

ON THE ITERATIVELY REGULARIZED  
GAUSS-NEWTON METHOD  
FOR SOLVING NONLINEAR ILL-POSED PROBLEMS

JIN QI-NIAN

ABSTRACT. The iteratively regularized Gauss-Newton method is applied to compute the stable solutions to nonlinear ill-posed problems  $F(x) = y$  when the data  $y$  is given approximately by  $y^\delta$  with  $\|y^\delta - y\| \leq \delta$ . In this method, the iterative sequence  $\{x_k^\delta\}$  is defined successively by

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

where  $x_0^\delta := x_0$  is an initial guess of the exact solution  $x^\dagger$  and  $\{\alpha_k\}$  is a given decreasing sequence of positive numbers admitting suitable properties. When  $x_k^\delta$  is used to approximate  $x^\dagger$ , the stopping index should be designated properly. In this paper, an a posteriori stopping rule is suggested to choose the stopping index of iteration, and with the integer  $k_\delta$  determined by this rule it is proved that

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

with a constant  $C$  independent of  $\delta$ , where  $x_k$  denotes the iterative solution corresponding to the noise free case. As a consequence of this result, the convergence of  $x_{k_\delta}^\delta$  is obtained, and moreover the rate of convergence is derived when  $x_0 - x^\dagger$  satisfies a suitable “source-wise representation”. The results of this paper suggest that the iteratively regularized Gauss-Newton method, combined with our stopping rule, defines a regularization method of optimal order for each  $0 < \nu \leq 1$ . Numerical examples for parameter estimation of a differential equation are given to test the theoretical results.

1. INTRODUCTION

Nonlinear inverse problems exist in a wide variety of problems in science and engineering, and many examples can be found in the monographs and surveys by Tikhonov and Arsenin [21], Hofmann [12], Banks and Kunisch [2], Engl [5], Groetsch [10], and Vasin and Ageev [23]. Such problems can be written as the operator equations

$$(1) \quad F(x) = y,$$

where  $F$  is a continuous and Fréchet differentiable nonlinear operator with domain  $D(F)$  in the real Hilbert space  $X$  and with its range  $R(F)$  in the real Hilbert space  $Y$ , and  $y$  is attainable, i.e.  $y \in R(F)$ . We call problem (1) ill-posed if its solution

---

Received by the editor March 17, 1998 and, in revised form, January 4, 1999.

1991 *Mathematics Subject Classification*. Primary 65J20, 45G10.

*Key words and phrases*. Nonlinear ill-posed problems, the iteratively regularized Gauss-Newton method, stopping rule, convergence, rates of convergence.

does not depend continuously on the right hand side  $y$ , which is often obtained by measurement and hence contains error. Let us assume that  $y^\delta$  is an approximate data of  $y$  and

$$(2) \quad \|y^\delta - y\| \leq \delta$$

with a given noise level  $\delta > 0$ . Then the computation of the stable solution of (1) from  $y^\delta$  becomes an important topic of ill-posed problems, and the regularization techniques have to be taken into account.

Tikhonov regularization is one of the best-known methods for solving nonlinear ill-posed problems, and it has received a lot of attention in recent years [20, 7, 19, 13]. In this method, the solution  $x_\alpha^\delta$  of the minimization problem

$$(3) \quad \min_{x \in D(F)} \{ \|F(x) - y^\delta\|^2 + \alpha \|x - x_0\|^2 \}$$

is used to approximate the solution of (1), where  $\alpha > 0$  is the regularization parameter and  $x_0$  is an a priori guess of the desired solution  $x^\dagger$  of (1). Iterative approaches are attractive alternatives to Tikhonov regularization, and some of them, for instance, Landweber iteration [11] and the steepest descent method [18], have been suggested to solve nonlinear ill-posed problems. In 1992, Bakushinskii [1] proposed the following iterative approach, namely, the iteratively regularized Gauss-Newton method

$$(4) \quad x_{k+1}^\delta = x_k^\delta - (\alpha_k I + F'(x_k^\delta)^* F'(x_k^\delta))^{-1} (F'(x_k^\delta)^* (F(x_k^\delta) - y^\delta) + \alpha_k (x_k^\delta - x_0))$$

with an initial guess  $x_0^\delta := x_0 \in D(F)$  to obtain the stable approximate solutions to nonlinear ill-posed problems, where  $\{\alpha_k\}$  is a sequence satisfying

$$(5) \quad \alpha_k > 0, \quad 1 \leq \frac{\alpha_k}{\alpha_{k+1}} \leq r \quad \text{and} \quad \lim_{k \rightarrow \infty} \alpha_k = 0$$

for some constant  $r > 1$ ,  $F'(x)$  is the Fréchet derivative of  $F$  at  $x \in D(F)$  and  $F'(x)^*$  is the adjoint of  $F'(x)$ . For some background on this method, please refer to [1, 23]. The convergence of this method has been considered in several papers [1, 3, 23] under certain conditions on  $F$ , and the rates of convergence have been derived by enforcing some conditions on  $x_0 - x^\dagger$ . It has been shown that if there exist a  $0 < \nu \leq 1$  and an element  $\omega \in \mathcal{N}(F'(x^\dagger))^\perp \subset X$  such that

$$(6) \quad x_0 - x^\dagger = (F'(x^\dagger)^* F'(x^\dagger))^\nu \omega,$$

then by choosing the integer  $N_\delta$  such that

$$\alpha_{N_\delta}^{\nu+\frac{1}{2}} \leq \frac{\delta}{\|\omega\|} < \alpha_k^{\nu+\frac{1}{2}}, \quad 0 \leq k < N_\delta,$$

the rate of convergence of  $x_{N_\delta}^\delta$  to  $x^\dagger$  can be established. This stopping rule, however, is an a priori one since it depends on  $\nu$ , which is difficult to know in practice. Therefore a wrong guess of the smoothness on  $x_0 - x^\dagger$  will lead to a bad choice of  $N_\delta$ , and consequently to a bad approximation to the exact solution  $x^\dagger$  of (1). Thus, this rule is of no practical interest, and an a posteriori criterion should be considered to choose the stopping index of iteration.

An a posteriori stopping rule has been proposed in [3] for the iteratively regularized Gauss-Newton method, and the stopping index of iteration  $n_\delta$  is chosen according to the discrepancy principle

$$(7) \quad \|F(x_{n_\delta}^\delta) - y^\delta\| \leq c\delta < \|F(x_k^\delta) - y^\delta\|, \quad 0 \leq k < n_\delta,$$

with  $c > 1$  chosen sufficiently large. Under certain conditions, the approximation property of  $x_{n_\delta}^\delta$  has been studied, and it has been proved that

$$(8) \quad \|x_{n_\delta}^\delta - x^\dagger\| \leq O(\delta^{\frac{2\nu}{1+2\nu}})$$

if  $x_0 - x^\dagger$  satisfies (6) with  $0 < \nu \leq \frac{1}{2}$ . Although they are interesting and useful, the results in [3] have the following disadvantages:

- With the  $n_\delta$  chosen from (7), one cannot expect to obtain a better rate than  $O(\delta^{1/2})$  even if  $x_0 - x^\dagger$  satisfies (6) with some  $\nu > \frac{1}{2}$ .
- The rates (8) were obtained under some conditions on  $F$  like

$$(9) \quad \begin{aligned} F'(x) &= R(x, z)F'(z) + Q(x, z), \\ \|I - R(x, z)\| &\leq C_R, & x, z \in B_{2\rho}(x_0), \\ \|Q(x, z)\| &\leq C_Q \|F'(x^\dagger)(x - z)\|, \end{aligned}$$

with  $\rho$ ,  $C_R$  and  $C_Q$  sufficiently small. Unfortunately, for many important inverse problems arising in medical imaging and nondestructive testing, condition (9) seems to be difficult to verify or even to be false.

Considering these aspects, it is natural to ask whether it is possible to give an a posteriori stopping rule yielding higher rates of convergence even under weaker conditions than (9). In this paper we try to answer this question. By making a comparison with Tikhonov regularization in Section 2, we find some similarities between these two methods. This observation leads us to propose a new rule for choosing the stopping index of iteration. With the index  $k_\delta$  chosen by our rule, we state some interesting results on  $x_{k_\delta}^\delta$ , including the convergence and rates of convergence, under a mild assumption in Section 2. Some numerical examples are given in Section 3 to verify the theoretical results. The proofs of the main results are given in Section 5, which is based on an important inequality obtained in Section 4.

## 2. THE STOPPING RULE AND MAIN RESULTS

As explained in the introduction, an a posteriori rule for choosing the stopping index of iteration is necessary when one wants to apply the iteratively regularized Gauss-Newton method to practical problems. Perhaps the discrepancy principle (7), which is frequently used in iterative regularization methods, is a natural one. However, as claimed in [3], with the stopping index chosen by this rule, the best possible rate of convergence cannot exceed  $O(\delta^{1/2})$ . So it is of interest to give an a posteriori rule yielding higher rates of convergence. To this end, let us compare Tikhonov regularization and the iteratively regularized Gauss-Newton method. If  $F$  is a linear bounded operator and  $x_0 = 0$ , then  $\{x_k^\delta\}$  is defined successively by

$$x_{k+1}^\delta = x_k^\delta - (\alpha_k I + F^* F)^{-1} (F^* (F x_k^\delta - y^\delta) + \alpha_k x_k^\delta).$$

From this one can easily get

$$x_k^\delta = (\alpha_k I + F^* F)^{-1} F^* y^\delta,$$

which indicates that  $x_k^\delta$  is nothing but the Tikhonov regularized solution corresponding to the regularization parameter  $\alpha = \alpha_k$  with  $\alpha_k$  chosen properly [9]. When  $F$  is a nonlinear operator,  $x_k^\delta$  is no longer the Tikhonov regularized solution, but we can conceive that there must exist some similarities between these two methods. Therefore it is helpful to recall the existing parameter choice strategy for Tikhonov regularization of nonlinear ill-posed problems. As we know, by

generalizing the idea developed in [8], Scherzer, Engl and Kunisch [19] proposed a rule to choose the regularization parameter for Tikhonov regularization of nonlinear ill-posed problems in 1993, and used the root  $\alpha := \alpha(\delta)$  of the equation

$$(10) \quad \alpha \left( F(x_\alpha^\delta) - y^\delta, (\alpha I + F'(x_\alpha^\delta)F'(x_\alpha^\delta)^*)^{-1} (F(x_\alpha^\delta) - y^\delta) \right) = c\delta^2$$

as the regularization parameter, and studied the convergence property of  $x_{\alpha(\delta)}^\delta$ . Further study of this strategy was given in [13], and it was pointed out that (10) can be applied to many concrete problems. From the above observation, by adapting (10) we propose the following stopping rule for the iteratively regularized Gauss-Newton method.

