3.9). In the meanwhile, (3.19) gives the proof of (3.10). \(\Box\)

From Proposition 1 and its proof it follows that \(x_n \in B_\rho(x^\ast)\) for \(0 \leq n \leq \hat{n}_\delta\) if \(\|w\|\) is sufficiently small. Furthermore, from (3.15) and (3.16) we have
\[
\|F(x_n) - y - Te_n\| \lesssim K_0 \|w\|^{1+\beta} s_n \frac{b+b \beta}{b+b \beta}
\]
and
\[
\|F(x_n) - y - Te_n\| \lesssim K_0 \|w\|^\beta \|Te_n\| \tag{3.21}
\]
for \(0 \leq n \leq \hat{n}_\delta\).

In the following we will show that \(n_\delta \leq \hat{n}_\delta\) for the integer \(n_\delta\) defined by (1.8) with \(\tau > 1\). Consequently, the method given by (1.7) and (1.8) is well-defined.
Lemma 4 Let all the conditions in Proposition 1 hold. Let $\tau > 1$ be a given number. If $c_0 \in X_\mu$ for some $(a - b)/\beta < \mu \leq b + 2s$ and if $K_0\|c_0\|_\beta$ is suitably small, then the discrepancy principle (1.8) defines a finite integer $n_\delta \leq \tilde{n}_\delta$.

Proof From Proposition 1 and $\mu > (a - b)/\beta$ it follows for $0 \leq n \leq \tilde{n}_\delta$ that
\[
\|F(x_n) - y^\delta\| \leq \|F(x_n) - y - T e_n\| + \|T e_n - y^\delta + y\|
\leq CK_0\|\omega\|^{1+\beta} s_n^{-a/(a+b)} + (c_0 + CK_0\|\omega\|^\beta) s_n^{-a/(a+b)} \|\omega\| + \delta
\leq (c_0 + CK_0\|\omega\|^\beta) s_n^{-a/(a+b)} \|\omega\| + \delta.
\]
By setting $n = \tilde{n}_\delta$ in the above inequality and using the definition of $\tilde{n}_\delta$ we obtain
\[
\|F(x_n) - y^\delta\| \leq \left(1 + \frac{\tau - 1}{2} + CK_0\|\omega\|^\beta\right) \delta \leq \tau \delta
\]
if $K_0\|\omega\|^\beta$ is suitably small. According to the definition of $n_\delta$ we have $n_\delta \leq \tilde{n}_\delta$. \qed

Now we are ready to prove the main result concerning the order optimal convergence rates for the method defined by (1.7) and (1.8) with $\tau > 1$.

Theorem 1 Let $F$ satisfy Assumptions \(3\) let $\{g_n\}$ satisfy Assumptions \(4\) and \(5\) and let $\{\alpha_n\}$ be a sequence of positive numbers satisfying (2.2). If $c_0 \in X_\mu$ for some $(a - b)/\beta < \mu \leq b + 2s$ and if $K_0\|c_0\|_\beta$ is suitably small, then for all $r \in [-a, \mu]$ there holds
\[
\|x_n - x^\dagger\|_r \leq C\|c_0\|_\mu^{\alpha_\tau} \delta^{\frac{a+\tau}{\alpha_\tau}}
\]
for the integer $n_\delta$ determined by the discrepancy principle (1.8) with $\tau > 1$, where $C > 0$ is a generic constant.

Proof It follows from (3.21) that if $K_0\|\omega\|^\beta$ is suitably small then
\[
\|F(x_n) - y - T e_n\| \leq \frac{1}{2}\|T e_n\|
\]
which implies $\|T e_n\| \leq 2\|F(x_n) - y\|$ for $0 \leq n \leq \tilde{n}_\delta$. Since Lemma 4 implies $n_\delta \leq \tilde{n}_\delta$, it follows from Assumption 3(a) and the definition of $n_\delta$ that
\[
\|e_n\|_a \leq \frac{1}{m}\|T e_n\| \leq \frac{2}{m} \left(\|F(x_n) - y^\delta\| + \delta\right) \leq \frac{2(1 + \tau)}{m} \delta.
\]
But from Proposition 1 we have $\|e_n\|_\mu \leq C_\mu\|\omega\|$. The desired estimate then follows from the interpolation inequality (2.23) and 3.7. \qed

