A convergence analysis of the iteratively regularized Gauss–Newton method under the Lipschitz condition

Qinian Jin

Department of Mathematics, The University of Texas at Austin, Austin, TX 78712, USA

E-mail: qjin@math.utexas.edu

Received 20 February 2008, in final form 7 April 2008
Published 22 May 2008
Online at stacks.iop.org/IP/24/045002

Abstract

In this paper we consider the iteratively regularized Gauss–Newton method for solving nonlinear ill-posed inverse problems. Under merely the Lipschitz condition, we prove that this method together with an \textit{a posteriori} stopping rule defines an order optimal regularization method if the solution is regular in some suitable sense.

1. Introduction

In this paper we will consider the nonlinear inverse problems which can be formulated as the operator equations

\[ F(x) = y, \quad (1.1) \]

where \( F : D(F) \subset X \to Y \) is a nonlinear operator between the Hilbert spaces \( X \) and \( Y \) with domain \( D(F) \) and range \( R(F) \). Such problems arise naturally from the parameter identification in partial differential equations. For instance, consider the identification of the parameter \( c \) in the boundary value problem

\[
\begin{cases}
-\Delta u + cu = f & \text{in } \Omega \\
u = g & \text{on } \partial \Omega
\end{cases} \quad (1.2)
\]

from the measurement of the state \( u \), where \( \Omega \subset \mathbb{R}^n, n \leq 3 \), is a bounded domain with smooth boundary \( \partial \Omega \), \( f \in L^2(\Omega) \) and \( g \in H^{3/2}(\partial \Omega) \). It is well known that (1.2) has a unique solution \( u := u(c) \in H^1(\Omega) \subset L^2(\Omega) \) for each

\[ c \in D := \{ c \in L^2(\Omega) : \| c - \hat{c} \|_{L^2(\Omega)} \leq \gamma \text{ for some } \hat{c} \geq 0 \text{ a.e.} \} \]

with some \( \gamma > 0 \). If we define the operator \( F \) as

\[ F : D \subset L^2(\Omega) \to L^2(\Omega), \quad c \to u(c), \]

then the problem of identifying \( c \) is reduced to solving (1.1).
Throughout this paper \( \| \cdot \| \) and \((\cdot, \cdot)\) will be used to denote the norms and inner products for both spaces \( X \) and \( Y \) since there is no confusion. The nonlinear operator \( F \) is always assumed to be Fréchet differentiable, the Fréchet derivative of \( F \) at \( x \in D(F) \) will be denoted as \( F'(x) \) and \( F'(x)^* \) will be used to denote the adjoint of \( F'(x) \). We will assume that \( y \) is attainable, i.e. problem (1.1) has a solution \( x^\dagger \in D(F) \) such that
\[
F(x^\dagger) = y.
\]

We say problem (1.1) is ill-posed if its solution does not depend continuously on the right-hand side \( y \), which is the characteristic property for most of the inverse problems. Since the right-hand side is usually obtained by measurement, thus, instead of \( y \) itself, the available data are an approximation \( y^\delta \) satisfying
\[
\| y^\delta - y \| \leq \delta
\]
with a given small noise level \( \delta > 0 \). Then the computation of a stable solution of (1.1) from \( y^\delta \) becomes an important issue of ill-posed problems, and the regularization techniques have to be taken into account.

Tikhonov regularization is one of the well-known methods that has been studied extensively in recent years. Several \textit{a posteriori} rules have been suggested to choose the regularization parameter. Besides the Morozov discrepancy principle, an \textit{a posteriori} rule has been proposed in [9] to yield optimal rates of convergence. It has been shown in [4, 10] that under some reasonable conditions Tikhonov regularization together with the rule in [9] indeed is order optimal. Moreover, it has also been proved in [10] that if the solution satisfies suitable source conditions, the same order optimal result is still true under merely the Lipschitz condition on \( F' \).

Iteration methods are also attractive since they are straightforward to implement for the numerical solution of nonlinear ill-posed problems. In [1], Bakushinskii proposed the iteratively regularized Gauss–Newton method in which the iterated solutions \( \{x^k\} \) are defined successively by
\[
x^k_{k+1} = x^k_k - (\alpha_k I + F'(x^k) F'(x^k)^*)^{-1} (F'(x^k)^* (F(x^k) - y^\delta) + \alpha_k (x^k_k - x_0)), \quad k \geq 0,
\]
where \( x^0_0 := x_0 \) is an initial guess of \( x^\dagger \), and \( \{\alpha_k\} \) is a given sequence of numbers such that
\[
\alpha_k > 0, \quad 1 \leq \frac{\alpha_k}{\alpha_{k+1}} \leq r \quad \text{and} \quad \lim_{k \to \infty} \alpha_k = 0
\]
for some constant \( r > 1 \). It has been shown in [1, 2] that if \( x_0 - x^\dagger = (F'(x^\dagger)^* F'(x^\dagger))^{1/2} \omega \) for some \( \omega \in X \) and \( 0 < \nu \leq 2 \), then for the stopping index \( N_\delta \) chosen by the \textit{a priori} rule
\[
\alpha_{N_\delta} \leq \left( \frac{\delta}{\| \omega \|} \right)^{2/(1+\nu)} < \alpha_k, \quad 0 \leq k < N_\delta
\]
there holds the order optimal convergence rate
\[
\| x^k_{N_\delta} - x^\dagger \| \leq C \| \omega \|^{1/(1+\nu)} \delta^{2/(1+\nu)}
\]
with some constant \( C \) independent of \( \delta \). This rule, however, depends on the knowledge on the smoothness of \( x_0 - x^\dagger \), which is difficult to check in practice. Thus a wrong guess of the smoothness will lead to a bad choice of \( N_\delta \), and consequently to a bad approximation to \( x^\dagger \). Therefore, \textit{a posteriori} rules, which use only quantities that arise during calculations, should be considered to choose the stopping index of iteration.

The generalized discrepancy principle
\[
\| F(x^k_{N_\delta}) - y^\delta \| \leq \tau \delta \leq \| F(x^k_{n_\delta}) - y^\delta \|, \quad 0 \leq k < n_\delta
\]
\( \{x^k\} \) denotes the sequence of iterates defined by
\[
x^k_{k+1} = x^k_k + \alpha_k (x^k_k - x_0) - \alpha_k F(x^k_k), \quad k \geq 0,
\]
where \( \alpha_k > 0 \) is a given sequence of numbers such that
\[
\alpha_k > 0, \quad 1 \leq \frac{\alpha_k}{\alpha_{k+1}} \leq r \quad \text{and} \quad \lim_{k \to \infty} \alpha_k = 0
\]
for some constant \( r > 1 \). It has been shown in [1, 2] that if \( x_0 - x^\dagger = (F'(x^\dagger)^* F'(x^\dagger))^{1/2} \omega \) for some \( \omega \in X \) and \( 0 < \nu \leq 2 \), then for the stopping index \( N_\delta \) chosen by the \textit{a priori} rule
\[
\alpha_{N_\delta} \leq \left( \frac{\delta}{\| \omega \|} \right)^{2/(1+\nu)} < \alpha_k, \quad 0 \leq k < N_\delta
\]
there holds the order optimal convergence rate
\[
\| x^k_{N_\delta} - x^\dagger \| \leq C \| \omega \|^{1/(1+\nu)} \delta^{2/(1+\nu)}
\]
with some constant \( C \) independent of \( \delta \). This rule, however, depends on the knowledge on the smoothness of \( x_0 - x^\dagger \), which is difficult to check in practice. Thus a wrong guess of the smoothness will lead to a bad choice of \( N_\delta \), and consequently to a bad approximation to \( x^\dagger \). Therefore, \textit{a posteriori} rules, which use only quantities that arise during calculations, should be considered to choose the stopping index of iteration.