**Rule 2.1.** Let  $c \geq 1$  be a given constant and  $x_0 \in D(F)$ . Then choose  $k_\delta$  to be the first integer such that

$$(11) \quad \alpha_{k_\delta} \left( F(x_{k_\delta}^\delta) - y^\delta, (\alpha_{k_\delta} I + F'(x_{k_\delta}^\delta)F'(x_{k_\delta}^\delta)^*)^{-1} (F(x_{k_\delta}^\delta) - y^\delta) \right) \leq c\delta^2.$$

With the above chosen  $k_\delta$ , we will use  $x_{k_\delta}^\delta$  to approximate the exact solution  $x^\dagger$  of (1). Before proceeding to argue the convergence behavior of  $x_{k_\delta}^\delta$ , we have to show the justification of Rule 2.1. To do this, we need the following restriction on  $F$ , which has been carefully interpreted in [19].

**Assumption 2.1.** There is a number  $p > 3\|x_0 - x^\dagger\|$  such that  $B_p(x^\dagger) := \{x \in X : \|x - x^\dagger\| \leq p\} \subset D(F)$ . Moreover, there exists a constant  $K_0$  such that for each pair  $x, z \in B_p(x^\dagger)$  and  $v \in X$  there is an element  $h(x, z, v) \in X$  such that

$$(F'(x) - F'(z))v = F'(z)h(x, z, v),$$

where

$$\|h(x, z, v)\| \leq K_0\|x - z\|\|v\|.$$

Now we can show that Rule 2.1 is well defined if  $c \geq \frac{25}{4}$  and  $12K_0\|x_0 - x^\dagger\| \leq 1$ . Obviously, all we have to do is to show that there is a finite integer  $k_\delta$  satisfying (11) if  $x_0 \neq x^\dagger$ . By denoting by  $\tilde{k}_\delta$  the integer such that

$$(12) \quad \alpha_{\tilde{k}_\delta} \leq \frac{(\sqrt{c} - 1)^2\delta^2}{4\|x_0 - x^\dagger\|^2} < \alpha_k, \quad 0 \leq k < \tilde{k}_\delta,$$

then we only need to prove that

$$(13) \quad a_{\tilde{k}_\delta} := \alpha_{\tilde{k}_\delta} \left( F(x_{\tilde{k}_\delta}^\delta) - y^\delta, (\alpha_{\tilde{k}_\delta} I + F'(x_{\tilde{k}_\delta}^\delta)F'(x_{\tilde{k}_\delta}^\delta)^*)^{-1} (F(x_{\tilde{k}_\delta}^\delta) - y^\delta) \right) \leq c\delta^2.$$

Let us first show that  $x_k^\delta$  is well defined for all integers  $0 \leq k \leq \tilde{k}_\delta$  by induction. Suppose  $x_k^\delta \in B_p(x^\dagger)$  for some  $0 \leq k < \tilde{k}_\delta$ ; then the definition of  $x_{k+1}^\delta$  gives

$$(14) \quad \begin{aligned} x_{k+1}^\delta - x^\dagger &= (\alpha_k I + F'(x_k^\delta)^*F'(x_k^\delta))^{-1} \left\{ \alpha_k(x_0 - x^\dagger) + F'(x_k^\delta)^*(y^\delta - y) \right. \\ &\quad \left. - F'(x_k^\delta)^*(F(x_k^\delta) - y - F'(x_k^\delta)(x_k^\delta - x^\dagger)) \right\}. \end{aligned}$$

Since Assumption 2.1 implies

$$(15) \quad F(x_k^\delta) - y - F'(x_k^\delta)(x_k^\delta - x^\dagger) = F'(x_k^\delta) \int_0^1 h_t^\delta dt$$

with  $h_t^\delta := h(x^\dagger + t(x_k^\delta - x^\dagger), x_k^\delta, x_k^\delta - x^\dagger)$  and  $\left\| \int_0^1 h_t^\delta dt \right\| \leq \frac{K_0}{2} \|x_k^\delta - x^\dagger\|^2$ , we have from (14) that

$$\|x_{k+1}^\delta - x^\dagger\| \leq \|\alpha_k(\alpha_k I + F'(x_k^\delta)^* F'(x_k^\delta))^{-1}(x_0 - x^\dagger)\| + \frac{\delta}{2\sqrt{\alpha_k}} + \frac{K_0}{2} \|x_k^\delta - x^\dagger\|^2.$$

From the definition of  $\tilde{k}_\delta$ , since  $c \geq \frac{25}{4}$ , we have for  $0 \leq k < \tilde{k}_\delta$

$$(16) \quad \|x_{k+1}^\delta - x^\dagger\| \leq \frac{5}{3} \|x_0 - x^\dagger\| + \frac{K_0}{2} \|x_k^\delta - x^\dagger\|^2.$$

By induction now we can prove if  $x_0$  is so close to  $x^\dagger$  that  $K_0 \|x_0 - x^\dagger\| \leq \eta$  with some  $\eta \leq \frac{8}{27}$ , then for all integers  $0 \leq k \leq \tilde{k}_\delta$

$$(17) \quad \|x_k^\delta - x^\dagger\| \leq \frac{10}{3 + \sqrt{9 - 30\eta}} \|x_0 - x^\dagger\| \leq 3 \|x_0 - x^\dagger\|.$$

Therefore  $x_k^\delta$  is well-defined for all  $0 \leq k \leq \tilde{k}_\delta$ . To make the following discussion laconic, we introduce the abbreviations

$$\mathcal{A}_k^\delta := F'(x_k^\delta)^* F'(x_k^\delta) \quad \text{and} \quad \mathcal{B}_k^\delta := F'(x_k^\delta) F'(x_k^\delta)^*.$$

Now from (15) and (17), and noting that  $K_0 \|x_0 - x^\dagger\| \leq \eta := \frac{1}{12}$ , it follows that

$$\begin{aligned} \sqrt{\alpha_{\tilde{k}_\delta}} &\leq \delta + \sqrt{\alpha_{\tilde{k}_\delta}} \|(\alpha_{\tilde{k}_\delta} I + \mathcal{B}_{\tilde{k}_\delta}^\delta)^{-\frac{1}{2}} (F(x_{\tilde{k}_\delta}^\delta) - y)\| \\ &\leq \delta + \sqrt{\alpha_{\tilde{k}_\delta}} \left( 1 + \frac{K_0}{2} \|x_{\tilde{k}_\delta}^\delta - x^\dagger\| \right) \|x_{\tilde{k}_\delta}^\delta - x^\dagger\| \\ &\leq \delta + \frac{10(1 + 5\eta/(3 + \sqrt{9 - 30\eta}))}{3 + \sqrt{9 - 30\eta}} \frac{\sqrt{c} - 1}{2} \delta \\ &\leq \sqrt{c}\delta. \end{aligned}$$

Therefore Rule 2.1 is well defined, and for the integer  $k_\delta$  determined by Rule 2.1 we always have  $k_\delta \leq \tilde{k}_\delta$ .

We are now in a position to state the main results. In order to formulate some conditions in a concise manner, throughout this paper we assume that the nonlinear operator  $F$  is properly scaled, i.e.

$$(18) \quad \|F'(x)\| \leq \sqrt{\frac{\alpha_0}{3}} \quad \forall x \in B_p(x^\dagger).$$

This scaling condition can always be fulfilled by multiplying both sides of (1) by a sufficiently small constant, which then appears as a relaxation parameter in the iteratively regularized Gauss-Newton method.

**Theorem 2.1.** *Let Assumption 2.1, (5) and (18) hold,  $12rK_0 \|x_0 - x^\dagger\| \leq 1$ ,  $c \geq \frac{25}{4}$ , and let  $k_\delta$  be the integer chosen from Rule 2.1. Then there is a constant  $C$ , independent of  $\delta$ , such that for all  $\delta > 0$*

$$(19) \quad \|x_{k_\delta}^\delta - x^\dagger\| \leq C \inf \left\{ \|x_k - x^\dagger\| + \frac{\delta}{\sqrt{\alpha_k}} : k = 0, 1, \dots \right\},$$

where  $\{x_k\}$  is the sequence defined by the iteratively regularized Gauss-Newton method (4) corresponding to the noise-free case.

The estimate (19) is quite useful; from it we can get a lot of information on  $x_{k_\delta}^\delta$ . In particular, we can use it to derive the convergence and rates of convergence for the iteratively regularized Gauss-Newton method.

**Corollary 2.1.** *Suppose the conditions in Theorem 2.1 are satisfied, and let  $k_\delta$  be the integer chosen by Rule 2.1. If  $x_0$  is chosen such that  $x_0 - x^\dagger \in \mathcal{N}(F'(x^\dagger))^\perp$ , then*

$$(20) \quad \lim_{\delta \rightarrow 0} x_{k_\delta}^\delta = x^\dagger.$$

Moreover, if  $x_0 - x^\dagger$  satisfies (6) with some  $0 < \nu \leq 1$ , then

$$(21) \quad \|x_{k_\delta}^\delta - x^\dagger\| \leq C_\nu \|\omega\|^{\frac{1}{1+2\nu}} \delta^{\frac{2\nu}{1+2\nu}}$$

with a constant  $C_\nu$  depending on  $\nu$  only.

Corollary 2.1 suggests that the iteratively regularized Gauss-Newton method together with Rule 2.1 defines a regularization method of optimal order for each  $0 < \nu \leq 1$  (see [22, 15]). The upper bound provided by (21) is of uniform nature without special regard for  $y$ . In a typical instance, however, the convergence of  $x_{k_\delta}^\delta$  to  $x^\dagger$  is faster than (21) claims, even under the slight weaker conditions

$$(22) \quad \int_0^\mu d\|E_\lambda(x_0 - x^\dagger)\|^2 = O(\mu^{2\nu})$$

and

$$(23) \quad \int_0^\mu d\|E_\lambda(x_0 - x^\dagger)\|^2 = o(\mu^{2\nu}),$$

where  $0 < \nu < 1$  and  $\{E_\lambda\}$  denotes the spectral family generated by the self-adjoint operator  $F'(x^\dagger)^*F'(x^\dagger)$ . These conditions were used first by Neubauer [16] to prove the converse and saturation results for Tikhonov regularization of linear ill-posed problems. The comparison of (22) and (23) with (6) can be seen from [16, Proposition 2.3].

**Corollary 2.2.** *Assume the conditions in Theorem 2.1 are satisfied, and let  $k_\delta$  be the integer defined by Rule 2.1. Then*

$$(24) \quad \|x_{k_\delta}^\delta - x^\dagger\| \leq \begin{cases} O(\delta^{\frac{2\nu}{1+2\nu}}) & \text{if } x_0 - x^\dagger \text{ satisfies (22),} \\ o(\delta^{\frac{2\nu}{1+2\nu}}) & \text{if } x_0 - x^\dagger \text{ satisfies (23).} \end{cases}$$

All the above results will be proved in Section 5. Some necessary preparation will be given in Section 4; in particular, an important inequality, which is the key to proving Theorem 2.1, will be presented. Please note results similar to (19) for some regularization methods for linear ill-posed problems have been obtained in several references [6, 17].