Remark 1 If $F$ satisfies (2.10) and $\{x_n\}$ is defined by (1.7) with $s > -a/2$, then the order optimal convergence rate holds for $x_0 - x^\dagger \in X_\mu$ with $0 < \mu \leq a + 2s$. On the other hand, if $F(x)$ satisfies the Lipschitz condition
\[
\|F'(x) - F'(x^\dagger)\| \leq K_0\|x - x^\dagger\|, \quad x \in B_\mu(x^\dagger)
\]
and $\{x_n\}$ is defined by (1.7) with $s > a/2$, then the order optimal convergence rate holds for $x_0 - x^\dagger \in X_\mu$ with $a < \mu \leq 2s.$
4 Examples

In this section we will give several important examples of \{g_\alpha\} that satisfy Assumptions 1 and 2. Thus, Theorem 1 applies to the corresponding methods if \(F\) satisfies Assumption 3 and \{\alpha_n\} satisfies (2.8). For all these examples, the functions \(g_\alpha\) are analytic at least in the domain

\[D_\alpha := \{z \in \mathbb{C} : z \neq -\alpha, -1\}.\]

Moreover, for each \(\alpha > 0\), we always take the closed contour \(\Gamma_\alpha\) to be (see [1])

\[\Gamma_\alpha = \Gamma_\alpha^{(1)} \cup \Gamma_\alpha^{(2)} \cup \Gamma_\alpha^{(3)} \cup \Gamma_\alpha^{(4)},\]

with

\[
\begin{align*}
\Gamma_\alpha^{(1)} &:= \{z = \frac{\alpha}{2} e^{i\phi} : \phi_0 \leq \phi \leq 2\pi - \phi_0\}, \\
\Gamma_\alpha^{(2)} &:= \{z = Re^{i\phi} : -\phi_0 \leq \phi \leq \phi_0\}, \\
\Gamma_\alpha^{(3)} &:= \{z = te^{i\phi_0} : \alpha/2 \leq t \leq R\}, \\
\Gamma_\alpha^{(4)} &:= \{z = te^{-i\phi_0} : \alpha/2 \leq t \leq R\},
\end{align*}
\]

where \(R > \max\{1, \alpha\}\) and \(0 < \phi_0 < \pi/2\) are fixed numbers. Clearly \(\Gamma_\alpha \subset D_\alpha\) and \([0, 1]\) lies inside \(\Gamma_\alpha\). It is straightforward to check that (2.9) is satisfied.

Example 1

We first consider for \(\alpha > 0\) the function \(g_\alpha\) given by

\[g_\alpha(\lambda) = \frac{(\alpha + \lambda)^N - \alpha^N}{\lambda(\alpha + \lambda)^N}\]

where \(N \geq 1\) is a fixed integer. This function arises from the iterated Tikhonov regularization of order \(N\) for linear ill-posed problems. The corresponding method (1.7) becomes

\[
\begin{align*}
 u_{n,0} &= x_n, \\
 u_{n,l+1} &= u_{n,l} - (\alpha_n L^{2s} + T_n T_n^*)^{-1} T_n^* (F(x_n) - y^\delta - T_n (x_n - u_{n,l})) , \\
 x_{n+1} &= u_{n,N},
\end{align*}
\]

where \(T_n : = F'(x_n)\). When \(N = 1\), this is the Levenberg-Marquardt method in Hilbert scales. The corresponding residual function is \(r_\alpha(\lambda) = \alpha N(\alpha + \lambda)^{-N}\). In order to verify Assumption 2 we recall the inequality (see [3] Lemma 3)

\[
\lambda \prod_{k=j}^{n} \frac{\alpha_k}{\alpha_k + \lambda} \leq (s_n - s_{j-1})^{-1} \quad \text{for all } \lambda \geq 0.
\]