The generalized discrepancy principle
\[
\| F(x^k_{N_\delta}) - y^\delta \| \leq \tau \delta \leq \| F(x^k_{n_\delta}) - y^\delta \|, \quad 0 \leq k < n_\delta
\]
has been considered in [2] as an \textit{a posteriori} rule for choosing the stopping index \( n_\delta \), where \( \tau > 1 \) is a sufficient large number. It has been shown that

\[
\| x_{n_\delta}^\delta - x^\dagger \| \leq O(\delta^{(1+\nu)})
\]

if \( x_0 - x^\dagger \) satisfies (1.6) with \( 0 < \nu \leq 1 \) and if \( F \) satisfies the condition

\[
F'(x) = R(x, z)F'(z) + Q(x, z),
\]

\[
\| I - R(x, z) \| \leq C_R, \quad x, z \in B_\rho(x^\dagger),
\]

\[
\| Q(x, z) \| \leq C_Q \| F'(x)(x - z) \|
\]

with \( \rho, C_Q \) and \( C_R \) being sufficiently small, where \( B_\rho(x^\dagger) \subset D(F) \) denotes a ball of radius \( \rho > 0 \) around \( x^\dagger \). However, with such \( n_\delta \), one cannot expect a better rate of convergence than \( O(\delta^{1/2}) \) even if \( x_0 - x^\dagger \) satisfies (1.6) with \( \nu > 1 \).

In order to prevent such saturation, an alternative \textit{a posteriori} rule has been suggested in [3] to choose the stopping index \( k_\delta \) as the first integer satisfying

\[
\alpha_{k_\delta} \left( F(x_{k_\delta}^\delta) - y^\delta, (\alpha_{k_\delta} I + F'(x_{k_\delta}^\delta)F'(x_{k_\delta}^\delta)^{-1}(F(x_{k_\delta}^\delta) - y^\delta) \right) \leq \tau^2 \delta^2,
\]

where \( \tau > 1 \) is a large number. It has been shown that if \( F \) satisfies the condition that there exists a constant \( K_0 \) such that

\[
F'(z) = F'(x)R(x, z) \quad \text{and} \quad \| I - R(x, z) \| \leq K_0 \| x - z \|
\]

for all \( x, z \in B_\rho(x^\dagger) \), then

\[
\| x_{k_\delta}^\delta - x^\dagger \| \leq C \| \alpha \|^{1/(1+\nu)} \delta^{(1+\nu)}
\]

as long as \( x_0 - x^\dagger \) satisfies (1.6) with \( 0 < \nu \leq 2 \). Moreover, the result in [3] implies the convergence rates under more general source conditions, thus it even applies to exponentially ill-posed problems.

Here are some comments on (1.8) and (1.10). By a first glimpse it seems they are quite similar, but in fact this is not the case. For many important inverse problems arising in medical imaging and nondestructive testing, condition (1.8) seems to be difficult to verify or even not hold, while (1.10) can often be verified easily. However, there are still some critical cases in which (1.10) is violated. Thus it would be useful to derive some conclusions on the stopping rule (1.9) under conditions different from or even weaker than (1.10).

In this paper we will establish some convergence rate results on the method defined by (1.4) and (1.9) under merely the Lipschitz condition on \( F' \). We will assume that

\[
B_\rho(x^\dagger) \subset D(F) \quad \text{for some} \ \rho > 4\tau \| x_0 - x^\dagger \|
\]

and that \( F' \) satisfies the Lipschitz condition, i.e. there exists a constant \( L \) such that

\[
\| F'(x) - F'(z) \| \leq L \| x - z \| \quad \text{for all} \ x, z \in B_\rho(x^\dagger).
\]

Moreover, we will assume that the nonlinear operator \( F \) is properly scaled, i.e.

\[
\| F'(x) \| \leq \alpha_0^{1/2}, \quad x \in B_\rho(x^\dagger).
\]

This scaling condition can always be fulfilled by multiplying the both sides of (1.1) by a sufficiently small constant, which then appears as a relaxation parameter in the iteratively regularized Gauss-Newton method.
Let $F$ satisfy (1.12), (1.13) and (1.14), let $x^*$ be determined by (1.4) with $\{a_k\}$ satisfying (1.5), and let $k_0$ be determined by (1.9) with $\tau > 1 + \sqrt{F}$. If $x_0 - x^1 = F(x^1)^* v$ for some $v \in N(F(x^1)^*)^\perp \subset Y$ and if $L \|v\| \leq \varepsilon_0$, then

$$
\|x_{k+1}^* - x^1\| \leq C \inf \left\{ \|x_k - x^1\| + \frac{\delta}{\sqrt{k}} : k = 0, 1, \ldots \right\},
$$

(1.15)

where $\varepsilon_0$ and $C$ are some positive constants depending only on $r$ and $\tau$.

In particular, if, in addition, $x_0 - x^1 = (F(x^1)^* F(x^1))^{\nu/2} \omega$ for some $\omega \in X$ and some $1 \leq v \leq 2$, then

$$
\|x_{k+1}^* - x^1\| \leq C_v \|\omega\|^{1/(1+v)} \delta^{(1+v)/2},
$$

(1.16)

where $C_v > 0$ is a constant depending only on $r, \tau$ and $v$.

The sequence $\{x_k\}$ appearing in theorem 1.1 is defined by the iteratively regularized Gauss-Newton method (1.4) corresponding to the noise-free case, that is $\{x_k\}$ is defined successively by

$$
x_{k+1} = x_k - (a_k I + F(x_k)^* F'(x_k))^{-1} [F'(x_k)^* (F(x_k) - y) + a_k (x_k - x_0)].
$$

(1.17)

Although smallness of $L \|v\|$ is not specified in the above result, all the required smallness conditions are indeed spelled out in the proof.

**Remark 1.** In [5, 6], Kaltenbacher considered a class of iterative regularization methods in which the iterated solutions $\{x_k\}$ are defined by

$$
x_{k+1} = x_0 - g_{\alpha_k} (F'(x_k)^* F'(x_k)) [F'(x_k)^* (F(x_k) - y) + \alpha_k (x_k - x_0)],
$$

(1.17)

where $\{a_k\}$ is a given sequence satisfying (1.5), and $\{g_{\alpha_k}\}$ is a family of functions satisfying suitable conditions. Convergence analysis of (1.17) has been given in [5] under *a priori* choice on the stopping index and in [6] under an *a posteriori* stopping rule which is a modified form of (1.7) and terminates the iteration as long as

$$
\max \left\{ \|F(x_k) - y\|, \|F'(x_k)\| \right\} \leq \tau \delta
$$

(1.18)

is satisfied for the first time. The order optimality of (1.18) has been established when $x_0 - x^1$ satisfies (1.6) with $0 \leq v \leq 1$ and $F$ satisfies (1.8). It is still unknown if it is possible to obtain order optimal convergence rates for $v > 1$ or under merely Lipschitz condition on $F'$. Note that the iteratively regularized Gauss–Newton method takes the form (1.17) with $g_{\alpha_k}(\lambda) = (\alpha + \lambda)^{-1}$ which has some special characters. Our argument on (1.4) and (1.9) is different from [5, 6]. The essential ingredients are proposition 2.1 and proposition 3.1 in which we employ the special properties of $(\alpha + \lambda)^{-1}$. The proof will make use of some ideas developed in [3, 4, 10].