Before concluding this section, let us make a comparison between Assumption 2.1 and (9). At first glance it seems that Assumption 2.1 is very similar to (9). But in fact this is not the case—Assumption 2.1 is always easier to verify than (9). For example, we consider the problem of estimating the coefficient  $a$  in the boundary value problem

$$(25) \quad \begin{cases} -\Delta u + au = f & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega, \end{cases}$$

from the additional measurement of the normal derivative of  $u$  on  $\partial\Omega$ , where  $\Omega$  is a bounded domain in  $R^3$  or  $R^2$  with smooth boundary,  $f \in L^2(\Omega)$  and  $g \in H^{3/2}(\partial\Omega)$ . Let  $\mathcal{T}$  be the trace operator  $\mathcal{T} : H^2(\Omega) \mapsto L^2(\partial\Omega)$ ,  $\mathcal{T}\phi = \frac{\partial\phi}{\partial n}|_{\partial\Omega}$ , and let  $G$  be the

parameter-to-solution mapping  $G : D(G) \subset L^2(\Omega) \mapsto H^2(\Omega)$ ,  $G(a) = u(a)$ , where  $u(a)$  is the unique solution of (25) and

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

with a suitable small constant  $\gamma > 0$ . Then we can define the nonlinear operator  $F$  as  $F = \mathcal{T} \circ G$ , which is well-defined on  $D(F) := D(G)$  (see [4]), and the Fréchet derivative of  $F$  is given by

$$F'(a)h = -\mathcal{T}A(a)^{-1}(hG(a)),$$

where  $A(a) : H^2 \cap H_0^1 \mapsto L^2$  is defined by  $A(a)u = -\Delta u + au$ . It has been shown (see [14]) that if  $|u(a^\dagger)(t)| \geq \kappa > 0$  for all  $t \in \partial\Omega$ , then Assumption 2.1 is true. However, it is difficult to verify (9) for this example. Indeed, the validity of (9) requires  $\mathcal{T}$  to commute with a family of linear operators, which is impossible in general.

### 3. NUMERICAL EXAMPLES

In this section we present some numerical results to test our assertion for Rule 2.1. For simplicity we just do the numerical experiments for the parameter estimation of ordinary differential equations. In all examples we always choose the stopping index  $k_\delta$  by Rule 2.1 with  $c = 1$ . Note that  $c = 1$  does not satisfy the lower bound  $\frac{25}{4}$  stated in Theorem 2.1. However, this bound mainly comes from the proof of the justification of Rule 2.1 and the proof of Lemma 5.2 (see Section 5). If Rule 2.1 is well-defined for smaller  $c$  and if we check the proof of Lemma 5.2 carefully, we can drop the requirement on  $c$  provided  $\|x_0 - x^\dagger\|$  sufficiently small. In numerical computation, one should use smaller  $c$  if possible, since the absolute error increases with  $c$ . In the following we also make a comparison between Rule 2.1 and the discrepancy principle (7); for the latter rule, we also choose  $c = 1$ .

We consider the identification of the coefficient  $a$  in the two-point boundary value problem

$$(26) \quad \begin{cases} -u'' + au = f, & t \in (0, 1), \\ u(0) = g_0, & u(1) = g_1, \end{cases}$$

from the measurement data  $u^\delta$  of the state variable  $u$ , where  $g_0, g_1$  and  $f \in L^2[0, 1]$  are given. Now the nonlinear operator  $F : D(F) \subset L^2[0, 1] \mapsto L^2[0, 1]$  is defined as the parameter-to-solution mapping  $F(a) = u(a)$  with  $u(a)$  being the unique solution of (26).  $F$  is well-defined (see [4]) on

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

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

$$\begin{aligned} F'(a)h &= -A(a)^{-1}(hu(a)), \\ F'(a)^*w &= -u(a)A(a)^{-1}w, \end{aligned}$$

where  $A(a) : H^2 \cap H_0^1 \mapsto L^2$  is defined by  $A(a)u = -u'' + au$ . It has been shown (see [19]) that Assumption 2.1 and (9) are valid if  $|u(a^\dagger)(t)| \geq \kappa > 0$  for all  $t \in [0, 1]$ .

**Example 3.1.** Here we estimate  $a$  in (26) by assuming  $f = 1 + t^2$  and  $g_0 = g_1 = 1$ . If  $u(a^\dagger) = 1$ , then the true solution is  $a^\dagger = 1 + t^2$ . In our computation, instead of  $u(a^\dagger)$  we use the special perturbation  $u^\delta = 1 + \delta\sqrt{2}\sin(10\pi t)$ . Clearly  $\|u^\delta - u(a^\dagger)\|_{L^2} = \delta$ . In order to apply the iteratively regularized Gauss-Newton

TABLE 1.a.  $\alpha_k = 0.1 \times 0.5^{k-1}$ 

| $\delta$    | Rule 2.1   |   |                    | Discrepancy Principle (7) |   |                    |
|-------------|------------|---|--------------------|---------------------------|---|--------------------|
|             | $k_\delta$ | $e_1 = \ a_{k_\delta}^\delta - a^\dagger\ $ | $e_1/\delta^{2/3}$ | $n_\delta$                | $e_2 = \ a_{n_\delta}^\delta - a^\dagger\ $ | $e_2/\delta^{1/2}$ |
| $0.10e + 0$ | 4          | $0.26e + 0$                                 | $0.12e + 1$        | 5                         | $0.19e + 0$                                 | $0.60e + 0$        |
| $0.10e - 1$ | 7          | $0.71e - 1$                                 | $0.15e + 1$        | 9                         | $0.29e - 1$                                 | $0.29e + 0$        |
| $0.10e - 2$ | 10         | $0.20e - 1$                                 | $0.20e + 1$        | 12                        | $0.15e - 1$                                 | $0.47e + 0$        |
| $0.10e - 3$ | 12         | $0.37e - 2$                                 | $0.17e + 1$        | 15                        | $0.11e - 1$                                 | $0.11e + 1$        |
| $0.10e - 4$ | 14         | $0.15e - 2$                                 | $0.32e + 1$        | 18                        | $0.56e - 2$                                 | $0.18e + 1$        |

TABLE 1.b.  $\alpha_k = 0.1 \times 0.25^{k-1}$ 

| $\delta$    | Rule 2.1   |   |                    | Discrepancy Principle (7) |   |                    |
|-------------|------------|---|--------------------|---------------------------|---|--------------------|
|             | $k_\delta$ | $e_1 = \ a_{k_\delta}^\delta - a^\dagger\ $ | $e_1/\delta^{2/3}$ | $n_\delta$                | $e_2 = \ a_{n_\delta}^\delta - a^\dagger\ $ | $e_2/\delta^{1/2}$ |
| $0.10e + 0$ | 2          | $0.33e + 0$                                 | $0.15e + 1$        | 3                         | $0.19e + 0$                                 | $0.60e + 0$        |
| $0.10e - 1$ | 4          | $0.72e - 1$                                 | $0.16e + 1$        | 5                         | $0.62e - 1$                                 | $0.62e + 0$        |
| $0.10e - 2$ | 5          | $0.21e - 1$                                 | $0.21e + 1$        | 7                         | $0.30e - 1$                                 | $0.95e + 0$        |
| $0.10e - 3$ | 6          | $0.61e - 2$                                 | $0.28e + 1$        | 8                         | $0.11e - 1$                                 | $0.11e + 1$        |
| $0.10e - 4$ | 7          | $0.21e - 2$                                 | $0.45e + 1$        | 10                        | $0.78e - 2$                                 | $0.25e + 1$        |

method, we choose the first guess as  $a_0 = 1 + t^2 - 2t(1-t)(1+t-t^2)$ . It is easy to know that  $a_0 - a^\dagger \in R(F'(a^\dagger)^*F'(a^\dagger))$  (see [7]), and thus the rate of convergence we can expect should be  $O(\delta^{2/3})$ .

In Tables 1.a and 1.b we report the numerical results obtained by using Rule 2.1 and the discrepancy principle (7) with different choices of the sequence  $\{\alpha_k\}$ . During the computation, the differential equations we met were solved approximately by the finite element method on the subspace of piecewise linear splines on a uniform grid with subinterval length  $\frac{1}{16}$ . Considering the discretization error, Tables 1.a and 1.b indicate that  $a_{k_\delta}$  converges to  $a^\dagger$  with a rate  $O(\delta^{2/3})$  if  $k_\delta$  is chosen by Rule 2.1, and only a convergence rate  $O(\delta^{1/2})$  can be seen for the discrepancy principle (7). This numerically illustrates the fact that the discrepancy principle (7) never yields a better convergence rate than  $O(\delta^{1/2})$ . At first glimpse, it seems that Rule 2.1 is more time-consuming than the discrepancy principle (7), since an additional operator  $\alpha_k I + \mathcal{B}_k^\delta$  has to be inverted in each iteration step. However, Tables 1.a and 1.b tell us that more iterations, which of course take time, have to be done for the discrepancy principle (7) to get the final results. In fact, the computational time for the discrepancy principle (7) is slightly longer than that for Rule 2.1 for small  $\delta$  in this example. Furthermore, we can see from Tables 1.a and 1.b that the results obtained by Rule 2.1 are better than those obtained by the discrepancy principle (7) if  $\delta > 0$  is quite small. Due to the observation given above, we can recommend Rule 2.1 in applications.

The results in Tables 1.a and 1.b also illustrate the influence of the choice of the sequence  $\{\alpha_k\}$ . The sequence  $\{\alpha_k\}$  used in Table 1.b decreases faster than that used in Table 1.a, so fewer iteration need to be done to get the final results, but the risk of worse convergence perhaps arises.

TABLE 2.  $\alpha_k = 0.1 \times 0.25^{k-1}$

| $\delta$    | Rule 2.1   |   | Discrepancy Principle (7) |   |
|-------------|------------|---|---------------------------|---|
|             | $k_\delta$ | $e_1 = \ a_{k_\delta}^\delta - a^\dagger\ $ | $n_\delta$                | $e_2 = \ a_{n_\delta}^\delta - a^\dagger\ $ |
| $0.10e + 0$ | 3          | $0.29e + 0$                                 | 3                         | $0.29e + 0$                                 |
| $0.10e - 1$ | 4          | $0.22e + 0$                                 | 6                         | $0.18e + 0$                                 |
| $0.10e - 2$ | 7          | $0.14e + 0$                                 | 8                         | $0.16e + 0$                                 |
| $0.10e - 3$ | 9          | $0.11e + 0$                                 | 9                         | $0.11e + 0$                                 |
| $0.10e - 4$ | 10         | $0.10e + 0$                                 | 11                        | $0.10e + 0$                                 |

**Example 3.2.** Here we continue the estimation of  $a$  in the problem (26) as described in Example 3.1, but use the first guess  $a_0 = 0.5 + t^2$ . Now

$$a_0 - a^\dagger \notin R(F'(a^\dagger)^*),$$

and in fact  $a_0 - a^\dagger$  has no sourcewise representation (6) with a good  $\nu > 0$ , so we cannot expect a good convergence rate either for Rule 2.1 or for the discrepancy principle (7), according to Corollary 2.1 and [3, Theorem 3.1]. However, we still have the convergence, which can be seen from Table 2, and the two stopping rules yield almost the same rates of convergence; here we choose  $\alpha_k = 0.1 \times 0.25^{k-1}$ . We also consider the choice  $\alpha_k = 0.1 \times 0.5^{k-1}$  for this example; the numerical results are essentially the same.

