Numer. Math. (2010) 115:229–259
DOI 10.1007/s00211-009-0275-x

Numerische Mathematik

On a regularized Levenberg–Marquardt method for solving nonlinear inverse problems

Qinian Jin

Received: 5 January 2009 / Revised: 8 June 2009 / Published online: 18 November 2009

Abstract We consider a regularized Levenberg–Marquardt method for solving nonlinear ill-posed inverse problems. We use the discrepancy principle to terminate the iteration. Under certain conditions, we prove the convergence of the method and obtain the order optimal convergence rates when the exact solution satisfies suitable source-wise representations.

Mathematics Subject Classification (2000) 65J15 · 65J20 · 47H17

1 Introduction

In this paper we will consider the equations

$$F(x) = y, \tag{1.1}$$

arising from nonlinear inverse problems, where $F : D(F) \subset X \mapsto Y$ is a nonlinear Fréchet differentiable operator between two Hilbert spaces X and Y whose norms and inner products are denoted as $\|\cdot\|$ and $(\cdot,\cdot)$ respectively. We assume that (1.1) has a solution $x^\dagger$ in the domain $D(F)$ of F such that

$$F(x^\dagger) = y.$$

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

Present Address:
Q. Jin (✉)
Department of Mathematics, Virginia Tech, Blacksburg, VA 24061, USA
e-mail: qnjin@math.vt.edu; qjin@math.utexas.edu

A characteristic property of such equations is their ill-posedness in the sense that their solutions do not depend continuously on the data. Since the right hand side is usually obtained by measurement, thus, instead of y itself, the available data is an approximation y^δ satisfying

$$\|y^\delta - y\| \le \delta \tag{1.2}$$

with a given small noise level $\delta > 0$. Due to the ill-posedness, the computation of a stable solution of (1.1) from y^δ becomes an important issue.

Many regularization methods have been suggested for solving (1.1) in a stable way using the available data y^δ, see [1,3,6–11] and the references therein. In [3] Hanke considered the Levenberg–Marquardt method which defines the iterative solutions $\{x_k^\delta\}$ by

$$x_{k+1}^\delta = x_k^\delta - \left(\alpha_k I + F'(x_k^\delta)^* F'(x_k^\delta)\right)^{-1} F'(x_k^\delta)^* \left(F(x_k^\delta) - y^\delta\right), \tag{1.3}$$

where $x_0^\delta := x_0 \in D(F)$ is an initial guess of $x^\dagger$, $F'(x)$ denotes the Fréchet derivative of F at $x \in D(F)$, and $F'(x)^*$ denotes the adjoint of $F'(x)$. Under certain conditions on F, it has been shown in [3] that if the sequence $\{\alpha_k\}$ of positive numbers is determined by a suitable adaptive strategy during computations and the discrepancy principle

$$\|F(x_{k_\delta}^\delta) - y^\delta\| \le \tau\delta < \|F(x_k^\delta) - y^\delta\|, \quad 0 \le k < k_\delta, \tag{1.4}$$

with $\tau > 1$, is used as a stopping rule, then $x_{k_\delta}^\delta$ converges to $x^\dagger$ as $\delta \to 0$. Further investigations have been made by Rieder in [11,12] by establishing rates of convergence under the source conditions

$$x_0 - x^\dagger = \left(F'(x^\dagger)^* F'(x^\dagger)\right)^\nu \omega \tag{1.5}$$

with $\nu > 0$. Under the condition

$$F'(z) = Q(x,z)F'(x) \quad \text{and} \quad \|I - Q(x,z)\| \le K_0\|x - z\| \tag{1.6}$$

for all x, z in a neighborhood of $x^\dagger$, it has been proved in [12] that there is a problem dependent number $0 < \eta < 1/2$ such that the convergence rate $O(\delta^{2(\nu-\eta)/(1+2\nu)})$ holds if $x_0 - x^\dagger$ satisfies (1.5) with $\eta < \nu \le 1/2$. This is not the optimal rate of convergence. It is not yet clear if the optimal convergence rate can be established under weaker source conditions. Thus, the investigation on convergence rates is still far from complete.[1]

1 Recently, Hanke established in [4] the order optimality of the regularizing Levenberg–Marquardt scheme proposed in [3].