**Remark 2.** In the method (1.17) or (1.4), the sequence $\{a_k\}$ is given a priori. This possibly could make an oversolving of (1.9) in the final step. In [7] Rieder developed an efficient inexact Newton-type method

$$
x_{k+1} = x_k^\delta - g_{\alpha_k} (F'(x_k)^* F'(x_k)) [F'(x_k)^* (F(x_k) - y) + \delta]
$$

(1.19)

in which the sequence $\{a_k\}$ is determined adaptively during computations. Under the condition $F'(z) = R(x, z) F'(x)$ and $\|I - R(x, z)\| \leq C_0 \|x - z\|$, $x, z \in B_\rho(x^1)$, (1.20) it has been proved in [8] that there is a problem-dependent number $0 \leq \eta < 1$ such that the convergence rate $O(\delta^{(2-\eta)/2})$ holds if $x_0 - x^1$ satisfies (1.6) with $\eta < v \leq 1$. Note that this is not the optimal rate. It is not yet clear if the convergence can be established under
weaker source conditions. Moreover, the condition (1.20) is indeed quite restrictive. Thus the convergence analysis is not quite complete.

This paper is organized as follows. In section 2 we first prove the convergence rate result corresponding to (1.16) with \( v = 1 \). In section 3, after deriving a stability estimate, we use the result from section 2 to establish an important inequality which connects the error terms at different levels. We then complete the proof of theorem 1.1 in section 4. Finally a numerical test is given in section 5 to test the theoretical results. For the ease of exposition, throughout this paper we will use \( C \) to denote a generic constant depending only on \( r \) and \( \tau \); we will also use the convention \( A \lesssim B \) to mean that \( A \leq CB \) for some generic constant \( C \).

2. A crucial estimate on convergence rates

Throughout this paper, we will use the notations

\[
A := F'(x^1)^* F'(x^1), \quad A_k := F'(x_k)^* F'(x_k), \quad A_k^\delta := F'(x_k^\delta)^* F'(x_k^\delta),
\]

\[
B := F'(x^1)^* F'(x^1)^*, \quad B_k := F'(x_k)^* F'(x_k)^*, \quad B_k^\delta := F'(x_k^\delta)^* F'(x_k^\delta)^*.
\]

We will also use the notations

\[
e_k := x_k - x^1, \quad e_k^\delta := x_k^\delta - x^1.
\]

The following elementary result will be used frequently.

**Lemma 2.1.** Let \( \{p_k\} \) be a sequence of positive numbers satisfying \( \frac{p_k}{p_{k+1}} \leq p \) with a constant \( p > 0 \). Suppose that the sequence \( \{\eta_k\} \) has the property \( \eta_k \leq \eta_{k+1} \leq \eta_k + \epsilon \eta_k \) for all \( k \). If \( \epsilon p < 1 \) and \( \eta_0 \leq \frac{p}{1 - \epsilon p} p_0 \), then \( \eta_k \leq \frac{p}{1 - \epsilon p} p_k \) for all \( k \).

The purpose of this section is to prove the following convergence rate result.

**Proposition 2.1.** Assume that (1.12), (1.13) and (1.14) hold and that \( x_0 - x^1 = F'(x^1)^* v \) for some \( v \in \mathcal{N}(F'(x^1)^*)^- \). Let \( \{x_k^\delta\} \) be defined by (1.4) with \( \{a_k\} \) satisfying (1.5) and let \( k_\delta \) be the integer determined by (1.9) with \( \tau > 1 + \sqrt{r} \). If \( L\|v\| \leq \epsilon_1 \), then

\[
\|x_k^\delta - x^1\| \leq C\|v\|^{1/2} \delta^{1/2},
\]

where \( \epsilon_1 \) and \( C \) are positive constants depending only on \( r \) and \( \tau \).

In order to prove proposition 2.1, we need to give an upper bound on \( k_\delta \). To this end, we introduce the integer \( \tilde{k}_\delta \) which is defined to be the first integer such that

\[
\alpha_{\tilde{k}_\delta} \leq \frac{c_0 \delta}{\|v\|} < \alpha_k, \quad 0 \leq k < \tilde{k}_\delta,
\]

where \( c_0 \) is a constant such that \( 0 < c_0 \leq \frac{\tau - 1 - \sqrt{r}}{2\sqrt{r}} \).

**Lemma 2.2.** Let all the conditions in proposition 2.1 be fulfilled. Then for all integers \( 0 \leq k \leq \tilde{k}_\delta \) there hold

\[
x_k^\delta \in B_\rho(x^1) \quad \text{and} \quad \|e_k^\delta\| \leq r^{1/2} \left( 1 + \frac{1}{c_0} \right) \alpha_k^{1/2} \|v\|.
\]

Moreover, for the integer \( k_\delta \) determined by (1.9) with \( \tau > 1 + \sqrt{r} \) there holds \( k_\delta \leq \tilde{k}_\delta \).

**Proof.** We first show that \( x_k^\delta \in B_\rho(x^1) \) for all integers \( 0 \leq k \leq \tilde{k}_\delta \). It is clear from (1.12) that this is true for \( k = 0 \). Now for any fixed integer \( 0 < l \leq \tilde{k}_\delta \), we assume that \( x_l^\delta \in B_\rho(x^1) \) for
all $0 \leq k < l$ and we are going to show $x_l \in B_p(x^1)$. To this end, from the definition of $\{x^l_k\}$ it follows that

$$
e_{k+1}^l = \alpha_k (\alpha_k I + A^l_k)^{-1} e_0 - (\alpha_k I + A^l_k)^{-1} F'(x^l_k)^* (F(x^l_k) - y^l - F'(x^l_k)e^l_k). \tag{2.3}$$

Using the condition $e_0 = F'(x^1)^* v$, we can write

$$
e_{k+1}^l = \alpha_k (\alpha_k I + A)^{-1} e_0 + \alpha_k \left[(\alpha_k I + A^l_k)^{-1} - (\alpha_k I + A)^{-1}\right] F'(x^1)^* v$$

$$- (\alpha_k I + A^l_k)^{-1} F'(x^l_k)^* (F(x^l_k) - y^l - F'(x^l_k)e^l_k).$$

Thus

$$\|\ne_{k+1}^l - \alpha_k (\alpha_k I + A)^{-1} e_0\| \leq \|\alpha_k \left[(\alpha_k I + A^l_k)^{-1} - (\alpha_k I + A)^{-1}\right] F'(x^1)^* v\|$$

$$+ \frac{1}{2\sqrt{\alpha_k}} \|F(x^l_k) - y^l - F'(x^l_k)e^l_k\|. \tag{2.4}$$

By using the Lipschitz condition (1.13) and (1.3) we have

$$\|F(x^l_k) - y^l - F'(x^l_k)e^l_k\| \leq \delta + \frac{1}{2} L \|e^l_k\|^2. \tag{2.5}$$