**Example 3.3.** Here we again estimate the parameter  $a$  in (26), but with  $g_0 = 0$ ,  $g_1 = 1$  and  $f = t$ . If  $u(a^\dagger) = t$ , the true solution is  $a^\dagger = 1$ . In our calculation we use the special perturbation  $u^\delta = t + \delta\sqrt{2}\sin(10\pi t)$ . As the first guess we choose  $a_0 = 1 + 0.4(7t^2 - 10t^4 + 3t^6)$ . It can be argued that

$$a_0 - a^\dagger \in R(F'(a^\dagger)^* F'(a^\dagger)).$$

In Table 3 we summarize the numerical results obtained by using Rule 2.1, and the discrepancy principle (7) with  $\alpha_k = 0.1 \times 0.25^{k-1}$ . The convergence rate  $O(\delta^{2/3})$  can be seen for Rule 2.1, and the rate  $O(\delta^{1/2})$  holds for the discrepancy principle (7) again. Note that we could not verify Assumption 2.1 for this example. Thus the results indicate that Rule 2.1 has a wider applicability than indicated by the conditions of Theorem 2.1. Recently we have obtained some results for Rule 2.1 under weaker conditions than Assumption 2.1, and more research is in progress now. Because of the different framework, we will report them in another paper.

TABLE 3.  $\alpha_k = 0.1 \times 0.25^{k-1}$

| $\delta$    | Rule 2.1   |   |                    | Discrepancy Principle (7) |   |                    |
|-------------|------------|---|--------------------|---------------------------|---|--------------------|
|             | $k_\delta$ | $e_1 = \ a_{k_\delta}^\delta - a^\dagger\ $ | $e_1/\delta^{2/3}$ | $n_\delta$                | $e_2 = \ a_{n_\delta}^\delta - a^\dagger\ $ | $e_2/\delta^{1/2}$ |
| $0.10e - 1$ | 5          | $0.52e - 1$                                 | $0.11e + 1$        | 6                         | $0.46e - 1$                                 | $0.46e + 0$        |
| $0.10e - 2$ | 6          | $0.15e - 1$                                 | $0.15e + 1$        | 7                         | $0.18e - 1$                                 | $0.37e + 0$        |
| $0.10e - 3$ | 7          | $0.41e - 2$                                 | $0.19e + 1$        | 9                         | $0.97e - 2$                                 | $0.97e + 0$        |
| $0.10e - 4$ | 8          | $0.12e - 2$                                 | $0.26e + 1$        | 10                        | $0.63e - 2$                                 | $0.20e + 1$        |
| $0.10e - 5$ | 9          | $0.34e - 3$                                 | $0.34e + 1$        | 11                        | $0.13e - 2$                                 | $0.13e + 1$        |

4. SOME RESULTS ASSOCIATED WITH THE NOISE FREE CASE

In this section we will give some investigation on the sequence  $\{x_k\}$  defined by (4) with  $y^\delta$  replaced by  $y$ . By assuming  $x_k \in B_p(x^\dagger)$  for some integer  $k$ , the definition of  $x_{k+1}$  gives

$$(27) \quad \begin{aligned} x_{k+1} - x^\dagger &= (\alpha_k I + F'(x_k)^* F'(x_k))^{-1} \left\{ \alpha_k (x_0 - x^\dagger) \right. \\ &\quad \left. - F'(x_k)^* (F(x_k) - y - F'(x_k)(x_k - x^\dagger)) \right\}. \end{aligned}$$

Since Assumption 2.1 implies

$$(28) \quad F(x_k) - y - F'(x_k)(x_k - x^\dagger) = F'(x_k) \int_0^1 h_t dt$$

with  $h_t = h(x^\dagger + t(x_k - x^\dagger), x_k, x_k - x^\dagger)$  and  $\left\| \int_0^1 h_t dt \right\| \leq \frac{K_0}{2} \|x_k - x^\dagger\|^2$ , we have from (27) that

$$(29) \quad \tilde{\beta}_k - \frac{K_0}{2} \|x_k - x^\dagger\|^2 \leq \|x_{k+1} - x^\dagger\| \leq \tilde{\beta}_k + \frac{K_0}{2} \|x_k - x^\dagger\|^2$$

with  $\tilde{\beta}_k := \|\alpha_k(\alpha_k I + F'(x_k)^* F'(x_k))^{-1}(x_0 - x^\dagger)\|$ . In particular, (29) implies

$$\|x_{k+1} - x^\dagger\| \leq \|x_0 - x^\dagger\| + \frac{K_0}{2} \|x_k - x^\dagger\|^2.$$

From this by induction we can show that if  $K_0 \|x_0 - x^\dagger\| \leq \eta$  with a constant  $\eta \leq \frac{1}{2}$ , then

$$(30) \quad \|x_k - x^\dagger\| \leq \frac{2}{1 + \sqrt{1 - 2\eta}} \|x_0 - x^\dagger\| \leq 2 \|x_0 - x^\dagger\|$$

for all integers  $k \geq 0$ . Therefore the sequence  $\{x_k\}$  is well defined.

The next lemma, although elementary, is very useful in the following discussions.

**Lemma 4.1.** *Let  $\{p_k\}_{k=0}^\infty$  be a sequence of positive numbers satisfying  $\frac{p_k}{p_{k+1}} \leq p$  with a constant  $p \geq 1$ . Suppose the sequence  $\{\eta_k\}_{k=0}^\infty$  has the property*

$$(31) \quad p_k - \tau \eta_k \leq \eta_{k+1} \leq p_k + \tau \eta_k, \quad k = 0, 1, \dots$$

*If  $\tau p < 1$  and  $\eta_0 \leq \frac{p}{1 - \tau p} p_0$ , then for all  $k$*

$$(32) \quad \eta_k \leq \frac{p}{1 - \tau p} p_k.$$

*If in addition,  $\{p_k\}_{k=0}^\infty$  is monotonically decreasing,  $\eta_0 \geq p_0$  and  $2\tau p < 1$ , then for all  $k$*

$$(33) \quad \eta_k \geq \frac{1 - 2\tau p}{1 - \tau p} p_k.$$

*Proof.* Assertion (32) can be proved by induction. In fact, it is trivial for  $k = 0$ . If it is true for  $k = j$ , then for  $k = j + 1$  we have

$$\eta_{j+1} \leq p_j + \tau \eta_j \leq \frac{1}{1 - \tau p} p_j = \frac{1}{1 - \tau p} \frac{p_j}{p_{j+1}} p_{j+1} \leq \frac{p}{1 - \tau p} p_{j+1}.$$

And hence (32) follows. Assertion (33) is an immediate consequence of (31) and (32). □

To continue our study, let us state a consequence of Assumption 2.1.

**Lemma 4.2.** *Let Assumption 2.1 hold. For each pair  $u, v \in B_p(x^\dagger)$  we denote  $\mathcal{A} = F'(u)^*F'(u)$  and  $\mathcal{B} = F'(v)^*F'(v)$ . Then, for all  $\alpha > 0$ ,*

$$(34) \quad \|\alpha((\alpha I + \mathcal{A})^{-1} - (\alpha I + \mathcal{B})^{-1})\| \leq 2K_0\|u - v\|.$$

*Proof.* Let  $a, b \in X$  be arbitrary. Then by Assumption 2.1,

$$\begin{aligned} & |(\alpha((\alpha I + \mathcal{A})^{-1} - (\alpha I + \mathcal{B})^{-1})a, b)| \\ & \leq |\alpha((\alpha I + \mathcal{A})^{-1}F'(u)^*(F'(u) - F'(v))(\alpha I + \mathcal{B})^{-1}a, b)| \\ & \quad + |\alpha(F'(v)(\alpha I + \mathcal{B})^{-1}a, (F'(u) - F'(v))(\alpha I + \mathcal{A})^{-1}b)| \\ & = |\alpha((\alpha I + \mathcal{A})^{-1}\mathcal{A}h(v, u, (\alpha I + \mathcal{B})^{-1}a), b)| \\ & \quad + |\alpha(\mathcal{B}(\alpha I + \mathcal{B})^{-1}a, h(u, v, (\alpha I + \mathcal{A})^{-1}b))| \\ & \leq 2K_0\|u - v\|\|a\|\|b\|, \end{aligned}$$

which gives (34) immediately. □

Now we introduce some notation by defining

$$\mathcal{C} := F'(x^\dagger)^*F'(x^\dagger), \quad \mathcal{D} := F'(x^\dagger)F'(x^\dagger)^* \quad \text{and} \quad \mathcal{A}_k := F'(x_k)^*F'(x_k) \quad \text{for all } k.$$

This helps make our statements more compact. Obviously, these operators are all self-adjoint and nonnegative definite.

**Lemma 4.3.** *Let Assumption 2.1, (5) and (18) hold and  $12rK_0\|x_0 - x^\dagger\| \leq 1$ . Then, for all  $k$ ,*

$$(35) \quad \frac{2}{3}\beta_k \leq \|x_k - x^\dagger\| \leq \frac{4}{3}r\beta_k,$$

$$(36) \quad \frac{1}{2r}\|x_k - x^\dagger\| \leq \|x_{k+1} - x^\dagger\| \leq 2\|x_k - x^\dagger\|,$$

where  $\beta_k$  is defined by  $\beta_k := \|\alpha_k(\alpha_k I + \mathcal{C})^{-1}(x_0 - x^\dagger)\|$ .

*Proof.* Since  $x_0 = x^\dagger$  implies  $x_k = x^\dagger$ , assertion (35) is trivial. Therefore in what follows we assume  $x_0 \neq x^\dagger$ .