On the other hand, one may consider the Levenberg–Marquardt method (1.3) with $\{\alpha_k\}$ given in an a priori way. In the recent book [10] Kaltenbacher et al. considered the choice

$$\alpha_k = \alpha_0 r^k, \quad k = 0, 1, \dots \tag{1.7}$$

for some $\alpha_0 > 0$ and $0 < r < 1$. By assuming the source condition (1.5) with $0 < \nu \le 1/2$, they terminated the iteration by the a priori stopping rule

$$\eta(1+k_\delta)^{-(1+\varepsilon)}\alpha_{k_\delta}^{\nu+1/2} \le \delta < \eta(1+k)^{-(1+\varepsilon)}\alpha_k^{\nu+1/2}, \quad 0 \le k < k_\delta \tag{1.8}$$

with $\eta > 0$ and $\varepsilon > 0$. Under the condition (1.6), they proved in [10, Theorem 4.7] that $x_{k_\delta}^\delta$ converges to $x^\dagger$ with the rate $O((\delta(1+|\log\delta|)^{1+\varepsilon})^{2\nu/(1+2\nu)})$ which is only almost optimal. Their result does not guarantee the convergence of $x_{k_\delta}^\delta$ to $x^\dagger$ if no source condition is assumed. Moreover, the stopping rule (1.8) depends heavily on the information from (1.5) which is difficult to check in practice. In order to make (1.3) and (1.7) to be a useful method in practical applications, (1.8) should be replaced by an a posteriori stopping rule.

In this paper we will consider the Levenberg–Marquardt method (1.3) with $\{\alpha_k\}$ chosen as in (1.7). We will terminate the iteration by the discrepancy principle (1.4) to produce an approximation $x_{k_\delta}^\delta$. Under the condition (1.6) on F, we will show that $x_{k_\delta}^\delta$ always converges to $x^\dagger$ as $\delta \to 0$, and, if, in addition, $x_0 - x^\dagger$ satisfies the source condition (1.5) with $0 < \nu \le 1/2$, we will establish the order optimal convergence rate. Consequently, we improve the corresponding result in [10] significantly. The method of the present paper is essentially different from those in [3,11,12] in that the sequence $\{\alpha_k\}$ is given in an a priori way. This has the advantage of saving the effort of computing $\{\alpha_k\}$ during computations.

This paper is organized as follows. In Sect. 2 we will state the main results on the method given by (2.1), (2.2) and (2.3). After some preliminary estimates given in Sect. 3, we then devote to the proof of the main results by establishing various estimates in Sects. 4–8. Finally, in Sect. 9 we report some numerical experiments to test our theoretical results.

2 Main results

The method we will consider is the Levenberg–Marquardt method

$$x_{k+1}^\delta = x_k^\delta - \left(\alpha_k I + F'(x_k^\delta)^* F'(x_k^\delta)\right)^{-1} F'(x_k^\delta)^* \left(F(x_k^\delta) - y^\delta\right), \tag{2.1}$$

where $\{\alpha_k\}$ is given by

$$\alpha_k = \alpha_0 r^k, \quad k = 0, 1, \dots \tag{2.2}$$

for some $\alpha_0 > 0$ and $0 < r < 1$. We will produce a reasonable approximation to $x^\dagger$ by terminating the iteration (2.1) according to the discrepancy principle

$$\|F(x_{k_\delta}^\delta) - y^\delta\| \le \tau\delta < \|F(x_k^\delta) - y^\delta\|, \quad 0 \le k < k_\delta, \tag{2.3}$$

where $\tau > 1$ is a given number.

In order to obtain the approximation property of $x_{k_\delta}^\delta$ to $x^\dagger$ as $\delta \to 0$, we need to impose certain conditions on F. We assume that

$$B_\rho(x^\dagger) := \left\{x \in X : \|x - x^\dagger\| < \rho\right\} \subset D(F) \tag{2.4}$$

for some $\rho > 0$ and there is a constant $K_0 \ge 0$ such that for any pair $x, z \in B_\rho(x^\dagger)$ there is a bounded linear operator $Q(x, z) : Y \to Y$ such that

$$F'(x) = Q(x, z)F'(z) \quad \text{and} \quad \|I - Q(x, z)\| \le K_0\|x - z\|. \tag{2.5}$$

For simplicity of the presentation, we assume that $F'(x^\dagger)$ is properly scaled so that

$$\|F'(x^\dagger)\| \le \alpha_0^{1/2}. \tag{2.6}$$

This scaling condition can always be fulfilled by multiplying the equation (1.1) by a sufficiently small constant.