Then for \(0 \leq \nu \leq 1\) and \(\lambda \geq 0\) we have

\[
\lambda^\nu \prod_{k=j}^{n} r_{\alpha_k}(\lambda) \leq \left( \lambda \prod_{k=j}^{n} \frac{\alpha_k}{\alpha_k + \lambda} \right)^\nu \leq (s_n - s_{j-1})^{-\nu}.
\]
and

\[ \lambda^\nu g_{\alpha_j}(\lambda) \prod_{k=j+1}^n r_{\alpha_k}(\lambda) = \frac{(\alpha_j + \lambda)^N - \alpha_j^N}{\alpha_j^{N-\nu}} \prod_{k=j}^n \left( \frac{\alpha_k}{\alpha_k + \lambda} \right)^N \]

\[ = \sum_{l=0}^{N-1} \binom{N}{l} \alpha_j^{l-N} \lambda^{N+N-1-1} \prod_{k=j}^n \left( \frac{\alpha_k}{\alpha_k + \lambda} \right)^N \]

\[ \leq \sum_{l=0}^{N-1} \binom{N}{l} \alpha_j^{l-N} \left( \lambda \prod_{k=j}^n \frac{\alpha_k}{\alpha_k + \lambda} \right)^{N+N-1} \]

\[ \leq \sum_{l=0}^{N-1} \binom{N}{l} \alpha_j^{l-N} (s_n - s_{j-1})^{N-1} \]

\[ \leq C_N \frac{1}{\alpha_j} (s_n - s_{j-1})^{-\nu}, \]

where \( C_N = 2^{N-1} \) and we used the fact \( \alpha_j^{-1} \leq s_n - s_{j-1} \). We therefore obtain (2.12) and (2.13) in Assumption 2.

Next we will verify (2.10) in Assumption 1. Note that

\[ \alpha \phi(z) = \alpha \phi(z) = \frac{1}{z(z + \alpha z)^N} \sum_{j=0}^{N-2} \binom{N-1}{j} \alpha_j^{j+1} z^{N-1-j}. \]

It is easy to check \(|\phi(z)| \leq \alpha^{-1} \) on \( R^{(1)}_\alpha \) and \(|\phi(z)| \leq 1 \) on \( R^{(2)}_\alpha \). Moreover, on \( R^{(3)}_\alpha \cup R^{(4)}_\alpha \) there holds

\[ |\phi(z)| \leq \frac{1}{t(\alpha + t)^N} \sum_{j=0}^{N-2} \alpha_j^{j+1} t^{N-1-j} \leq \sum_{j=0}^{N-2} \alpha_j^{j+1} t^{-2-j}. \]

Therefore

\[ \int_{r_\alpha} |\phi(z)| dz = \int_{r_\alpha^{(1)}} |\phi(z)| dz + \int_{r_\alpha^{(2)}} |\phi(z)| dz + \int_{r_\alpha^{(3)} \cup r_\alpha^{(4)}} |\phi(z)| dz \]

\[ \leq \alpha^{-1} \int_{\phi_0}^{2\pi - \phi_0} a d\phi + \int_{-\phi_0}^{\phi_0} d\phi + \sum_{j=0}^{N-2} \alpha_j^{j+1} \int_{r_\alpha/2}^{R} t^{-2-j} dt \]

\[ \lesssim 1. \]

Assumption 1 is therefore verified.

**Example 2** We consider the method (1.7) with \( g_\alpha \) given by

\[ g_\alpha(\lambda) = \frac{1}{\lambda} \left( 1 - e^{-\lambda/\alpha} \right) \]

which arises from the asymptotic regularization for linear ill-posed problems. In this method, the iterative sequence \( \{x_n\} \) is equivalently defined as \( x_{n+1} := x(1/\alpha_n) \), where \( x(t) \) is the unique solution of the initial value problem

\[ \frac{d}{dt} x(t) = \lambda^{-2} F'(x_n)^* (y - F(x_n) + F'(x_n)(x_n - x(t))), \quad t > 0, \]