With an application of (34) we have  $|\tilde{\beta}_k - \beta_k| \leq 2K_0\|x_0 - x^\dagger\|\|x_k - x^\dagger\|$ . Hence, noting that  $12rK_0\|x_0 - x^\dagger\| \leq 1$ , from (29) and (30) it follows that

$$(37) \quad \beta_k - \frac{1}{4r}\|x_k - x^\dagger\| \leq \|x_{k+1} - x^\dagger\| \leq \beta_k + \frac{1}{4r}\|x_k - x^\dagger\|.$$

Let  $\{E_\lambda\}$  be the spectral family generated by  $\mathcal{C}$ . Then

$$\begin{aligned} (38) \quad \beta_k^2 &= \int_0^\infty \frac{\alpha_k^2}{(\lambda + \alpha_k)^2} d\|E_\lambda(x_0 - x^\dagger)\|^2 \\ &\leq \left(\frac{\alpha_k}{\alpha_{k+1}}\right)^2 \int_0^\infty \frac{\alpha_{k+1}^2}{(\lambda + \alpha_{k+1})^2} d\|E_\lambda(x_0 - x^\dagger)\|^2 \\ &\leq r^2\beta_{k+1}^2. \end{aligned}$$

Since (18) implies  $\|x_0 - x^\dagger\| \geq \beta_0 \geq \frac{3}{4}\|x_0 - x^\dagger\|$ , from (37), (38), the monotonicity of  $\{\beta_k\}$  and Lemma 4.1 we can obtain (35). Assertion (36) is a direct consequence of (35) and (37). □

Now let us digress a moment and give some converse and saturation results for  $\{x_k\}$  by using Lemma 4.3. As we know, it has been proved in [3] that

$$(39) \quad \|x_k - x^\dagger\| \leq \begin{cases} o(\alpha_k^\nu) & \text{if } 0 < \nu < 1, \\ O(\alpha_k) & \text{if } \nu = 1, \end{cases}$$

if  $x_0 - x^\dagger$  satisfies (6). Now we wonder whether (6) is necessary to derive (39) and whether  $O(\alpha_k)$  is the optimal rate. Neubauer [16] has pointed out that (6) is not necessary for the expected rates in general for Tikhonov regularization of linear ill-posed problems, and instead of (6), he has used the characterizations (22) and (23) of the true solution. In the following we use the recent results in [16] to show that (22) and (23) are necessary to derive the corresponding rates in (39), i.e. we have

**Proposition 4.1.** *Under the assumptions in Lemma 4.3, some converse results for  $\{x_k\}$  hold, i.e.,*

$$(40) \quad \|x_k - x^\dagger\| = O(\alpha_k) \iff x_0 - x^\dagger \text{ satisfies (6) with } \nu = 1$$

and, for  $0 < \nu < 1$ ,

$$(41) \quad \|x_k - x^\dagger\| = O(\alpha_k^\nu) \iff x_0 - x^\dagger \text{ satisfies (22),}$$

$$(42) \quad \|x_k - x^\dagger\| = o(\alpha_k^\nu) \iff x_0 - x^\dagger \text{ satisfies (23).}$$

Moreover, the saturation result holds:

$$(43) \quad \|x_k - x^\dagger\| = o(\alpha_k) \implies x_0 = x^\dagger.$$

*Proof.* Let us prove (41) first. Obviously (35) has the immediate consequence

$$(44) \quad \|x_k - x^\dagger\| = O(\alpha_k^\nu) \iff \beta_k = O(\alpha_k^\nu).$$

Now suppose  $\beta_k = O(\alpha_k^\nu)$  with some  $0 < \nu < 1$ . Since for any  $0 < \alpha \leq \alpha_0$  there exists an integer  $k$  such that  $\alpha_{k+1} < \alpha \leq \alpha_k$ , we have  $\alpha \leq \alpha_k \leq r\alpha$  and

$$\begin{aligned} \|\alpha(\alpha I + \mathcal{C})^{-1}(x_0 - x^\dagger)\|^2 &= \int_0^\infty \frac{\alpha^2}{(\alpha + \lambda)^2} d\|E_\lambda(x_0 - x^\dagger)\|^2 \\ &\leq \int_0^\infty \frac{\alpha_k^2}{(\alpha_{k+1} + \lambda)^2} d\|E_\lambda(x_0 - x^\dagger)\|^2 \\ &= \int_0^\infty \frac{\alpha_k^2}{(\alpha_k + \lambda)^2} \left(\frac{\alpha_k + \lambda}{\alpha_{k+1} + \lambda}\right)^2 d\|E_\lambda(x_0 - x^\dagger)\|^2 \\ &\leq r^2 \beta_k^2 = O(\alpha_k^{2\nu}) = O(\alpha^{2\nu}). \end{aligned}$$

Therefore we have in fact shown that

$$(45) \quad \beta_k = O(\alpha_k^\nu) \iff \|\alpha(\alpha I + \mathcal{C})^{-1}(x_0 - x^\dagger)\| = O(\alpha^\nu),$$

since the other direction is obvious. The combination of (44) and (45) gives

$$\|x_k - x^\dagger\| = O(\alpha_k^\nu) \iff \|\alpha(\alpha I + \mathcal{C})^{-1}(x_0 - x^\dagger)\| = O(\alpha^\nu).$$

Thus [16, Theorem 2.1] can be used to obtain (41). Assertions (40) and (42) can be proved in the same way.

Using the same argument in the above, we also have

$$\|x_k - x^\dagger\| = o(\alpha_k) \implies \|\alpha(\alpha I + \mathcal{C})^{-1}(x_0 - x^\dagger)\| = o(\alpha).$$

Therefore by using [9, Theorem 3.2.1] we can obtain (43). □

Now we can give the following inequality, which plays a significant role in the proof of Theorem 2.1.

**Lemma 4.4.** *Let Assumption 2.1, (5) and (18) hold and let  $12rK_0\|x_0 - x^\dagger\| \leq 1$ . Then, for all integers  $k \geq l \geq 0$ ,*

$$(46) \quad \|x_l - x^\dagger\| \leq C_0 \left\{ \|x_k - x^\dagger\| + \frac{\|\alpha_l^{\frac{1}{2}}(\alpha_l I + \mathcal{D})^{-\frac{1}{2}}(F(x_l) - y)\|}{\sqrt{\alpha_k}} \right\}$$

with a generic constant  $C_0$  independent of  $k$  and  $l$ .

*Proof.* We first consider the case  $l > 0$ . By setting  $k$  in (27) to be  $k - 1$  and  $l - 1$ , respectively, and then subtracting them, it follows that  $x_k - x_l = Q_1 + Q_2 + Q_3$ , where

$$\begin{aligned} Q_1 &:= \{ \alpha_{k-1}(\alpha_{k-1}I + \mathcal{A}_{k-1})^{-1} - \alpha_{l-1}(\alpha_{l-1}I + \mathcal{A}_{l-1})^{-1} \} (x_0 - x^\dagger), \\ Q_2 &:= (\alpha_{l-1}I + \mathcal{A}_{l-1})^{-1} F'(x_{l-1})^* (F(x_{l-1}) - y - F'(x_{l-1})(x_{l-1} - x^\dagger)), \\ Q_3 &:= (\alpha_{k-1}I + \mathcal{A}_{k-1})^{-1} F'(x_{k-1})^* (y - F(x_{k-1}) - F'(x_{k-1})(x^\dagger - x_{k-1})). \end{aligned}$$

By using Assumption 2.1 and (30) and noting that  $K_0\|x_0 - x^\dagger\| \leq \frac{1}{12r}$ , we can obtain, with  $\tau := \sqrt{6r}/(\sqrt{6r} + \sqrt{6r - 1})$ ,

$$(47) \quad \|Q_2\| \leq \frac{1}{2}K_0\|x_{l-1} - x^\dagger\|^2 \leq \tau K_0\|x_0 - x^\dagger\|\|x_{l-1} - x^\dagger\|,$$

$$(48) \quad \|Q_3\| \leq \frac{1}{2}K_0\|x_{k-1} - x^\dagger\|^2 \leq \tau K_0\|x_0 - x^\dagger\|\|x_{k-1} - x^\dagger\|.$$

And from (34) we also have

$$(49) \quad \begin{aligned} \|Q_1\| &\leq \|J\| + \|\alpha_{l-1}((\alpha_{l-1}I + \mathcal{A}_{l-1})^{-1} - (\alpha_{l-1}I + \mathcal{C})^{-1})(x_0 - x^\dagger)\| \\ &\quad + \|\alpha_{k-1}((\alpha_{k-1}I + \mathcal{A}_{k-1})^{-1} - (\alpha_{k-1}I + \mathcal{C})^{-1})(x_0 - x^\dagger)\| \\ &\leq \|J\| + 2K_0\|x_0 - x^\dagger\|(\|x_{k-1} - x^\dagger\| + \|x_{l-1} - x^\dagger\|), \end{aligned}$$

where  $J := (\alpha_{l-1}(\alpha_{l-1}I + \mathcal{C})^{-1} - \alpha_{k-1}(\alpha_{k-1}I + \mathcal{C})^{-1})(x_0 - x^\dagger)$ .

Combining (47), (48) and (49) gives

$$(50) \quad \|x_k - x_l\| \leq \|J\| + (2 + \tau)K_0\|x_0 - x^\dagger\|(\|x_{k-1} - x^\dagger\| + \|x_{l-1} - x^\dagger\|).$$

Next we estimate  $J$ . By introducing the notation

$$\begin{aligned} J_1 &:= \left(1 - \frac{\alpha_{k-1}}{\alpha_{l-1}}\right) (\alpha_{k-1}I + \mathcal{C})^{-1} F'(x^\dagger)^* (F(x_l) - y), \\ J_2 &:= \left(1 - \frac{\alpha_{k-1}}{\alpha_{l-1}}\right) (\alpha_{k-1}I + \mathcal{C})^{-1} F'(x^\dagger)^* (F'(x^\dagger)(x_l - x^\dagger) - F(x_l) + y), \\ J_3 &:= \left(1 - \frac{\alpha_{k-1}}{\alpha_{l-1}}\right) (\alpha_{k-1}I + \mathcal{C})^{-1} \mathcal{C} (\alpha_{l-1}(\alpha_{l-1}I + \mathcal{C})^{-1}(x_0 - x^\dagger) - (x_l - x^\dagger)), \end{aligned}$$

we have  $J = J_1 + J_2 + J_3$ . Obviously Assumption 2.1 and (30) imply

$$(51) \quad \|J_2\| \leq \frac{K_0}{2}\|x_l - x^\dagger\|^2 \leq \tau K_0\|x_0 - x^\dagger\|\|x_l - x^\dagger\|.$$

By inserting the expression of  $x_l - x^\dagger$  (i.e. (27) with  $k = l - 1$ ) into  $J_3$  we have from Assumption 2.1, Lemma 4.2 and (30) that

$$\begin{aligned}
 \|J_3\| &\leq \|\alpha_{l-1}(\alpha_{l-1}I + \mathcal{C})^{-1}(x_0 - x^\dagger) - (x_l - x^\dagger)\| \\
 &\leq \|\alpha_{l-1}((\alpha_{l-1}I + \mathcal{C})^{-1} - (\alpha_{l-1}I + \mathcal{A}_{l-1})^{-1})(x_0 - x^\dagger)\| \\
 (52) \quad &+ \|(\alpha_{l-1}I + \mathcal{A}_{l-1})^{-1}F'(x_{l-1})^*(F(x_{l-1}) - y - F'(x_{l-1})(x_{l-1} - x^\dagger))\| \\
 &\leq 2K_0\|x_0 - x^\dagger\|\|x_{l-1} - x^\dagger\| + \frac{K_0}{2}\|x_{l-1} - x^\dagger\|^2 \\
 &\leq (2 + \tau)K_0\|x_0 - x^\dagger\|\|x_{l-1} - x^\dagger\|.
 \end{aligned}$$