Moreover, note that

$$\alpha_k \left[(\alpha_k I + A^l_k)^{-1} - (\alpha_k I + A)^{-1}\right] F'(x^1)^* v$$

$$= \alpha_k (\alpha_k I + A^l_k)^{-1} (A - A^l_k) (\alpha_k I + A)^{-1} F'(x^1)^* v$$

$$= \alpha_k (\alpha_k I + A^l_k)^{-1} F'(x^l_k)^* (F(x^l_k) - F'(x^l_k))(\alpha_k I + A)^{-1} F'(x^1)^* v$$

$$+ \alpha_k (\alpha_k I + A^l_k)^{-1} (F'(x^1)^* - F'(x^l_k)^*) (F'(x^1)) (\alpha_k I + A)^{-1} F'(x^1)^* v.$$

We then use (1.13) to obtain

$$\|\alpha_k \left[(\alpha_k I + A^l_k)^{-1} - (\alpha_k I + A)^{-1}\right] F'(x^1)^* v\| \leq \frac{5}{4} L \|v\| \|e^l_k\|. \tag{2.6}$$

Combining (2.4), (2.5) and (2.6) gives

$$\|\ne_{k+1}^l - \alpha_k (\alpha_k I + A)^{-1} e_0\| \leq \frac{5}{4} L \|v\| \|e^l_k\| + \frac{1}{4\sqrt{\alpha_k}} L \|e^l_k\|^2 + \frac{\delta}{2\sqrt{\alpha_k}}.$$  

Let

$$\beta_k := \|\alpha_k (\alpha_k I + A)^{-1} e_0\|. \tag{2.7}$$

Then we have

$$\|\ne_{k+1}^l\| \leq \beta_k + \frac{5}{4} L \|v\| \|e^l_k\| + \frac{1}{4\sqrt{\alpha_k}} L \|e^l_k\|^2 + \frac{\delta}{2\sqrt{\alpha_k}}. \tag{2.8}$$

Note that for $0 \leq k < \hat{k}$ we have $\frac{1}{2\sqrt{\alpha_0}} \leq \frac{1}{2\sqrt{\alpha_k}} \|e^l_k\|$; note also that $\beta_k \leq \frac{1}{2\alpha_k} \|v\|$. We thus obtain

$$\|\ne_{k+1}^l\| \leq \left(\alpha_k \|v\| + \frac{5}{4} L \|v\| \|e^l_k\| + \frac{1}{4\sqrt{\alpha_k}} L \|e^l_k\|^2. \tag{2.9}$$

This and (1.5) imply

$$\|\ne_{k+1}^l\| \leq \frac{1}{2} \left(\alpha_k \|v\| + \frac{5}{4} L \|v\| \|e^l_k\| + \frac{1}{4\sqrt{\alpha_k}} L \|e^l_k\|^2\right)^{1/2}. \tag{2.10}$$

Note that (1.14) and $e_0 = F'(x^1)^* v$ imply $\|e^l_0\|/\sqrt{\alpha_0} \leq \|v\|$. Thus, by induction, we can show that if $\|v\|$ is so small that

$$\left[5 + r^{1/2} \left(\frac{1}{2} + \frac{1}{c_0}\right)\right] \|v\| \leq \frac{1}{r}.$$
then

$$\frac{\|e_k^2\|}{\sqrt{\alpha_k}} \leq r^{1/2} \left(1 + \frac{1}{c_0}\right) \|v\| \quad \text{for} \quad 0 \leq k \leq l. \quad (2.11)$$

Combining this with (2.8), noting that $\frac{4}{\sqrt{c_0}} \leq \frac{1}{\sqrt{c_0}} \|v\|^{1/2} \delta^{1/2}$ and using the smallness condition (2.10), we have for $0 \leq k < l$ that

$$\|e_{k+1}\| \leq \beta_k + \frac{1}{2\sqrt{c_0}} \|v\|^{1/2} \delta^{1/2} + \frac{1}{4r} e_k^2.$$ 

Recall that $\beta_k \leq r \beta_{k+1}$ which was proved in [3, lemma 3.4]. We may apply lemma 2.1 to conclude

$$\|e_k^2\| \leq 4r \beta_k + \frac{2r}{3\sqrt{c_0}} \|v\|^{1/2} \delta^{1/2},$$

since, due to (1.14), this is true for $k = 0$. Note that $\beta_k \leq \|e_0\| \leq \frac{\rho}{c_0}$, the above inequality implies that $x_k^3 \in B_{\rho}(x^1)$ for all $0 \leq k \leq l$. We thus obtain $x_k^3 \in B_{\rho}(x^1)$ for all $0 \leq k \leq k_3$. In the meanwhile, (2.11) gives the desired estimates in (2.2).

In order to prove $k_3 \leq k_3$, we use (1.13) and (2.1) to obtain

$$\|\alpha_k^{1/2} (\alpha_k I + B_{k}^1)^{-1/2} (F(x_k^3) - y^3)\|$$

$$\leq \delta + \|\alpha_k^{1/2} (\alpha_k I + B_{k}^1)^{-1/2} F'(x_k^3) e_k^3\| + \|F(x_k^3) - F'(x_k^3) e_k^3\|$$

$$\leq \delta + \alpha_k^{1/2} \|e_k^3\| + \frac{1}{2} L \|e_k^3\|^2$$

$$\leq \delta + \rho^{1/2} \left[1 + \frac{1}{c_0}\right] \|v\| \|\alpha_k\| + \frac{1}{2} r \left(1 + \frac{1}{c_0}\right)^2 L \|v\|^2 \|\alpha_k\|$$

$$\leq \delta + \rho^{1/2} \left[1 + \frac{1}{c_0}\right] c_0 \delta + \frac{1}{2} r \left(1 + \frac{1}{c_0}\right)^2 c_0 L \|v\| \delta.$$ 

Recall that $r > 1 + \sqrt{r}$ and $0 < c_0 \leq \frac{\sqrt{1+\sqrt{r}}}{2\sqrt{\rho}}$. If $L \|v\|$ is so small that

$$r^{1/2} \left[1 + \frac{1}{c_0}\right]^2 L \|v\| < 2, \quad (2.12)$$

then there holds

$$\|\alpha_k^{1/2} (\alpha_k I + B_{k}^1)^{-1/2} (F(x_k^3) - y^3)\| \leq \tau \delta.$$ 

By the definition of $k_3$, we thus conclude that $k_3 \leq k_3$. \qed

**Proof of proposition 2.1.** Recall that in the proof of lemma 2.2 we have obtained the following two estimates,

$$\|e_{k+1}\| \leq \beta_k + \frac{1}{2\sqrt{c_0}} \|v\|^{1/2} \delta^{1/2} + \frac{1}{4r} e_k^2 \quad (2.13)$$

and

$$\|e_k^2\| \leq 4r \beta_k + \frac{2r}{3\sqrt{c_0}} \|v\|^{1/2} \delta^{1/2} \quad (2.14)$$

for all $0 \leq k < k_3$. 