To formulate the main results on $x_{k_\delta}^\delta$ precisely, we introduce the two constants

$$c_0 := \frac{1}{\sqrt{r}} \quad \text{and} \quad c_1 := \frac{1}{1 - \sqrt{r}}.$$

For the sequence $\{\alpha_k\}$ given by (2.2), it is easy to verify that

$$\left(\sum_{j=0}^{k} \alpha_j^{-1}\right)^{-1/2} \le c_0\alpha_{k+1}^{1/2} \tag{2.7}$$

and

$$\sum_{m=0}^{k} \alpha_m^{-1/2} \left(\sum_{j=m}^{k} \alpha_j^{-1}\right)^{-1/2} \le c_1 \tag{2.8}$$

for all integers $k \ge 0$.

Now we are ready to state the main result of this paper.[2]

[2] The choice of $\{\alpha_k\}$ in (2.2) is the most important case in applications. The main result, however, is still true for any sequence $\{\alpha_k\}$ of positive numbers satisfying $\alpha_{k+1}^{-1} \le c\sigma_k$ for all k, where $\sigma_k = \sum_{j=0}^{k} \alpha_j^{-1}$ and c is a positive constant independent of k. This can be established by modifying slightly the arguments in the present paper.

Theorem 1 *Let F satisfy (2.4)–(2.6) for some $\rho > 0$, let $\{\alpha_k\}$ be given by (2.2), let $\tau > 1$ be a given number and let $\gamma_0 > c_0/(\tau - 1)$. There is a positive constant η_0 depending only on τ and r such that if $(2+c_1\gamma_0)\|x_0 - x^\dagger\| < \rho$ and $K_0\|x_0 - x^\dagger\| \le \eta_0$ then the method (2.1)–(2.3) is well-defined, the integer k_δ determined by the discrepancy principle (2.3) satisfies*

$$k_\delta = O(1 + |\log \delta|), \tag{2.9}$$

and if $x_0 - x^\dagger \in \mathcal{N}(F'(x^\dagger))^\perp$ then

$$\lim_{\delta\to 0} x_{k_\delta}^\delta = x^\dagger. \tag{2.10}$$

If, in addition, $x_0 - x^\dagger = (F'(x^\dagger)^* F'(x^\dagger))^\nu \omega$ for some $0 < \nu \le 1/2$ and $\omega \in \mathcal{N}(F'(x^\dagger))^\perp \subset X$, and if $K_0\|\omega\| \le \eta_1$ for some $\eta_1 > 0$ depending only on α_0 and r, then

$$\|x_{k_\delta}^\delta - x^\dagger\| \le C_\nu \|\omega\|^{1/(1+2\nu)} \delta^{2\nu/(1+2\nu)}, \tag{2.11}$$

where C_ν is a constant depending only on τ, r and ν.

Theorem 1 shows the order optimality of the method defined by (2.1)–(2.3) for each $0 < \nu \le 1/2$. The inequality (2.9) indicates that the method has fast convergence property. We emphasize that our main result is established for any fixed $\tau > 1$ for the number τ in the discrepancy principle (2.3) provided that $\|x_0 - x^\dagger\|$ is suitably small, which is important in practical applications.

When F is a linear operator, the method defined by (2.1)–(2.3) has been considered in [5] in which it is called the nonstationary iterated Tikhonov regularization. The detailed convergence analysis has been carried out there and the order optimality has even been established for all $\nu > 0$. Due to the nonlinearity of F, in Theorem 1 we are only able to show the order optimality for $0 < \nu \le 1/2$.