To estimate the term  $J_1$ , we use the abbreviations

$$\mathcal{K} := (\alpha_{k-1}I + \mathcal{C})^{-\frac{1}{2}}F'(x^\dagger)^*$$

and

$$\mathcal{L} := \left(\frac{\alpha_{k-1}}{\alpha_{l-1}}\right)^{\frac{1}{2}}(\alpha_{k-1}I + \mathcal{D})^{-\frac{1}{2}}(\alpha_{l-1}I + \mathcal{D})^{\frac{1}{2}},$$

and write  $J_1$  as

$$J_1 = \frac{1}{\sqrt{\alpha_{k-1}}}\left(1 - \frac{\alpha_{k-1}}{\alpha_{l-1}}\right)\mathcal{K}\mathcal{L}\alpha_{l-1}^{\frac{1}{2}}(\alpha_{l-1}I + \mathcal{D})^{-\frac{1}{2}}(F(x_l) - y).$$

Let  $\{\hat{E}_\lambda\}$  be the spectral family generated by  $\mathcal{D}$ ; then for any  $v \in Y$  we have

$$\|\mathcal{L}v\|^2 = \int_0^\infty \frac{\alpha_{k-1}(\alpha_{l-1} + \lambda)}{\alpha_{l-1}(\alpha_{k-1} + \lambda)}d\|\hat{E}_\lambda v\|^2.$$

Since  $\alpha_{l-1} \geq \alpha_{k-1}$ , the function

$$g(\lambda) := \frac{\alpha_{k-1}(\alpha_{l-1} + \lambda)}{\alpha_{l-1}(\alpha_{k-1} + \lambda)}$$

is monotonically decreasing on  $[0, \infty)$  and attains its maximum 1 at  $\lambda = 0$ . Therefore

$$\|\mathcal{L}v\|^2 \leq \int_0^\infty d\|\hat{E}_\lambda v\|^2 = \|v\|^2, \quad \forall v \in Y.$$

This implies  $\|\mathcal{L}\| \leq 1$ . By the same procedure we have  $\|\mathcal{K}\| \leq 1$ . Hence

$$\|J_1\| \leq \frac{1}{\sqrt{\alpha_{k-1}}}\|\alpha_{l-1}^{\frac{1}{2}}(\alpha_{l-1}I + \mathcal{D})^{-\frac{1}{2}}(F(x_l) - y)\|.$$

Similarly to the derivation of (38), we have

$$\|\alpha_{l-1}^{\frac{1}{2}}(\alpha_{l-1}I + \mathcal{D})^{-\frac{1}{2}}(F(x_l) - y)\| \leq r^{\frac{1}{2}}\|\alpha_l^{\frac{1}{2}}(\alpha_lI + \mathcal{D})^{-\frac{1}{2}}(F(x_l) - y)\|.$$

Therefore, by noting that  $\alpha_k \leq \alpha_{k-1}$  it follows that

$$(53) \quad \|J_1\| \leq \frac{r^{\frac{1}{2}}}{\sqrt{\alpha_k}}\|\alpha_l^{\frac{1}{2}}(\alpha_lI + \mathcal{D})^{-\frac{1}{2}}(F(x_l) - y)\|.$$

Thus the combination of (50)–(53) gives

$$\begin{aligned}
 \|x_k - x_l\| &\leq \frac{r^{\frac{1}{2}}}{\sqrt{\alpha_k}}\|\alpha_l^{\frac{1}{2}}(\alpha_lI + \mathcal{D})^{-\frac{1}{2}}(F(x_l) - y)\| + (2 + \tau)K_0\|x_0 - x^\dagger\|\|x_{k-1} - x^\dagger\| \\
 &+ (4 + 2\tau)K_0\|x_0 - x^\dagger\|\|x_{l-1} - x^\dagger\| + \tau K_0\|x_0 - x^\dagger\|\|x_l - x^\dagger\|.
 \end{aligned}$$

Noting that  $12rK_0\|x_0 - x^\dagger\| \leq 1$ , Lemma 4.3 can be used to obtain

$$\begin{aligned} \|x_k - x_l\| &\leq \frac{r^{\frac{1}{2}}}{\sqrt{\alpha_k}} \|\alpha_l^{\frac{1}{2}}(\alpha_l I + \mathcal{D})^{-\frac{1}{2}}(F(x_l) - y)\| + \frac{2 + \tau}{6} \|x_k - x^\dagger\| \\ &\quad + \left(\frac{\tau}{12r} + \frac{\tau + 2}{3}\right) \|x_l - x^\dagger\|. \end{aligned}$$

Since  $\frac{\tau}{12r} + \frac{\tau+2}{3} < 1$ , assertion (46) follows.

For the case  $l = 0$ , we can assume  $k \geq 1$ . Since (46) is valid for  $l = 1$ , we can use (36) to assert that (46) is also true for  $l = 0$ .  $\square$

### 5. PROOFS OF MAIN RESULTS

In this section we shall prove Theorem 2.1 and its corollaries. The proof is based on Lemma 4.4 and other two auxiliary results given below. The first concerns the stability estimate for the iteratively regularized Gauss-Newton method. It is obvious that  $x_k^\delta \rightarrow x_k$  as  $\delta \rightarrow 0$  for each fixed  $k$ , which can be confirmed by induction since Assumption 2.1 implies the continuity of the mapping  $x \mapsto F'(x)$  on  $B_p(x^\dagger)$ . This, however, is not sufficient for our purpose, we hope to obtain a finer estimate on  $\|x_k^\delta - x_k\|$ . The following result gives a satisfactory answer.

**Lemma 5.1.** *Let Assumption 2.1 hold,  $12K_0\|x_0 - x^\dagger\| \leq 1$ , and let  $\tilde{k}_\delta$  be the integer defined by (12). Then, for all  $0 \leq k \leq \tilde{k}_\delta$ ,*

$$(54) \quad \|x_k^\delta - x_k\| \leq \frac{\delta}{\sqrt{\alpha_k}}.$$

*Proof.* Since (54) is trivial for  $k = 0$ , therefore if we can establish the estimate

$$(55) \quad \|x_{k+1}^\delta - x_{k+1}\| \leq \frac{\delta}{2\sqrt{\alpha_k}} + \frac{1}{2} \|x_k^\delta - x_k\|$$

for all integers  $0 \leq k < \tilde{k}_\delta$ , then the proof can be complete by a simple application of Lemma 4.1.

To prove (55), we subtract (27) from (14), to obtain

$$\begin{aligned} (56) \quad x_{k+1}^\delta - x_{k+1} &= ((\alpha_k I + \mathcal{A}_k)^{-1} F'(x_k)^* u_k - (\alpha_k I + \mathcal{A}_k^\delta)^{-1} F'(x_k^\delta)^* u_k^\delta) \\ &\quad + \alpha_k ((\alpha_k I + \mathcal{A}_k^\delta)^{-1} - (\alpha_k I + \mathcal{A}_k)^{-1}) (x_0 - x^\dagger) \\ &\quad + (\alpha_k I + \mathcal{A}_k^\delta)^{-1} F'(x_k^\delta)^* (y^\delta - y) \\ &=: I_1 + I_2 + I_3, \end{aligned}$$

where we used the abbreviations

$$\begin{aligned} u_k &= F(x_k) - y - F'(x_k)(x_k - x^\dagger), \\ u_k^\delta &= F(x_k^\delta) - y - F'(x_k^\delta)(x_k^\delta - x^\dagger). \end{aligned}$$

In what follows we estimate the three terms  $I_1, I_2$  and  $I_3$ . Obviously we have

$$(57) \quad \|I_2\| \leq 2K_0\|x_0 - x^\dagger\| \|x_k^\delta - x_k\|, \quad \|I_3\| \leq \frac{\delta}{2\sqrt{\alpha_k}};$$

here we used (34) to obtain the first estimate. To estimate  $I_1$ , we write

$$\begin{aligned} I_1 &= \{(\alpha_k I + \mathcal{A}_k)^{-1} F'(x_k)^* u_k - (\alpha_k I + \mathcal{A}_k^\delta)^{-1} F'(x_k^\delta)^* u_k\} \\ &\quad + (\alpha_k I + \mathcal{A}_k^\delta)^{-1} F'(x_k^\delta)^* (u_k - u_k^\delta) \\ &=: I_1^{(1)} + I_1^{(2)}. \end{aligned}$$

Since Assumption 2.1 implies

$$(F'(x_k^\delta) - F'(x_k))(x_k - x^\dagger) = F'(x_k^\delta)h(x_k, x_k^\delta, x^\dagger - x_k)$$

and

$$(58) \quad F(x_k) - F(x_k^\delta) - F'(x_k^\delta)(x_k - x_k^\delta) = F'(x_k^\delta) \int_0^1 m_t^\delta dt$$

with  $m_t^\delta = h(x_k^\delta + t(x_k - x_k^\delta), x_k^\delta, x_k - x_k^\delta)$  and  $\|\int_0^1 m_t^\delta dt\| \leq \frac{K_0}{2} \|x_k - x_k^\delta\|^2$ , we can obtain

$$(59) \quad \begin{aligned} \|I_1^{(2)}\| &\leq \left\| (\alpha_k I + \mathcal{A}_k^\delta)^{-1} \mathcal{A}_k^\delta \left( h(x_k, x_k^\delta, x^\dagger - x_k) + \int_0^1 m_t^\delta dt \right) \right\| \\ &\leq \frac{K_0}{2} (2\|x_k - x^\dagger\| + \|x_k^\delta - x_k\|) \|x_k^\delta - x_k\|. \end{aligned}$$

By applying (28) and Assumption 2.1 we also have

$$\begin{aligned} I_1^{(1)} &= ((\alpha_k I + \mathcal{A}_k)^{-1} \mathcal{A}_k - (\alpha_k I + \mathcal{A}_k^\delta)^{-1} \mathcal{A}_k^\delta) \int_0^1 h_t dt \\ &\quad + (\alpha_k I + \mathcal{A}_k^\delta)^{-1} F'(x_k^\delta)^* \int_0^1 (F'(x_k^\delta) - F'(x_k)) h_t dt \\ &= \alpha_k \{(\alpha_k I + \mathcal{A}_k^\delta)^{-1} - (\alpha_k I + \mathcal{A}_k)^{-1}\} \int_0^1 h_t dt \\ &\quad - (\alpha_k I + \mathcal{A}_k^\delta)^{-1} \mathcal{A}_k^\delta \int_0^1 h(x_k, x_k^\delta, h_t) dt. \end{aligned}$$

Hence the application of (34) gives

$$(60) \quad \begin{aligned} \|I_1^{(1)}\| &\leq 2K_0 \|x_k^\delta - x_k\| \left\| \int_0^1 h_t dt \right\| + \left\| \int_0^1 h(x_k, x_k^\delta, h_t) dt \right\| \\ &\leq \frac{3}{2} K_0^2 \|x_k - x^\dagger\|^2 \|x_k^\delta - x_k\|. \end{aligned}$$

Combining (59) and (60) and noting that  $12K_0 \|x_0 - x^\dagger\| \leq 1$ , from (17) and (30) we have

$$(61) \quad \|I_1\| \leq 4K_0 \|x_0 - x^\dagger\| \|x_k - x_k^\delta\|.$$

Now (55) follows from the combination of (56), (57) and (61). □

Our next auxiliary result contributes to the estimates of some terms.