Now we set
\[ \beta_k^\delta := \|a_k(a_kI + A_k^\delta)^{-1}e_0\|. \]
Then it follows from \( e_0 = F'(x^\delta)^*v \) and (2.6) that
\[ |\beta_k - \beta_k^\delta| \leq \|a_k[(a_kI + A_k^\delta)^{-1} - (a_kI + A)^{-1}]e_0\| \leq \frac{2}{5}L\|v\|\|e_k^\delta\|. \]
By assuming that \( 10rL\|v\| \leq 1 \), this together with (2.13) and (2.14) implies
\[ \|e_k^\delta\| \leq \frac{8}{5}r\beta_k^\delta + C\|v\|^{1/2}\delta^{1/2}, \]
\[ \|e_{k+1}^\delta\| \leq \frac{8}{5}\beta_k^\delta + C\|v\|^{1/2}\delta^{1/2}, \]
(2.15)
(2.16)
We need to estimate \( \beta_k^\delta \). We first have
\[ (\beta_k^\delta)^2 = (a_k(a_kI + A_k^\delta)^{-1}e_0, a_k(a_kI + A_k^\delta)^{-1}F'(x^\delta)^*v) \]
\[ = (a_k(a_kI + A_k^\delta)^{-1}e_0, a_k(a_kI + A_k^\delta)^{-1}[F'(x_k^\delta)^* + (F'(x^\delta)^* - F'(x_k^\delta)^*)]v) \]
\[ = (a_k^{1/2}(a_kI + B_k^\delta)^{-1/2}F'(x_k^\delta)e_0, a_k^{1/2}(a_kI + B_k^\delta)^{-1/2}v) \]
\[ + (a_k(a_kI + A_k^\delta)^{-1}e_0, a_k(a_kI + A_k^\delta)^{-1}[F'(x^\delta)^* - F'(x_k^\delta)^*]v) \]
\[ \leq \gamma_k^\delta\|v\| + \beta_k^\deltaL\|v\|\|e_k^\delta\|, \]
where
\[ \gamma_k^\delta := \|a_k^{1/2}(a_kI + B_k^\delta)^{-1/2}F'(x_k^\delta)e_0\|. \]
Therefore
\[ \beta_k^\delta \leq \sqrt{\gamma_k^\delta\|v\|^{1/2} + L\|v\|\|e_k^\delta\|}. \]
(2.17)
In order to estimate \( \gamma_k^\delta \), we observe that (2.3) implies
\[ a_k^{1/2}(a_kI + B_k^\delta)^{-1/2}F'(x_k^\delta)e_0 = a_k^{1/2}(a_kI + B_k^\delta)^{-1/2}B_k^\delta(F(x_k^\delta) - y^\delta - F'(x_k^\delta)e_k^\delta) + a_k^{1/2}(a_kI + B_k^\delta)^{-1/2}F'(x_k^\delta)e_{k+1}^\delta. \]
Thus
\[ \gamma_k^\delta \leq \|a_k^{1/2}(a_kI + B_k^\delta)^{-1/2}F'(x_k^\delta)e_{k+1}^\delta\| + \|F(x_k^\delta) - y^\delta - F'(x_k^\delta)e_k^\delta\| \]
\[ \leq \|a_k^{1/2}(a_kI + B_k^\delta)^{-1/2}(F(x_{k+1}^\delta) - y^\delta)\| + \|F'(x_{k+1}^\delta) - F'(x_k^\delta)\|e_{k+1}^\delta\| \]
\[ + \|F(x_k^\delta) - y - F'(x_k^\delta)e_k^\delta\| + \|F(x_k^\delta) - y - F'(x_k^\delta)e_k^\delta\| + 2\delta. \]
It then follows from (1.13) that
\[ \gamma_k^\delta \leq \|a_k^{1/2}(a_kI + B_k^\delta)^{-1/2}(F(x_{k+1}^\delta) - y^\delta)\| + 2\delta + L\|e_k^\delta\|^2 + 2L\|e_{k+1}^\delta\|^2 . \]
Now we assume further that
\[ 3r^{1/2}\left(1 + \frac{1}{c_0}\right)L\|v\| \leq 1. \]
(2.18)
Then by (2.2) we have
\[ L\|x_{k+1}^\delta - x_k^\delta\| \leq L\left(\|e_k^\delta\| + \|e_{k+1}^\delta\|\right) \leq 2r^{1/2}\left(1 + \frac{1}{c_0}\right)L\|v\|\|a_k^{1/2}\| \leq \frac{2}{3}\alpha_k^{1/2}. \]
Thus we can conclude, by using \[10, \text{proposition 3.4}\], that there hold
\[
\|a_k^{1/2}(a_kI + B_k^I)^{-1/2}(F(x_k^I) - y^I)\| \lesssim \|a_k^{1/2}(a_kI + B_k^I)^{-1/2}(F(x_k^I) - y^I)\|
\]
\[
\lesssim \|a_k^{1/2}(a_kI + B_k^I)^{-1/2}(F(x_k^I) - y^I)\|.
\]
Therefore
\[
\gamma_k^{\delta} \leq C \|a_k^{1/2}(a_kI + B_k^I)^{-1/2}(F(x_k^I) - y^I)\| + 2\delta + L\|e_k^I\|^2 + 2L\|e_k^I\|^2.
\]
This together with (2.17) gives
\[
\beta_k^{\delta} \leq C \|a_k^{1/2}(a_kI + B_k^I)^{-1/2}(F(x_k^I) - y^I)\|^{1/2}\|v\|^{1/2} + C\|v\|^{1/2}\delta^{1/2} + 2\sqrt{L\|v\|}(\|e_k^I\| + \|e_k^I\|).
\]
Combining this with (2.15) and (2.16) yields
\[
\beta_k^{\delta} \leq C \|a_k^{1/2}(a_kI + B_k^I)^{-1/2}(F(x_k^I) - y^I)\|^{1/2}\|v\|^{1/2} + C\|v\|^{1/2}\delta^{1/2} + \frac{16}{2\tau}(1 + r)\sqrt{L\|v\|}\beta_k^{\delta}.
\]
If we assume further that
\[
\frac{1}{4\tau}(1 + r)\sqrt{L\|v\|} \leq 1,
\]
then we can obtain
\[
\beta_k^{\delta} \lesssim a_k^{-1/2}(a_kI + B_k^I)^{-1/2}(F(x_k^I) - y^I)\|^{1/2}\|v\|^{1/2} + \|v\|^{1/2}\delta^{1/2}.
\]
It then follows from (2.16) that for all \(0 < k \leq k_0\) holds
\[
\|e_k^I\| \lesssim a_k^{-1/2}(a_kI + B_k^I)^{-1/2}(F(x_k^I) - y^I)\|^{1/2}\|v\|^{1/2} + \|v\|^{1/2}\delta^{1/2}.
\]
Thus, by setting \(k = k_0\) in the above inequality and using the definition of \(k_0\), we obtain the desired estimate. \(\square\)

3. A key inequality

The main result of this section is the following inequality.

**Proposition 3.1.** Assume that (1.12), (1.13) and (1.14) hold and that \(x_0 - x^I = F'(x^I)^*v\) for some \(v \in \mathcal{N}(F'(x^I)^*)^\perp\). Then for any integer \(k \geq k_0\) there holds
\[
\|e_k\| \lesssim \|e_k\| + \frac{a_k^{-1/2}(a_kI + B_k)^{-1/2}(F(x_k^I) - y)}{\sqrt{a_k}} + \frac{\delta}{\sqrt{a_k}}
\]
if \(L\|v\| \leq \varepsilon_2\) for a small number \(\varepsilon_2 > 0\) depending only on \(r\) and \(\tau\).