The proof of Theorem 1 will occupy the remaining part of this paper. Throughout the paper we will use $\{x_k\}$ to denote the iterative solutions defined by (2.1) corresponding to the noise free case, i.e.

$$x_{k+1} = x_k - \left(\alpha_k I + F'(x_k)^* F'(x_k)\right)^{-1} F'(x_k)^* \left(F(x_k) - y\right). \tag{2.12}$$

We will use the notations

$$\begin{aligned} e_k &:= x_k - x^\dagger, & \mathcal{A} &:= F'(x^\dagger)^* F'(x^\dagger), & \mathcal{A}_k &:= F'(x_k)^* F'(x_k),\\ e_k^\delta &:= x_k^\delta - x^\dagger, & \mathcal{B} &:= F'(x^\dagger) F'(x^\dagger)^*, & \mathcal{A}_k^\delta &:= F'(x_k^\delta)^* F'(x_k^\delta). \end{aligned}$$

For ease of exposition, we will use C to denote a generic constant depending only on τ and r, we will also use the convention $\Phi \lesssim \Psi$ to mean that $\Phi \le C\Psi$ for some generic constant C. Moreover, when we say $K_0\|e_0\|$ is suitably small we will mean that $K_0\|e_0\| \le \eta$ for some small positive constant η depending only on τ and r.

Since the proof is somewhat involved, we give a brief outline of the arguments.

- We first show the method given by (2.1)–(2.3) is well-defined. This is done in Sect. 4.
- We establish in Sect. 5 the stability estimate $\|x_k^\delta - x_k\| \lesssim \delta/\sqrt{\alpha_k}$ together with other crucial estimates. This enables us to write

$$\|e_{k_\delta}^\delta\| \lesssim \|e_{k_\delta}\| + \frac{\delta}{\sqrt{\alpha_{k_\delta}}}.$$

- We then establish in Sect. 6 the various estimates on the noise-free iterates $\{x_k\}$, in particular we show that $\|e_k\| \to 0$ as $k \to \infty$. This, together with the above estimates, is enough to prove $x_{k_\delta}^\delta \to x^\dagger$ as $\delta \to 0$.
- In order to obtain the optimal convergence rates, in Sect. 7 we connect $\|e_{k_\delta}\|$ with $\|e_k\|$ for $k \geq k_\delta$ by establishing the inequality

$$\|x_{k_\delta} - x_k\| \lesssim \frac{1}{\sqrt{\alpha_k}}\|F(x_{k_\delta}) - y\|.$$

  Consequently, with the help of the definition of k_δ, we derive

$$\|e_{k_\delta}^\delta\| \lesssim \|e_k\| + \frac{\delta}{\sqrt{\alpha_k}}, \quad k \geq k_\delta.$$

  When e_0 satisfies (1.5) with $0 < \nu \leq 1/2$, the optimal rate of convergence can be obtained by choosing a $k \geq k_\delta$ carefully.

3 Preliminary estimates

In this section we will give some preliminary estimates which will be frequently used during the proof of Theorem 1.

It is well-known that the condition (2.5) on F implies for $x, z \in B_\rho(x^\dagger)$ that (see [6])

$$\|F(x) - F(z) - F'(z)(x-z)\| \leq \frac{1}{2}K_0\|x-z\|\|F'(z)(x-z)\| \tag{3.1}$$

and

$$\|F(x) - F(z) - F'(x)(x-z)\| \leq \frac{3}{2}K_0\|x-z\|\|F'(z)(x-z)\|. \tag{3.2}$$

The following result gives some additional consequences of (2.5).

Lemma 1 *Let F satisfy (2.5). For any $x, z \in B_\rho(x^\dagger)$ let $\mathcal{A}_x := F'(x)^*F'(x)$ and $\mathcal{A}_z := F'(z)^*F'(z)$. Then, for any $\alpha > 0$, there hold*

$$(\alpha I + \mathcal{A}_x)^{-1} - (\alpha I + \mathcal{A}_z)^{-1} = (\alpha I + \mathcal{A}_z)^{-1}F'(z)^*R_\alpha(x,z) \tag{3.3}$$

and

$$(\alpha I + \mathcal{A}_x)^{-1} F'(x)^* - (\alpha I + \mathcal{A}_z)^{-1} F'(z)^* = (\alpha I + \mathcal{A}_z)^{-1} F'(z)^* S_\alpha(x, z), \tag{3.4}$$

where $R_\alpha(x, z) : X \to Y$ *and* $S_\alpha(x, z) : Y \to Y$ *are two bounded linear operators satisfying*

$$\|R_\alpha(x, z)\| \leq K_0 \|x - z\| \alpha^{-1/2} \quad and \quad \|S_\alpha(x, z)\| \leq 3K_0 \|x - z\|. \tag{3.5}$$