\[ x(0) = x_n. \]
The corresponding residual function is \( r_\alpha (\lambda) = e^{-\lambda/\alpha} \). We first verify Assumption 2. It is easy to see
\[
\lambda^n \prod_{k=1}^{n} r_{\alpha_k}(\lambda) = \lambda^n e^{-\lambda (s_n - s_{n-1})} \leq \nu^n e^{-\nu (s_n - s_{n-1})} \leq (s_n - s_{n-1})^{-\nu}
\]
for \( 0 \leq \nu \leq 1 \) and \( \lambda \geq 0 \). This shows (2.12). By using the elementary inequality \( e^{-p\lambda} - e^{-q\lambda} \leq (q - p)/q \) for \( 0 < p \leq q \) and \( \lambda \geq 0 \) and observing that \( 0 \leq r_\alpha (\lambda) \leq 1 \) and \( 0 \leq g_\alpha (\lambda) \leq 1/\alpha \), we have for \( 0 \leq \nu \leq 1 \) and \( \lambda \geq 0 \) that
\[
\lambda^n g_{\alpha_j}(\lambda) \prod_{k=j+1}^{n} r_{\alpha_k}(\lambda) \leq \frac{1}{\alpha_j} \left( \lambda g_{\alpha_j}(\lambda) \prod_{k=j+1}^{n} r_{\alpha_k}(\lambda) \right)^\nu 
\]
which gives (2.14). In order to verify (2.10) in Assumption 1 we note that
\[
\varphi_\alpha(z) = \frac{1 - e^{-z/\alpha}}{z} \leq 1 \frac{\alpha + z e^{-z/\alpha}}{z(\alpha + z)}.
\]
It is easy to see that \( |\varphi_\alpha(z)| \lesssim z^{-1} \) on \([1/\alpha, 1]\), \( |\varphi_\alpha(z)| \lesssim 1 \) on \([1/\alpha, 2]\) and
\[
|\varphi_\alpha(z)| \lesssim \frac{\alpha + (\alpha + t) e^{-z \cos \phi_\alpha}}{t(\alpha + t)} \lesssim \alpha t^{-2}
\]
on \([1/\alpha, 3] \cup [1/\alpha, 4]\). Therefore
\[
\int_{[1/\alpha]} |\varphi_\alpha(z)||dz| \lesssim 1 + \int_{\alpha/2} \alpha t^{-2} dt \lesssim 1.
\]

Example 3 We consider for \( 0 < \alpha \leq 1 \) the function \( g_\alpha \) given by
\[
g_\alpha(\lambda) = \sum_{l=0}^{[1/\alpha] - 1} (1 - \lambda)^l = \frac{1 - (1 - \lambda)^{[1/\alpha]}}{\lambda}
\]
which arises from the linear Landweber iteration, where \([1/\alpha]\) denotes the largest integer not greater than \(1/\alpha\). The method (1.7) then becomes
\[
\begin{align*}
u_{n,1} &= u_n, \\
u_{n,t+1} &= u_n, - L^{-2s} T_n \left( F(x_n) - y^\delta - T_n (x_n - u_{n,t}) \right), \quad 0 \leq t \leq [1/\alpha] - 1, \\
u_{n+1} &= u_{n,[1/\alpha]},
\end{align*}
\]
where \( T_n := F'(x_n) \). When \( \alpha_n = 1 \) for all \( n \), this method reduces to the Landweber iteration in Hilbert scales proposed in [13]. The corresponding residual function is \( r_\alpha (\lambda) = (1 - \lambda)^{1/[1/\alpha]} \). We first verify Assumption 2 when the sequence \( \{\alpha_n\} \) is given by \( \alpha_n = 1/k_n \) for some integers \( k_n \geq 1 \). Then for \( 0 \leq \nu \leq 1 \) and \( 0 \leq \lambda \leq 1 \) we have
\[
\lambda^n \prod_{k=1}^{n} r_{\alpha_k}(\lambda) = \lambda^n (1 - \lambda)^{s_n - s_{n-1}} \leq \nu^n (s_n - s_{n-1})^{-\nu} \leq (s_n - s_{n-1})^{-\nu}.
\]
We thus obtain (2.12). Observing that \( 0 \leq r_{\alpha_j}(\lambda) \leq 1 \) and \( 0 \leq g_{\alpha_j}(\lambda) \leq 1/\alpha_j \) for \( 0 \leq \lambda \leq 1 \), we have
\[
\lambda^\nu g_{\alpha_j}(\lambda) \prod_{k=j+1}^n r_{\alpha_k}(\lambda) \leq \frac{1}{\alpha_j^{1-\nu}} \left( \lambda g_{\alpha_j}(\lambda) \prod_{k=j+1}^n r_{\alpha_k}(\lambda) \right)^\nu
= \frac{1}{\alpha_j^{1-\nu}} \left( (1 - \lambda)^{s_{\alpha_j} - 1} - (1 - \lambda)^{s_{\alpha_{j-1}}} \right)^\nu.
\]
Thus, (2.13) follows from the elementary inequality \( t^p - t^q \leq (q - p)/q \) for \( 0 < p \leq q \) and \( 0 \leq t \leq 1 \).