The proof of this result will employ proposition 2.1 and the following two auxiliary results which are of independent interest.

**Lemma 3.1.** Let all the conditions in proposition 3.1 be fulfilled. Then for all \(k \geq 0\) there hold
\[
x_k \in B_\rho(x^I) \quad \text{and} \quad \|e_k\| \leq r^{1/2}a_k^{1/2}\|v\|.
\]
Moreover, for all integer \( k \geq 0 \) there hold
\[
\frac{2}{3} \beta_k \leq \|e_k\| \leq \frac{4}{3} \beta_k \quad \text{and} \quad \frac{1}{2r} \|e_k\| \leq \|e_{k+1}\| \leq 2 \|e_k\|, \tag{3.2}
\]
where \( \beta_k \) is defined as in (2.7).

**Proof.** From the definition of \( \{x_k\} \) it follows easily that
\[
e_{k+1} = \alpha_k (\alpha_k I + A_k)^{-1} e_0 - (\alpha_k I + A_k)^{-1} F'(x_k)^* (F(x_k) - y - F'(x_k) e_k). \tag{3.3}
\]
By the smallness condition (2.10), then we can use the similar argument in the proof of lemma 2.2 to conclude (3.1) and the estimate
\[
\|e_{k+1}\| - \beta_k \leq \|e_{k+1}\| - \alpha_k (\alpha_k I + A)^{-1} e_0 \leq \frac{1}{4r} \|e_k\| \tag{3.4}
\]
Thus, by lemma 2.1, we have
\[
\|e_k\| \leq \frac{4}{3} \beta_k.
\]
Note that \( \beta_k \) is non-increasing, we can use (3.4) again to obtain
\[
\|e_{k+1}\| \geq \beta_k - \frac{1}{4r} \|e_k\| \geq \frac{2}{3} \beta_k - \frac{2}{3} \beta_{k+1}.
\]
Therefore \( \frac{2}{3} \beta_k \leq \|e_k\| \leq \frac{4}{3} \beta_k \). This together with (3.4) then implies
\[
\frac{2}{3} \beta_k \leq \|e_{k+1}\| \leq \beta_k + \frac{1}{4r} \|e_k\| \leq \frac{3}{2} \|e_k\| + \frac{1}{4r} \|e_k\| \leq 2 \|e_k\|
\]
and
\[
\|e_{k+1}\| \geq \beta_k - \frac{1}{4r} \|e_k\| \geq \frac{3}{4} \|e_k\| - \frac{1}{4r} \|e_k\| = \frac{1}{2r} \|e_k\|.
\]
The proof is complete. \( \square \)

**Lemma 3.2.** Let all the conditions in proposition 3.1 be fulfilled. Then for all integers \( 0 \leq k \leq \tilde{k}_\delta \) there hold
\[
\|x_k - x^\delta\| \leq \frac{\delta}{\sqrt{\alpha_k}}. \tag{3.5}
\]

**Proof.** By setting
\[
u_k := F(x_k) - y - F'(x_k) e_k, \quad u_k^\delta := F(x_k^\delta) - y - F'(x_k^\delta) e_k^\delta,
\]
it then follows from (2.3) and (3.3) that
\[
x_k^{\delta+1} - x_k = I_1 + I_2 + I_3 + I_4,
\]
where
\[
I_1 := \alpha_k \left[ (\alpha_k I + A_k^\delta)^{-1} - (\alpha_k I + A_k)^{-1} \right] e_0,
I_2 := (\alpha_k I + A_k^\delta)^{-1} F'(x_k^\delta)^* (y^\delta - y),
I_3 := \left[ (\alpha_k I + A_k)^{-1} F'(x_k)^* - (\alpha_k I + A_k^\delta)^{-1} F'(x_k^\delta)^* \right] u_k^\delta,
I_4 := (\alpha_k I + A_k)^{-1} F'(x_k)^* (u_k - u_k^\delta).
\]
It is clear that \( \|I_2\| \leq \frac{\delta}{\sqrt{\alpha_k}} \). In order to estimate \( I_1, I_3 \) and \( I_4 \), we assume further that \( L\|v\| \) is so small that
\[
\left[ r^{1/2} \left( 1 + \frac{1}{c_0} \right) + 3 r^{1/2} \right] L\|v\| \leq \frac{2}{5}.
\]

For the term $I_1$, recall that $e_0 = F'(x^*) v$, we can write

$$I_1 = \alpha_k \left[ (\alpha_k I + A_k^T)^{-1} - (\alpha_k I + A_k)^{-1} \right] F'(x_k)^* v$$

$$+ \alpha_k \left[ (\alpha_k I + A_k^T)^{-1} - (\alpha_k I + A_k)^{-1} \right] (F'(x_k)^* - F'(x_k^*) v$$

We may use (1.13), the similar argument in deriving (2.6), and lemma 3.1 to obtain

$$\| I_1 \| \leq \frac{5}{4} L \| v \| \| x_k^\delta - x_k \| + \frac{1}{\sqrt{\alpha_k}} L^2 \| v \| \| e_k \| \| x_k^\delta - x_k \|$$

Similarly, for $I_3$ we have

$$\| I_3 \| \leq \frac{9}{4} \alpha_k^{-1} L \| u_k \| \| x_k^\delta - x_k \| \leq \frac{9}{8} \alpha_k^{-1} L \| e_k \| \| x_k^\delta - x_k \|$$

By using (1.13), lemma 3.1 and lemma 2.2 we have

$$\| I_k \| \leq \frac{1}{\sqrt{\alpha_k}} \| u_k^\delta - u_k \| \leq \frac{3}{4} \alpha_k^{-1} L \| e_k \| \| x_k^\delta - x_k \| \leq \frac{1}{16} \| x_k^\delta - x_k \|$$

Combining the above estimates on $I_1, I_2, I_3$ and $I_4$ we conclude for all $0 \leq k < \hat{k}$ that

$$\| x_{k+1}^\delta - x_{k+1} \| \leq \delta \frac{\sqrt{\alpha_k}}{2} + \frac{1}{2} \| x_k^\delta - x_k \|$$

An application of lemma 2.1 gives the desired estimate. □

Now we are in a position to prove proposition 3.1.