Proof With the help of (2.5), we have

$$\begin{aligned}
&(\alpha I + \mathcal{A}_x)^{-1} - (\alpha I + \mathcal{A}_z)^{-1} \\
&\quad = (\alpha I + \mathcal{A}_z)^{-1} \left[F'(z)^* \left(F'(z) - F'(x)\right) + \left(F'(z)^* - F'(x)^*\right) F'(x)\right] (\alpha I + \mathcal{A}_x)^{-1} \\
&\quad = (\alpha I + \mathcal{A}_z)^{-1} F'(z)^* \left[(Q(z, x) - I) + \left(I - Q(x, z)^*\right)\right] F'(x) (\alpha I + \mathcal{A}_x)^{-1}
\end{aligned}$$

and

$$\begin{aligned}
&(\alpha I + \mathcal{A}_x)^{-1} F'(x)^* - (\alpha I + \mathcal{A}_z)^{-1} F'(z)^* \\
&\quad = (\alpha I + \mathcal{A}_z)^{-1} \left(F'(x)^* - F'(z)^*\right) + \left[(\alpha I + \mathcal{A}_x)^{-1} - (\alpha I + \mathcal{A}_z)^{-1}\right] F'(x)^* \\
&\quad = (\alpha I + \mathcal{A}_z)^{-1} F'(z)^* \left(Q(x, z)^* - I\right) \\
&\qquad + (\alpha I + \mathcal{A}_z)^{-1} F'(z)^* \left[(Q(z, x) - I) + \left(I - Q(x, z)^*\right)\right] \mathcal{B}_x (\alpha I + \mathcal{B}_x)^{-1},
\end{aligned}$$

where $\mathcal{B}_x := F'(x) F'(x)^*$. We thus obtain (3.3) and (3.4) with

$$\begin{aligned}
R_\alpha(x, z) &:= \left[(Q(z, x) - I) + \left(I - Q(x, z)^*\right)\right] F'(x) (\alpha I + \mathcal{A}_x)^{-1}, \\
S_\alpha(x, z) &:= \left(Q(x, z)^* - I\right) + \left[(Q(z, x) - I) + \left(I - Q(x, z)^*\right)\right] \mathcal{B}_x (\alpha I + \mathcal{B}_x)^{-1}
\end{aligned}$$

which clearly verify (3.5). □

Lemma 2 *Let* $\{\alpha_k\}$ *be a sequence of positive numbers. Then for any bounded linear operator* $A : X \to Y$ *there holds*

$$\left\| \prod_{j=m}^{k} \alpha_j \left(\alpha_j I + A^* A\right)^{-1} (A^* A)^\nu \right\| \leq \left(\sum_{j=m}^{k} \alpha_j^{-1} \right)^{-\nu} \tag{3.6}$$

for any $0 \leq \nu \leq 1$ *and any integers* $0 \leq m \leq k < \infty$.

Proof It is easy to see that

$$\prod_{j=m}^{k} (\alpha_j + \lambda) \geq \lambda \left(\sum_{j=m}^{k} \alpha_j^{-1} \prod_{i=m}^{k} \alpha_i \right)$$

for all $\lambda \geq 0$ since the right hand side is the first order term of the polynomial in λ on the left hand side. This inequality implies immediately that

$$\lambda \prod_{j=m}^{k} \alpha_j (\alpha_j + \lambda)^{-1} \leq \left(\sum_{j=m}^{k} \alpha_j^{-1} \right)^{-1}.$$

Thus for $0 \leq \nu \leq 1$ and $\lambda \geq 0$ we have

$$\lambda^{\nu} \prod_{j=m}^{k} \alpha_j (\alpha_j + \lambda)^{-1} \leq \left(\lambda \prod_{j=m}^{k} \alpha_j (\alpha_j + \lambda)^{-1} \right)^{\nu} \leq \left(\sum_{j=m}^{k} \alpha_j^{-1} \right)^{-\nu}.$$

Now the desired inequality (3.6) follows easily from the spectral theory of self-adjoint operators. □

Remark 1 More general inequalities than (3.6) have been derived in [5], see also [2], when $\{\alpha_k\}$ satisfies certain conditions. We will not use such general inequalities since (3.6) is already enough for our purpose.