In order to verify (2.10) in Assumption 1, in the definition of \( \Gamma \), we pick \( R > 1 \) and \( 0 < \phi_0 < \pi/2 \) such that \( R < 2 \cos \phi_0 \). Note that
\[
\varphi_{\alpha}(z) = 1 - \frac{(1 - z)^{[1/\alpha]}}{z} - \frac{1}{\alpha + z} = \frac{\alpha - (\alpha + z)(1 - z)^{[1/\alpha]}}{z(\alpha + z)}.
\]
By using the fact \((1 + \alpha)^{1/\alpha} \leq e\) we can see
\[
|\varphi_{\alpha}(z)| \leq \alpha^{-1}(1 + \alpha/2)^{1/\alpha} \lesssim \alpha^{-1} \quad \text{on } \Gamma_{\alpha}^{(1)}.
\]
According to the choice of \( R \) and \( \phi_0 \), we have \( 1 + R^2 - 2R \cos \phi_0 < 1 \). Thus
\[
|\varphi_{\alpha}(z)| \lesssim \frac{\alpha + (\alpha + R)(1 + R^2 - 2R \cos \phi_0)^{1/[\alpha]/2}}{R(R + \alpha)} \lesssim 1 \quad \text{on } \Gamma_{\alpha}^{(2)}.
\]
Furthermore, on \( \Gamma_{\alpha}^{(3)} \cup \Gamma_{\alpha}^{(4)} \) we have
\[
|\varphi_{\alpha}(z)| \lesssim \frac{\alpha + (\alpha + t)(1 + t^2 - 2t \cos \phi_0)^{1/(2\alpha)}}{t(\alpha + t)}.
\]
Therefore
\[
\int_{\Gamma_{\alpha}} |\varphi_{\alpha}(z)||dz| \lesssim 1 + \int_{\alpha/2}^R \frac{\alpha + (\alpha + t)(1 + t^2 - 2t \cos \phi_0)^{1/(2\alpha)}}{t(\alpha + t)} dt
= 1 + \int_{1/2}^{R/\alpha} \frac{1 + (1 + t)(1 + \alpha^2 t^2 - 2\alpha t \cos \phi_0)^{1/(2\alpha)}}{t(1 + t)} dt
\lesssim 1 + \int_{1/2}^{R/\alpha} (1 + \alpha^2 t^2 - 2\alpha t \cos \phi_0)^{1/(2\alpha)} dt.
\]
Observe that for \( 1/2 \leq t \leq R/\alpha \) there holds
\[
(1 + \alpha^2 t^2 - 2\alpha t \cos \phi_0)^{1/(2\alpha)} \leq (1 - \mu_0 t)^{1/(2\alpha)} \leq e^{-\mu t/2}
\]
with \( \mu_0 := 2 \cos \phi_0 - R > 0 \). Thus
\[
\int_{\Gamma_{\alpha}} |\varphi_{\alpha}(z)||dz| \lesssim 1 + \int_{1/2}^{R/\alpha} e^{-\mu t/2} dt \lesssim 1.
\]