Proof of proposition 3.1. Let

$$J := (\alpha_{k-1}(\alpha_k I + A)^{-1} - (\alpha_{k-1} I + A)^{-1}) e_0.$$

Then it follows from (3.3), (1.13), and lemma 3.1 that

$$\| x_k - x_k \| \leq \| J \| + \| (\alpha_{k-1}(\alpha_k I + A_k)^{-1} - (\alpha_{k-1} I + A)^{-1}) \| F'(x_k)^* v \|

+ \| (\alpha_{k-1}(\alpha_k I + A_k^T)^{-1} - (\alpha_{k-1} I + A_k)^{-1}) \| F'(x_k^*) v \|

+ \frac{1}{2 \sqrt{\alpha_k}} \| F(x_{k-1}) - y - F'(x_{k-1}) e_{k-1} \|

+ \frac{1}{2 \sqrt{\alpha_k}} \| F(x_{k-1}) - y - F'(x_{k-1}) e_{k-1} \|$$
where we assume that $L\|v\|$ satisfies

$$r(5 + r^{1/2}/L)\|v\| \leq \frac{4}{7}. \tag{3.7}$$

In order to estimate $J$, we write

$$J = J_1 + J_2 + J_3,$$

where

$$J_1 := \left(1 - \frac{\alpha_{k_1 - 1}}{\alpha_{k_1 - 1}}\right) (\alpha_{k_1 - 1} I + A)^{-1} F' \left(x^1 \right)^* F \left(x_{k_1} \right) - y),$$

$$J_2 := \left(1 - \frac{\alpha_{k_2 - 1}}{\alpha_{k_2 - 1}}\right) (\alpha_{k_2 - 1} I + A - 1) F' \left(x^1 \right)^* F \left(x_{k_2} \right) + y),$$

$$J_3 := \left(1 - \frac{\alpha_{k_3 - 1}}{\alpha_{k_3 - 1}}\right) (\alpha_{k_3 - 1} I + A) - 1 A \left[\alpha_{k_3 - 1} (\alpha_{k_3 - 1} I + A) - 1 e_0 - e_{k_3}.\right.$$}

By using the argument in the proof of [3, lemma 4.4] we can see that

$$\|J_1\| \leq \frac{1}{\sqrt{\alpha_{k_1}}} \|\alpha_{k_1} I + B\|^{-1/2} \|F \left(x_{k_1} \right) - y\|.$$

Also, by using (1.13) and noting $\alpha_{k_2 - 1} \leq \alpha_{k_2 - 1}$, it is easy to see that

$$\|J_2\| \leq \frac{1}{4\sqrt{\alpha_{k_2 - 1}}} \|\alpha_{k_2 - 1} I + A\|^{-1} \|F \left(x_{k_2} \right) - y\|.$$

From proposition 2.1 and lemma 3.2 we have

$$\|e_{k_3}\| \leq \|e_{k_1}\| + \|x_{k_3} - x_{k_1}\| \leq \|v\|^{1/2} \delta^{1/2} + \frac{\delta}{\sqrt{\alpha_{k_3}}}.$$

Recall that $k_3 \leq \tilde{k}_3$ which implies $\frac{\delta}{\sqrt{\alpha_{k_3}}} \leq \|v\|^{1/2} \delta^{1/2}$. Thus

$$\|e_{k_3}\| \leq \|v\|^{1/2} \delta^{1/2}.$$

Therefore

$$\|J_3\| \leq \frac{1}{\sqrt{\alpha_{k_3}}} \|\alpha_{k_3} I + B\|^{-1/2} \|F \left(x_{k_3} \right) - y\| \leq \frac{\delta}{\sqrt{\alpha_{k_3}}}.$$

By using (3.4) and (3.2), $J_3$ can be estimated as

$$\|J_3\| \leq \|\alpha_{k_3 - 1} (\alpha_{k_3 - 1} I + A) - 1 e_0 - e_{k_3}\| \leq \frac{1}{4\alpha_{k_3}} \|e_{k_3 - 1}\| \leq \frac{1}{2} \|e_{k_3}\|.$$

Combining the above estimates on $J_1$, $J_2$, and $J_3$, we obtain

$$\|J\| \leq C \left[\alpha_{k_1} \left(\alpha_{k_1} I + B\right)^{-1/2} \|F \left(x_{k_1} \right) - y\| \right] + \frac{\delta}{\sqrt{\alpha_{k_3}}} + \frac{1}{2} \|e_{k_3}\|.$$

This together with (3.6) gives

$$\|x_{k_3} - x_{k}\| \leq C \left[\alpha_{k_2} \left(\alpha_{k_2} I + B\right)^{-1/2} \|F \left(x_{k_2} \right) - y\| \right] + \frac{\delta}{\sqrt{\alpha_{k_3}}} + \frac{1}{2} \|e_{k_3}\| + \frac{2}{5} \|e_{k}\| + \frac{9}{10} \|e_{k}\|.$$

The desired assertion thus follows. \qed
4. Proof of theorem 1.1

In this section we will complete the proof of the main result, theorem 1.1. In order to apply proposition 3.1, we need the following estimates.

**Lemma 4.1.** Assume that (1.12), (1.13) and (1.14) hold and that \( x_0 - x^\dagger = F'(x^\dagger)^* v \) for some \( v \in \mathcal{N}(F'(x^\dagger)^*)^\perp \). Let \( k_3 \) be the integer determined by (1.9) with \( \tau > 1 + \sqrt{r} \). If \( L \|v\| \leq \varepsilon_3 \) for some number \( \varepsilon_3 > 0 \) depending only on \( r \) and \( \tau \), then we have

\[
\left\| \alpha_k (\alpha_k I + B)^{-1/2} (F(x_k) - y) \right\| \lesssim \delta
\]

and

\[
\delta \lesssim \left\| \alpha_k^{1/2} (\alpha_k I + B)^{-1/2} (F(x_k) - y) \right\|
\]

for all \( 0 \leq k < k_3 \).

**Proof.** For \( 0 \leq k \leq k_3 \) we set

\[
a_k := \left\| \alpha_k^{1/2} (\alpha_k I + B)^{-1/2} (F(x_k) - y) \right\|^2,
\]

\[
b_k := \left\| \alpha_k^{1/2} (\alpha_k I + B_k^\dagger)^{-1/2} (F(x_k) - y) \right\|^2.
\]

It then follows from [10, proposition 3.4], lemma 2.2, and the smallness condition (2.10) that

\[
|a_k - b_k| \leq \frac{1}{\sqrt{\alpha_k}} L \left\| \alpha_k^2 \right\| (a_k + b_k) \leq r^{1/2} \left( 1 + \frac{1}{c_0} \right) L \|v\| (a_k + b_k) \leq \frac{1}{r} (a_k + b_k).
\]

This implies \( b_k \lesssim a_k \lesssim b_k \). Thus it suffices to show that

\[
\sqrt{b_k} \sim \delta \quad \text{and} \quad \delta \lesssim \sqrt{b_k} \quad \text{for} \quad 0 \leq k < k_3.
\]

By assuming that \( L \|v\| \leq (\tau - 2)c_0 \), and using (1.13), lemma 3.2 and (2.1) we have for \( 0 \leq k < k_3 \)

\[
\sqrt{b_k} \geq \left\| \alpha_k^{1/2} (\alpha_k I + B_k^\dagger)^{-1/2} (F(x_k^\dagger) - y^\dagger) \right\| \geq \left\| F(x_k) - F(x_k^\dagger) - F'(x_k^\dagger)(x_k - x_k^\dagger) \right\| - \left\| \alpha_k^{1/2} (\alpha_k I + B_k^\dagger)^{-1/2} F'(x_k^\dagger)(x_k - x_k^\dagger) \right\| - \delta
\]

\[
\geq (\tau - 1)\delta - \alpha_k^{1/2} \left\| x_k - x_k^\dagger \right\| - \frac{1}{2} L \left\| x_k - x_k^\dagger \right\|^2
\]

\[
\geq (\tau - 2)\delta - \frac{1}{2} L \delta^2 \geq \left( \tau - 2 - \frac{1}{2c_0} L \|v\| \right) \delta
\]

\[
\geq \frac{\tau - 2}{2} \delta.
\]