During the proof of Theorem 1, we need the estimates on various sums formed by the sequence $\{\alpha_k\}$ satisfying (2.2). All these estimates can be verified easily. In the following result we give two such estimates which will be used in Sect. 6. These two estimates are not sharp, but good enough for our purpose.

Lemma 3 *Let $\{\alpha_k\}$ be given by* (2.2). *For any $0 \leq \nu \leq 1/2$ there hold*

$$\sum_{m=0}^{k} \alpha_m^{2\nu-1/2} \left(\sum_{j=m}^{k} \alpha_j^{-1} \right)^{-1/2} \leq C_0 \alpha_{k+1}^{\nu} \tag{3.7}$$

and

$$\sum_{m=0}^{k} \alpha_m^{2\nu-1/2} \left(\sum_{j=m}^{k} \alpha_j \right)^{-1} \leq C_1 \alpha_{k+1}^{\nu+1/2}, \tag{3.8}$$

where C_0 and C_1 are positive constants depending only on r and α_0.

Proof The argument is elementary, we will only indicate the proof of (3.7). Let $J_k(\nu)$ denote the left hand side. Since $\alpha_j = \alpha_0 r^j$ and $0 < r < 1$, we have

$$\sum_{j=m}^{k} \alpha_j^{-1} = \alpha_{k+1}^{-1} \frac{1 - r^{k+1-m}}{r^{-1} - 1} \geq \alpha_{k+1}^{-1} \frac{1 - r}{r^{-1} - 1} = r \alpha_{k+1}^{-1}.$$

This implies

$$J_k(\nu) \le \frac{1}{\sqrt{r}} \alpha_{k+1}^{1/2} \sum_{m=0}^{k} \alpha_m^{2\nu-1/2}.$$

When $2/5 \le \nu \le 1/2$, we have $2\nu - 1/2 \ge 3/10$. Thus

$$\sum_{m=0}^{k} \alpha_m^{2\nu-1/2} \le \frac{\alpha_0^{2\nu-1/2}}{1-r^{2\nu-1/2}} \le \frac{\alpha_0^{2\nu-1/2}}{1-r^{3/10}}.$$

Therefore, with $C_2 := 1/(\sqrt{r}(1-r^{3/10}))$, we obtain

$$J_k(\nu) \le C_2 \alpha_0^{2\nu-1/2} \alpha_{k+1}^{1/2} = C_2 \alpha_0^{\nu} r^{1/2-\nu} \alpha_{k+1}^{\nu} \le C_2 \alpha_0^{\nu} \alpha_{k+1}^{\nu}.$$

When $1/5 \le \nu \le 2/5$, we have $2\nu - 1/2 \ge -1/10$. Thus

$$\sum_{m=0}^{k} \alpha_m^{2\nu-1/2} \le \alpha_0^{2\nu-1/2} \sum_{m=0}^{k} r^{-m/10} \le \alpha_0^{2\nu-1/2} \frac{r^{-(k+1)/10}}{r^{-1/10}-1}.$$

Consequently, with $C_3 := 1/(\sqrt{r}(r^{-1/10}-1))$,

$$J_k(\nu) \le C_3 \alpha_{k+1}^{1/2} \alpha_0^{2\nu-1/2} r^{-(k+1)/10} = C_3 \alpha_0^{\nu} r^{(2/5-\nu)(k+1)} \alpha_{k+1}^{\nu} \le C_3 \alpha_0^{\nu} \alpha_{k+1}^{\nu}.$$

When $0 \le \nu \le 1/5$, we have $2\nu - 1/2 \le -1/10$. Thus

$$\sum_{m=0}^{k} \alpha_m^{2\nu-1/2} \le \alpha_0^{2\nu-1/2} \frac{r^{(2\nu-1/2)(k+1)}}{r^{2\nu-1/2}-1} \le \frac{1}{r^{-1/10}-1} \alpha_0^{2\nu-1/2} r^{(2\nu-1/2)(k+1)}$$

Therefore

$$J_k(\nu) \le C_3 \alpha_0^{2\nu-1/2} r^{(2\nu-1/2)(k+1)} \alpha_{k+1}^{1/2} = C_3 \alpha_0^{\nu} r^{\nu(k+1)} \alpha_{k+1}^{\nu} \le C_3 \alpha_0^{\nu} \alpha_{k+1}^{\nu}.$$