**Lemma 5.2.** *Let Assumption 2.1 hold,  $12K_0 \|x_0 - x^\dagger\| \leq 1$ ,  $c \geq \frac{25}{4}$ , and let  $k_\delta$  be the integer determined by Rule 2.1. Then*

$$(62) \quad \alpha_{k_\delta} (F(x_{k_\delta}) - y, (\alpha_{k_\delta} I + \mathcal{D})^{-1} (F(x_{k_\delta}) - y)) \leq c_1^2 \delta^2.$$

Moreover, if  $k_\delta > 0$  then for all integers  $0 \leq k < k_\delta$ ,

$$(63) \quad \alpha_k (F(x_k) - y, (\alpha_k I + \mathcal{D})^{-1} (F(x_k) - y)) \geq c_2^2 \delta^2,$$

where

$$c_1 := \sqrt{\frac{5}{3}} \left( \sqrt{c} + 2 + \frac{r^{\frac{1}{2}}}{12(\sqrt{c}-1)} \right) \quad \text{and} \quad c_2 := \sqrt{\frac{3}{5}} \left( \sqrt{c} - 2 - \frac{1}{12(\sqrt{c}-1)} \right).$$

*Proof.* Let  $\tilde{k}_\delta$  be the integer defined by (12); then  $k_\delta \leq \tilde{k}_\delta$ . By using (58) and Lemma 5.1 it follows that, for all integers  $0 \leq k \leq \tilde{k}_\delta$ ,

$$\begin{aligned} \sqrt{\alpha_k} \|(\alpha_k I + \mathcal{B}_k^\delta)^{-\frac{1}{2}}(F(x_k^\delta) - F(x_k))\| &\leq \sqrt{\alpha_k} \left( 1 + \frac{K_0}{2} \|x_k^\delta - x_k\| \right) \|x_k^\delta - x_k\| \\ &\leq \left( 1 + \frac{K_0 \delta}{2\sqrt{\alpha_k}} \right) \delta. \end{aligned}$$

Since the definition of  $\tilde{k}_\delta$  and (5) imply

$$\frac{\delta}{\sqrt{\alpha_{\tilde{k}_\delta}}} \leq \frac{r^{\frac{1}{2}} \delta}{\sqrt{\alpha_{\tilde{k}_\delta-1}}} \leq \frac{2r^{\frac{1}{2}} \|x_0 - x^\dagger\|}{\sqrt{c}-1} \quad \text{and} \quad \frac{\delta}{\sqrt{\alpha_k}} \leq \frac{2\|x_0 - x^\dagger\|}{\sqrt{c}-1}$$

for all  $0 \leq k < \tilde{k}_\delta$ , we have for  $0 \leq k < \tilde{k}_\delta$

$$\begin{aligned} \sqrt{\alpha_{k_\delta}} \|(\alpha_{k_\delta} I + \mathcal{B}_{k_\delta}^\delta)^{-\frac{1}{2}}(F(x_{k_\delta}^\delta) - F(x_{k_\delta}))\| &\leq \left( 1 + \frac{r^{\frac{1}{2}}}{12(\sqrt{c}-1)} \right) \delta, \\ \sqrt{\alpha_k} \|(\alpha_k I + \mathcal{B}_k^\delta)^{-\frac{1}{2}}(F(x_k^\delta) - F(x_k))\| &\leq \left( 1 + \frac{1}{12(\sqrt{c}-1)} \right) \delta. \end{aligned}$$

Thus we can use the definition of  $k_\delta$  and (2) to obtain for  $0 \leq k < k_\delta$  that

$$(64) \quad \sqrt{\alpha_{k_\delta}} \|(\alpha_{k_\delta} I + \mathcal{B}_{k_\delta}^\delta)^{-\frac{1}{2}}(F(x_{k_\delta}) - y)\| \leq \left( \sqrt{c} + 2 + \frac{r^{\frac{1}{2}}}{12(\sqrt{c}-1)} \right) \delta,$$

$$(65) \quad \sqrt{\alpha_k} \|(\alpha_k I + \mathcal{B}_k^\delta)^{-\frac{1}{2}}(F(x_k) - y)\| \geq \left( \sqrt{c} - 2 - \frac{1}{12(\sqrt{c}-1)} \right) \delta.$$

Let us now introduce for  $0 \leq k \leq k_\delta$  the notation

$$\begin{aligned} a_k &:= \alpha_k (F(x_k) - y, (\alpha_k I + \mathcal{B}_k^\delta)^{-1}(F(x_k) - y)), \\ b_k &:= \alpha_k (F(x_k) - y, (\alpha_k I + \mathcal{D})^{-1}(F(x_k) - y)). \end{aligned}$$

Since (17) implies  $K_0 \|x_k^\delta - x^\dagger\| \leq 3K_0 \|x_0 - x^\dagger\| \leq \frac{1}{4}$ , we can exploit [19, Lemma 3.6] to obtain

$$\begin{aligned} |a_k - b_k| &= |\alpha_k (F(x_k) - y, ((\alpha_k I + \mathcal{B}_k^\delta)^{-1} - (\alpha_k I + \mathcal{D})^{-1})(F(x_k) - y))| \\ &= |\alpha_k (F(x_k) - y, (\alpha_k I + \mathcal{D})^{-1}(\mathcal{D} - \mathcal{B}_k^\delta)(\alpha_k I + \mathcal{B}_k^\delta)^{-1}(F(x_k) - y))| \\ &= |\alpha_k ((\alpha_k I + \mathcal{D})^{-\frac{1}{2}}(F(x_k) - y), (\alpha_k I + \mathcal{D})^{-\frac{1}{2}}(\mathcal{D} - \mathcal{B}_k^\delta)(\alpha_k I + \mathcal{B}_k^\delta)^{-\frac{1}{2}} \\ &\quad \times (\alpha_k I + \mathcal{B}_k^\delta)^{-\frac{1}{2}}(F(x_k) - y))| \\ &\leq 2K_0 \alpha_k \|x_k^\delta - x^\dagger\| \|(\alpha_k I + \mathcal{D})^{-\frac{1}{2}}(F(x_k) - y)\| \\ &\quad \times \|(\alpha_k I + \mathcal{B}_k^\delta)^{-\frac{1}{2}}(F(x_k) - y)\| \\ &\leq \frac{1}{4} \alpha_k \left\{ (F(x_k) - y, (\alpha_k I + \mathcal{D})^{-1}(F(x_k) - y)) \right. \\ &\quad \left. + (F(x_k) - y, (\alpha_k I + \mathcal{B}_k^\delta)^{-1}(F(x_k) - y)) \right\} \\ &= \frac{1}{4} (a_k + b_k), \end{aligned}$$

which implies  $\frac{2}{3}b_k \leq a_k \leq \frac{5}{3}b_k$ . This together with (64) and (65) gives the desired results.  $\square$

Now we are ready to give the proofs of Theorem 2.1 and its corollaries.

*Proof of Theorem 2.1.* Since

$$\frac{\delta}{\sqrt{\alpha_k}} \geq \frac{\delta}{\sqrt{\alpha_{\tilde{k}_\delta}}} \geq \frac{2}{\sqrt{c-1}} \|x_0 - x^\dagger\|$$

for all integers  $k > \tilde{k}_\delta$ , and since  $\|x_{\tilde{k}_\delta}^\delta - x^\dagger\| \leq 3\|x_0 - x^\dagger\|$ , we need only to prove there is a constant  $C > 0$  independent of  $\delta$  such that

$$\|x_{\tilde{k}_\delta}^\delta - x^\dagger\| \leq C \inf \left\{ \|x_k - x^\dagger\| + \frac{\delta}{\sqrt{\alpha_k}} : 0 \leq k \leq \tilde{k}_\delta \right\},$$

which, using Lemma 5.1, can be confirmed by showing that for all integers  $k \leq \tilde{k}_\delta$

$$(66) \quad \|x_{k_\delta} - x^\dagger\| + \frac{\delta}{\sqrt{\alpha_{k_\delta}}} \leq C \left\{ \|x_k - x^\dagger\| + \frac{\delta}{\sqrt{\alpha_k}} \right\}.$$

In the following, we carry out the proof of (66) by considering the two cases  $k_\delta \leq k \leq \tilde{k}_\delta$  and  $0 \leq k < k_\delta$  separately.

(i) For the case  $k_\delta \leq k \leq \tilde{k}_\delta$ , we obviously have  $\delta/\sqrt{\alpha_{k_\delta}} \leq \delta/\sqrt{\alpha_k}$ . Since (62) implies

$$\|\alpha_{\tilde{k}_\delta}^{\frac{1}{2}}(\alpha_{k_\delta}I + \mathcal{D})^{-\frac{1}{2}}(F(x_{k_\delta}) - y)\| \leq c_1\delta,$$

we can use Lemma 4.4 to obtain

$$\|x_{k_\delta} - x^\dagger\| \leq C_0 \left\{ \|x_k - x^\dagger\| + \frac{c_1\delta}{\sqrt{\alpha_k}} \right\}.$$

Therefore (66) is true for this case.

(ii) Next we consider the case  $0 \leq k < k_\delta$ . By using the well-known fact that the function  $\alpha \mapsto \|\alpha(\alpha I + \mathcal{C})^{-1}(x_0 - x^\dagger)\|$  is monotonically increasing on  $[0, \infty)$ , from Lemma 4.3 we have for all integers  $l \geq m$  that

$$(67) \quad \begin{aligned} \|x_l - x^\dagger\| &\leq \frac{4}{3}r \|\alpha_l(\alpha_l I + \mathcal{C})^{-1}(x_0 - x^\dagger)\| \\ &\leq \frac{4}{3}r \|\alpha_m(\alpha_m I + \mathcal{C})^{-1}(x_0 - x^\dagger)\| \\ &\leq 2r \|x_m - x^\dagger\|, \end{aligned}$$

which in particular implies

$$(68) \quad \|x_{k_\delta} - x^\dagger\| \leq 2r \|x_k - x^\dagger\|.$$

Since  $12rK_0\|x_0 - x^\dagger\| \leq 1$ , we can exploit (63), (67) and Assumption 2.1 to obtain

$$\begin{aligned} c_2\delta &\leq \sqrt{\alpha_{k_\delta-1}} \|(\alpha_{k_\delta-1}I + \mathcal{D})^{-\frac{1}{2}}(F(x_{k_\delta-1}) - y)\| \\ &\leq \sqrt{\alpha_{k_\delta-1}} \|(\alpha_{k_\delta-1}I + \mathcal{D})^{-\frac{1}{2}}F'(x^\dagger)(x_{k_\delta-1} - x^\dagger + \int_0^1 \tilde{h}_t dt)\| \\ &\leq \sqrt{\alpha_{k_\delta-1}} \left(1 + \frac{K_0}{2}\|x_{k_\delta-1} - x^\dagger\|\right) \|x_{k_\delta-1} - x^\dagger\| \\ &\leq \sqrt{\alpha_{k_\delta-1}} (1 + K_0\|x_0 - x^\dagger\|) \|x_{k_\delta-1} - x^\dagger\| \\ &\leq \frac{12r^{\frac{3}{2}} + r^{\frac{1}{2}}}{6} \sqrt{\alpha_{k_\delta}} \|x_k - x^\dagger\|, \end{aligned}$$

where  $\tilde{h}_t := h(x^\dagger + t(x_{k_\delta-1} - x^\dagger), x^\dagger, x_{k_\delta-1} - x^\dagger)$ . Clearly, this implies

$$(69) \quad \frac{\delta}{\sqrt{\alpha_{k_\delta}}} \leq \frac{12r^{\frac{3}{2}} + r^{\frac{1}{2}}}{6c_2} \|x_k - x^\dagger\|.$$