Similarly, we have

\[
\sqrt{b_k} \leq (\tau + 1)\delta + \alpha_k^{1/2} \left\| x_k - x_k^\dagger \right\| + \frac{1}{2} L \left\| x_k^\dagger - x_k \right\|^2 \lesssim \delta + \frac{\delta^2}{\alpha_k} \lesssim \delta.
\]

The proof is thus complete. \( \square \)
Proof of theorem 1.1. We first prove (1.15). Note that for $k \geq k_3$, we have from lemma 3.2, proposition 3.1 and lemma 4.1 that
\[
\|e_{k}\| \lesssim \|e_k\| + \frac{\delta}{\sqrt{\alpha_k}} \\
\lesssim \|e_k\| + \frac{\|a_{k}^{1/2}(a_{k-1}I + B)^{-1/2}F(x_k) - y\|}{\sqrt{\alpha_k}} + \frac{\delta}{\sqrt{\alpha_k}} \approx \|e_k\| + \frac{\delta}{\sqrt{\alpha_k}},
\]
while for $0 \leq k < k_3$ we have from lemma 3.1 and lemma 4.1 that
\[
\|e_{k}\| \lesssim \|e_k\| + \frac{\delta}{\sqrt{\alpha_k}} \\
\lesssim \|e_k\| + \frac{\|a_{k}^{1/2}(a_{k-1}I + B)^{-1/2}F(x_k) - y\|}{\sqrt{\alpha_k}} \approx \|e_k\| + \frac{\delta}{\sqrt{\alpha_k}}.
\]

The proof of (1.15) is complete.

Next we prove (1.16). Note that $\beta_k \leq \alpha_k^{1/2}\|\omega\|$ under the condition on $x_0 - x^\dagger$. We have from lemma 3.1 that $\|e_k\| \lesssim \alpha_k^{1/2}\|\omega\|$. Thus it follows from (1.15) that
\[
\|e_{k}\| \lesssim \inf \left\{ \alpha_k^{1/2}\|\omega\| + \frac{\delta}{\sqrt{\alpha_k}} : k = 0, 1, \ldots \right\}.
\]

Now we introduce the integer $\hat{k}_3$ such that
\[
\alpha_\hat{k}_3 \leq \left( \frac{\delta}{\|\omega\|} \right)^{2/(1+\nu)} < \alpha_k, \quad 0 \leq k < \hat{k}_3.
\]

Then it is readily seen that
\[
\|e_{k}\| \lesssim \alpha_k^{1/2}\|\omega\| + \frac{\delta}{\sqrt{\alpha_k}} \lesssim \|\omega\|^{1/(1+\nu)}\delta^{\nu/(1+\nu)}
\]
which gives the desired convergence rates.

5. A numerical test

In this section we present a numerical example to test our convergence rate result given in theorem 1.1 for the stopping rule (1.9) by considering the identification of the coefficient $c$ in the two-point boundary value problem
\[
\begin{cases}
-u'' + cu = f, & t \in (0, 1) \\
u(0) = 0, & u(1) = 1
\end{cases}
\]
from the measurement data $u^\delta$ of the state variable $u$, where $f \in L^2[0, 1]$ is given. This is exactly the example given at the beginning of this paper with $n = 1$ and $\Omega = (0, 1)$. Thus the nonlinear operator $F : D(F) \subset L^2[0, 1] \mapsto L^2[0, 1]$ is defined as the parameter-to-solution mapping $F(c) = u(c)$ with $u(c)$ being the unique solution of (5.1). $F$ is well defined on

$$D(F) := \{ c \in L^2[0, 1] : ||c - \hat{c}||_{L^2} \leq \gamma \text{ for some } \hat{c} \geq 0 \ a.e. \}$$

with some $\gamma > 0$. Moreover, $F$ is Fréchet differentiable, the Fréchet derivative and its adjoint are given by

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

where $A(c) : H^2 \cap H^1_0 \mapsto L^2$ is defined by $A(c)u = -u'' + cu$. Since $u(0) = 0$ is assumed in (5.1), we are not able to verify either (1.8) or (1.10). However, it is easy to see that $F'$ satisfies the Lipschitz condition (1.13) locally.

In our numerical experiment, we estimate $c$ in (5.1) by assuming $f(t) = t$. If $u(c^\dagger) = t$, then $c^\dagger = 1$ is a true solution. In our computation, instead of $u(c^\dagger)$, we use the special perturbation

$$u^\delta = t + \delta \sqrt{\frac{2}{\pi}} \cos(10\pi t)$$

with high frequency. Clearly $||u^\delta - u(c^\dagger)||_{L^2[0, 1]} = \delta$. As a first guess we choose

$$c_0 = 1 + (15t^2 - 10t^4 + 3t^6),$$

One can show that $c_0 - c^\dagger \in R(F(c^\dagger)^*F'(c^\dagger))$. Thus, according to theorem 1.1, the expected rate of convergence should be $O(\delta^{3/2})$. During the computation, all differential equations are solved approximately by the finite difference method by dividing the interval $[0, 1]$ into $n + 1$ subintervals with equal length $h = 1/(n + 1)$; we take $n = 199$ in our actual computation.

In table 1 we summarized the numerical results obtained by the method defined by (1.4) and the stopping rule (1.9) with $\alpha_k = 1.0 \times (1.5)^{1-k}$ and $\tau = 2.5$. In order to indicate the dependence of the convergence rates on the noise level, different values of $\delta$ are selected. The rates in table 1 coincide with theorem 1.1 very well.

As a comparison, in table 1 we also include the computational results by the method defined by (1.4) and the discrepancy principle (1.7) with the same sequence $\{\alpha_k\}$ and $\tau$. It clearly indicates that $O(\delta^{1/2})$ is the best possible rate for (1.7).

**Acknowledgments**

The author wishes to thank one of the referees for his careful reading and valuable comments on the original manuscript.
References

[1] Bakushinskii A B 1992 The problems of the convergence of the iteratively regularized Gauss–Newton method
Comput. Math. Math. Phys. 32 1353–9

[2] Blaschke B, Neubauer A and Scherzer O 1997 On convergence rates for the iteratively regularized Gauss–
Newton method IMA J. Numer. Anal. 17 421–36

[3] Jin Q N 2000 On the iteratively regularized Gauss–Newton method for solving nonlinear ill-posed problems
Math. Comp. 69 1603–23

[4] Jin Q N and Hou Z Y 1999 On an a posteriori parameter choice strategy for Tikhonov regularization of nonlinear
ill-posed problems Numer. Math. 83 139–59

[5] Kaltenbacher B 1997 Some Newton-type methods for the regularization of nonlinear illposed problems Inverse
Problems 13 729–53

[6] Kaltenbacher B 1998 A posteriori choice strategies for some Newton type methods for the regularization of
nonlinear ill-posed problems Numer. Math. 79 501–28

[7] Rieder A 1999 On the regularization of nonlinear ill-posed problems via inexact Newton iterations Inverse
Problems 15 369–27

[8] Rieder A 2001 On convergence rates of inexact Newton regularizations Numer. Math. 88 347–65

[9] Scherzer O, Engl H W and Kunisch K 1993 Optimal a posteriori parameter choice for Tikhonov regularization
for solving nonlinear ill-posed problems SIAM J. Numer. Anal. 30 1796–838

[10] Tautenhahn U and Jin Q N 2003 Tikhonov regularization and a posteriori rules for solving nonlinear ill posed
problems Inverse Problems 19 1–21