Note that $\alpha_0^{\nu} \le \max\{1, \alpha_0^{1/2}\}$ for $0 \le \nu \le 1/2$, we complete the proof of (3.7). □

4 Justification of the method

In this section we will show, under the conditions in Theorem 1, that the method given by (2.1)–(2.3) is well-defined. We will achieve this by proving that $x_k^{\delta} \in B_{\rho}(x^{\dagger})$ for $0 \le k \le k_{\delta}$ and k_{δ} is finite. To this end, we introduce the integer $\tilde{k}_{\delta}$ defined by

$$\alpha_{\tilde{k}_\delta} \le \left(\frac{\delta}{\gamma_0\|e_0\|}\right)^2 < \alpha_k, \quad 0 \le k < \tilde{k}_\delta, \tag{4.1}$$

where, as in Theorem 1, $\gamma_0 > c_0/(\tau-1)$ is a fixed number.

Lemma 4 *Let F satisfy (2.4)–(2.6) for some $\rho > 0$, let $\{\alpha_k\}$ be given by (2.2), let $\tau > 1$ be a given number, and let $(2 + c_1\gamma_0)\|e_0\| < \rho$. If $K_0\|e_0\|$ is suitably small, then*

$$x_k^\delta \in B_\rho(x^\dagger), \quad \|e_k^\delta\| \lesssim \|e_0\| \quad and \quad \|F'(x^\dagger)e_k^\delta\| \lesssim \alpha_k^{1/2}\|e_0\|$$

for all integers $0 \le k \le \tilde{k}_\delta$.

Proof It suffices to show for $0 \le k \le \tilde{k}_\delta$ that

$$\|e_k^\delta\| \le (2 + c_1\gamma_0)\,\|e_0\| \quad \text{and} \quad \|F'(x^\dagger)e_k^\delta\| \le (2c_0 + c_2\gamma_0)\,\|e_0\|\alpha_k^{1/2}, \tag{4.2}$$

where c_0 and c_1 are given in Sect. 2 such that (2.7) and (2.8) hold, and $c_2 > 0$ is a constant such that

$$\sum_{m=0}^{k} \alpha_m^{-1/2} \left(\sum_{j=m}^{k} \alpha_j^{-1}\right)^{-1} \le c_2\alpha_{k+1}^{1/2} \tag{4.3}$$

for all $k \ge 0$. By using (2.2) we can take $c_2 = 1/(\sqrt{r}(1-\sqrt{r}))$.

From (2.6) it is easy to see that (4.2) is trivial for $k = 0$. Now for any fixed $0 \le l < \tilde{k}_\delta$ we assume that (4.2) is true for every $0 \le k \le l$, and we will show it is also true for $k = l + 1$.

We set for each $0 \le k \le l$

$$u_k^\delta := F(x_k^\delta) - y - F'(x_k^\delta)e_k^\delta.$$

It follows from (2.1) that

$$\begin{aligned} e_{k+1}^\delta &= \alpha_k\left(\alpha_k I + \mathcal{A}_k^\delta\right)^{-1} e_k^\delta - \left(\alpha_k I + \mathcal{A}_k^\delta\right)^{-1} F'(x_k^\delta)^*\left(y - y^\delta + u_k^\delta\right) \\ &= \alpha_k\left(\alpha_k I + \mathcal{A}\right)^{-1} e_k^\delta + \alpha_k\left[\left(\alpha_k I + \mathcal{A}_k^\delta\right)^{-1} - \left(\alpha_k I + \mathcal{A}\right)^{-1}\right] e_k^\delta \\ &\quad - \left(\alpha_k I + \mathcal{A}\right)^{-1} F'(x^\dagger)^*\left(y - y^\delta + u_k^\delta\right) \\ &\quad - \left[\left(\alpha_k I + \mathcal{A}_k^\delta\right)^{-1} F'(x_k^\delta)^* - \left(\alpha_k I + \mathcal{A}\right)^{-1} F'(x^\dagger)^*\right]\left(y - y^\delta + u_k^\delta\right). \end{aligned}$$