The combination of (68) and (69) gives the proof of (66) again. □

*Proof of Corollary 2.1.* We first prove assertion (20). By using (35) we can write (19) in the following form:

$$(70) \quad \|x_{k_\delta}^\delta - x^\dagger\| \leq C \inf \left\{ \|\alpha_k(\alpha_k I + \mathcal{C})^{-1}(x_0 - x^\dagger)\| + \frac{\delta}{\sqrt{\alpha_k}} : k=0, 1, \dots \right\}.$$

Here and later  $C$  denotes a generic constant independent of  $\delta$ . If we choose  $m_\delta$  to be the first integer such that  $\alpha_{m_\delta} \leq \delta$ , then

$$(71) \quad \|x_{k_\delta}^\delta - x^\dagger\| \leq C \left\{ \|\alpha_{m_\delta}(\alpha_{m_\delta} I + \mathcal{C})^{-1}(x_0 - x^\dagger)\| + \frac{\delta}{\sqrt{\alpha_{m_\delta}}} \right\}.$$

Since  $m_\delta \rightarrow \infty$  as  $\delta \rightarrow 0$ , we have  $\delta/\sqrt{\alpha_{m_\delta}} \rightarrow 0$  and  $\|\alpha_{m_\delta}(\alpha_{m_\delta} I + \mathcal{C})^{-1}(x_0 - x^\dagger)\| \rightarrow 0$ . Therefore  $x_{k_\delta}^\delta \rightarrow x^\dagger$  which follows from (71).

To prove assertion (21), recalling the well-known fact that

$$(72) \quad \|\alpha_k(\alpha_k I + \mathcal{C})^{-1}(x_0 - x^\dagger)\| \leq \|\omega\| \alpha_k^\nu$$

if  $x_0 - x^\dagger$  satisfies (6) with  $0 < \nu \leq 1$ , we have from (70) that

$$\|x_{k_\delta}^\delta - x^\dagger\| \leq C \inf \left\{ \|\omega\| \alpha_k^\nu + \frac{\delta}{\sqrt{\alpha_k}} : k = 0, 1, \dots \right\}.$$

This suggests that if we choose the integer  $\bar{k}_\delta$  to be such that

$$\alpha_{\bar{k}_\delta}^{\nu+\frac{1}{2}} \leq \frac{\delta}{\|\omega\|} < \alpha_k^{\nu+\frac{1}{2}}, \quad 0 \leq k < \bar{k}_\delta,$$

then

$$\|x_{k_\delta}^\delta - x^\dagger\| \leq C \frac{\delta}{\sqrt{\alpha_{\bar{k}_\delta}}} \leq C \frac{\delta}{\sqrt{\alpha_{\bar{k}_\delta-1}}} \leq C \|\omega\|^{\frac{1}{1+2\nu}} \delta^{\frac{2\nu}{1+2\nu}},$$

and the proof follows. □

*Proof of Corollary 2.2.* If  $x_0 = x^\dagger$ , then (70) says that  $\|x_{k_\delta}^\delta - x^\dagger\| \leq O(\delta)$ , and the assertion is trivial. Therefore in what follows we assume  $x_0 \neq x^\dagger$ . We choose  $\tilde{m}_\delta$  to be the first integer such that

$$(73) \quad \sqrt{\alpha_{\tilde{m}_\delta}} \|x_{\tilde{m}_\delta} - x^\dagger\| \leq \delta.$$

The existence of  $\tilde{m}_\delta$  is guaranteed because the sequence  $\{\rho_k\} := \{\sqrt{\alpha_k} \|x_k - x^\dagger\|\}$  has the property  $\rho_k \rightarrow 0$  as  $k \rightarrow \infty$ . Moreover we have  $\tilde{m}_\delta \rightarrow \infty$  by an easy exercise. With this  $\tilde{m}_\delta$  we have from (19) that

$$(74) \quad \|x_{k_\delta}^\delta - x^\dagger\| \leq \frac{C\delta}{\sqrt{\alpha_{\tilde{m}_\delta}}} \leq \frac{C\delta}{\sqrt{\alpha_{\tilde{m}_\delta-1}}}.$$

By using the notation  $c_\nu(k) := \|x_k - x^\dagger\|/\alpha_k^\nu$ , from (73) it follows that

$$\alpha_{\tilde{m}_\delta-1} \geq \left( \frac{\delta}{c_\nu(\tilde{m}_\delta-1)} \right)^{\frac{1+2\nu}{2}}.$$

Therefore (74) gives

$$(75) \quad \|x_{k_\delta}^\delta - x^\dagger\| \leq C c_\nu (\tilde{m}_\delta - 1)^{\frac{1}{1+2\nu}} \delta^{\frac{2\nu}{1+2\nu}}.$$

Since  $\tilde{m}_\delta \rightarrow \infty$ , by using Proposition 4.1 we have  $c_\nu(\tilde{m}_\delta - 1) = O(1)$  if  $x_0 - x^\dagger$  satisfies (22), and  $c_\nu(\tilde{m}_\delta - 1) = o(1)$  if  $x_0 - x^\dagger$  satisfies (23). Therefore (75) implies (21).  $\square$

#### ACKNOWLEDGMENTS

I would like to thank Dr. U. Amato for teaching me the Matlab program used in the numerical computation during my visit to the Istituto per Applicazioni della Matematica CNR in Napoli, which was sponsored by the Italian Ministry of Foreign Affairs. Sincere thanks also goes to the referee for constructive suggestions which lead to this improved version. This work is supported by the National Natural Science Foundation of China under grant 19801018.

#### REFERENCES

- [1] A. B. Bakushinskii, *The problems of the convergence of the iteratively regularized Gauss-Newton method*, Comput. Math. Math. Phys., 32(1992), 1353–1359. MR **93k**:65049
- [2] H. T. Banks and K. Kunisch, *Estimation Techniques for Distributed Parameter Systems*, Basel: Birkhäuser, 1989. MR **91b**:93085
- [3] B. Blaschke, A. Neubauer and O. Scherzer, *On convergence rates for the iteratively regularized Gauss-Newton method*, IMA J. Numer. Anal., 17(1997), 421–436. MR **98f**:65066
- [4] F. Colonus and K. Kunisch, *Stability for parameter estimation in two point boundary value problems*, J. Reine Angew. Math., 370(1986), 1–29. MR **88j**:93027
- [5] H. W. Engl, *Regularization methods for the stable solutions of inverse problems*, Surv. Math. Ind., 3(1993), 71–143. MR **94g**:65064
- [6] H. W. Engl and H. Gfrerer, *A posteriori parameter choice for general regularization methods for solving ill-posed problems*, Appl. Numer. Math., 4(1988), 395–417. MR **89i**:65060
- [7] H. W. Engl, K. Kunisch and A. Neubauer, *Convergence rates for Tikhonov regularization of nonlinear ill-posed problems*, Inverse Problems, 5(1989), 523–540. MR **91k**:65102
- [8] H. Gfrerer, *An a posteriori parameter choice for ordinary and iterated Tikhonov regularization of ill-posed problems leading to optimal convergence rates*, Math. Comp., 49(1987), 507–522. MR **88k**:65049
- [9] C. W. Groetsch, *The Theory of Tikhonov Regularization for Fredholm Equation of the First Kind*, (Boston, MA: Pitman), 1984. MR **85k**:45020
- [10] C. W. Groetsch, *Inverse Problems in Mathematical Sciences*, Vieweg, Wiesbaden, 1993. MR **94m**:00008
- [11] M. Hanke, A. Neubauer and O. Scherzer, *A convergence analysis of Landweber iteration for nonlinear ill-posed problems*, Numer. Math., 72(1995), 21–37. MR **96i**:65046
- [12] B. Hofmann, *Regularization for Applied Inverse and Ill-Posed Problems*, Leipzig, Teubner, 1986. MR **88i**:65001
- [13] Q. N. Jin and Z. Y. Hou, *On an a posteriori parameter choice strategy for Tikhonov regularization of nonlinear ill-posed problems*, Numer. Math., 83(1999), 139–159. CMP 99:16
- [14] B. Kaltenbacher, *Some Newton-type methods for the regularization of nonlinear ill-posed problems*, Inverse Problems, 13(1997), 729–754. MR **98h**:65025
- [15] A. K. Louis, *Inverse und Schlecht Gestellte Probleme*, Teubner, Stuttgart, 1989. MR **90g**:65075
- [16] A. Neubauer, *On converse and saturation results for Tikhonov regularization of linear ill-posed problems*, SIAM J. Numer. Anal., 34(1997), 517–527. MR **98d**:65081
- [17] R. Plato and H. Hämarik, *On pseudo-optimal parameter choice and stopping rules for regularization methods in Banach spaces*, Numer. Funct. Anal. Optimiz., 17(1996), 181–195. MR **97g**:65124
- [18] O. Scherzer, *A convergence analysis of a method of steepest descent and a two-step algorithm for nonlinear ill-posed problems*, Numer. Funct. Anal. Optimiz., 17(1996), 197–214. MR **97g**:65125

- [19] O. Scherzer, H. W. Engl and K. Kunisch, *Optimal a posteriori parameter choice for Tikhonov regularization for solving nonlinear ill-posed problems*, SIAM J. Numer. Anal., 30(1993), 1796–1838. MR **95a**:65104
- [20] T. I. Seidman and C. R. Vogel, *Well-posedness and convergence of some regularization methods for nonlinear ill-posed problems*, Inverse Problems, 5(1989), 227–238. MR **90d**:65117
- [21] A. N. Tikhonov and V. Y. Arsenin, *Solutions of ill-posed problems*. Winston, Washington, DC, 1977. MR **56**:13604
- [22] G. M. Vainikko and A. Y. Veretennikov, *Iteration Procedures in Ill-Posed Problems* (in Russian), Nauka, Moscow, 1986. MR **88c**:47019
- [23] V. V. Vasin and A. L. Ageev, *Ill-Posed Problems with A Priori Information*, Inverse and Ill-Posed Problems Series, VSP, Utrecht, The Netherlands, 1995. MR **97j**:65100

INSTITUTE OF MATHEMATICS, NANJING UNIVERSITY, NANJING 210008, P. R. CHINA  
E-mail address: galgebra@nju.edu.cn