**Example 4** We consider for \( 0 < \alpha \leq 1 \) the function \( g_{\alpha} \) given by
\[
g_{\alpha}(\lambda) = \sum_{i=1}^{[1/\alpha]} (1 + \lambda)^{-1} = \frac{1 - (1 + \lambda)^{-[1/\alpha]}}{\lambda}
\]
which arises from the Lardy method for linear inverse problems. Then the method (1.7) becomes

\[
\begin{align*}
    u_{n,0} &= x_n, \\
    u_{n,l+1} &= u_{n,l} - (L^2s + T_n^* T_n)^{-1} T_n^* \left( F(x_n) - y^\delta - T_n(x_n - u_{n,l}) \right), \\
    x_{n+1} &= u_{n,[1/\alpha_n]},
\end{align*}
\]

where \( T_n = F'(x_n) \). The residual function is \( r_\alpha(\lambda) = (1 + \lambda)^{-[1/\alpha]} \). Assumption \( \mathbb{H} \) and Assumption \( \mathbb{I} \) can be verified similarly as in Example 3 when the sequence \( \{\alpha_n\} \) is given by \( \alpha_n = 1/k_n \) for some integers \( k_n \geq 1 \).

References

1. A. B. Bakushinsky and M. Yu. Kokurin, *Iterative Methods for Approximate Solutions of Inverse Problems*, Mathematics and its applications, Springer, 2004.
2. R. S. Dembo, S. C. Eisenstat and T. Steihaug, *Inexact Newton methods*, SIAM J. Numer. Anal., 19 (1982), 400–408.
3. H. W. Engl, M. Hanke and A. Neunauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.
4. M. Hanke, *A regularizing Levenberg-Marquardt scheme with applications to inverse groundwater filtration problems*, Inverse Problems, 13(1997), 79–95.
5. M. Hanke, *The regularizing Levenberg-Marquardt scheme is of optimal order*, J. Integral Equations and Applications, 22 (2010), no. 2, 259–283.
6. M. Hanke, A. Neubauer and O. Scherzer, *A convergence analysis of the Landweber iteration for nonlinear ill-posed problems*, Numer. Math., 72 (1995), 21–37.
7. M. Hochbruck, M. Hönig and A. Ostermann, *A convergence analysis of the exponential Euler iteration for nonlinear ill-posed problems*, Inverse Problems, 25 (2009), no.7, article no. 075009.
8. Q. Jin, *Error estimates of some Newton-type methods for solving nonlinear inverse problems in Hilbert scales*, Inverse Problems, 16 (2000), no. 1, 187–197.
9. Q. Jin, *On a regularized Levenberg-Marquardt method for solving nonlinear inverse problems*, Numer. Math., 115 (2010), no. 2, 229–259.
10. J. Köhler and U. Tautenhahn, *Error bounds for regularized solutions of nonlinear ill-posed problems*, J. Inv. Ill-Posed Problems 3 (1995), 47–74.
11. A. Lechleiter and A. Rieder, *Towards a general convergence theory for inexact Newton regularizations*, Numer. Math. 114 (2010), no. 3, 521–548.
12. F. Natterer, *Error bounds for Tikhonov regularization in Hilbert scales*, Appl. Anal., 18 (1984), 29–37.
13. A. Neubauer, *On Landweber iteration for nonlinear ill-posed problems in Hilbert scales*, Numer. Math., 85 (2000), 309–328.
14. A. Rieder, *On the regularization of nonlinear ill-posed problems via inexact Newton iterations*, Inverse Problems, 15(1999), 309–327.
15. A. Rieder, *On convergence rates of inexact Newton regularizations*, Numer. Math. 88(2001), 347–365.
16. U. Tautenhahn, *Error estimates for regularization methods in Hilbert scales*, SIAM J. Numer Anal., 33 (1996), 2120–2130.
17. U. Tautenhahn, *On a general regularization scheme for nonlinear ill-posed problems: II. regularization in Hilbert scales*, Inverse Problems, 14 (1998), 1607–1616.