With the help of Lemma 1 we then have

$$e_{k+1}^\delta = \alpha_k\left(\alpha_k I + \mathcal{A}\right)^{-1} e_k^\delta + \left(\alpha_k I + \mathcal{A}\right)^{-1} F'(x^\dagger)^* w_k^\delta, \tag{4.4}$$

where

$$w_k^\delta := \alpha_k R_{\alpha_k}(x_k^\delta, x^\dagger) e_k^\delta + (y^\delta - y) - u_k^\delta - S_{\alpha_k}(x_k^\delta, x^\dagger)\left(y - y^\delta + u_k^\delta\right).$$

By telescoping (4.4) we obtain

$$e_{l+1}^\delta = \prod_{j=0}^{l} \alpha_j \left(\alpha_j I + \mathcal{A}\right)^{-1} e_0 + \sum_{m=0}^{l} \alpha_m^{-1} \prod_{j=m}^{l} \alpha_j \left(\alpha_j I + \mathcal{A}\right)^{-1} F'(x^\dagger)^* w_m^\delta. \tag{4.5}$$

We multiply (4.5) by $F'(x^\dagger)$ and obtain

$$F'(x^\dagger) e_{l+1}^\delta = \prod_{j=0}^{l} \alpha_j \left(\alpha_j I + \mathcal{B}\right)^{-1} F'(x^\dagger) e_0 + \sum_{m=0}^{l} \alpha_m^{-1} \prod_{j=m}^{l} \alpha_j \left(\alpha_j I + \mathcal{B}\right)^{-1} \mathcal{B} w_m^\delta. \tag{4.6}$$

By applying (3.2) to estimate u_m^δ, and then using (3.5) and (1.2) we have

$$\|w_m^\delta\| \le K_0 \|e_m^\delta\|^2 \alpha_m^{1/2} + \left(1 + 3K_0\|e_m^\delta\|\right)\left(\delta + \frac{3}{2} K_0 \|e_m^\delta\| \|F'(x^\dagger) e_m^\delta\|\right).$$

From (4.1) we note that $\delta < \gamma_0 \|e_0\| \alpha_m^{1/2}$ for $0 \le m \le l$, therefore we may use the induction hypothesis to conclude

$$\|w_m^\delta\| \le (\gamma_0 + C K_0 \|e_0\|) \|e_0\| \alpha_m^{1/2}. \tag{4.7}$$

By using (3.6), (4.7), (2.7), (2.8) and (4.3), we can obtain from (4.5) and (4.6) that

$$\begin{aligned}
\|e_{l+1}^\delta\| &\le \|e_0\| + \sum_{m=0}^{l} \alpha_m^{-1} \left(\sum_{j=m}^{l} \alpha_j^{-1}\right)^{-1/2} \|w_m^\delta\| \\
&\le \|e_0\| + (\gamma_0 + C K_0 \|e_0\|) \|e_0\| \sum_{m=0}^{l} \alpha_m^{-1/2} \left(\sum_{j=m}^{l} \alpha_j^{-1}\right)^{-1/2} \\
&\le (1 + c_1 \gamma_0 + C K_0 \|e_0\|) \|e_0\| \\
&\le (2 + c_1 \gamma_0) \|e_0\|
\end{aligned}$$

and

$$\begin{aligned}
\|F'(x^\dagger)e_{l+1}^\delta\| &\le \left(\sum_{j=0}^{l}\alpha_j^{-1}\right)^{-1/2}\|e_0\| + \sum_{m=0}^{l}\alpha_m^{-1}\left(\sum_{j=m}^{l}\alpha_j^{-1}\right)^{-1}\|w_m^\delta\| \\
&\le c_0\alpha_{l+1}^{1/2}\|e_0\| + (\gamma_0 + CK_0\|e_0\|)\|e_0\|\sum_{m=0}^{l}\alpha_m^{-1/2}\left(\sum_{j=m}^{l}\alpha_j^{-1}\right)^{-1} \\
&\le (c_0 + c_2\gamma_0 + CK_0\|e_0\|)\|e_0\|\alpha_{l+1}^{1/2} \\
&\le (2c_0 + c_2\gamma_0)\|e_0\|\alpha_{l+1}^{1/2}
\end{aligned}$$

if $K_0\|e_0\|$ is suitably small. The proof is therefore complete. □

By applying the same argument to the noise-free iterative solutions $\{x_k\}$ defined by (2.12) we can obtain

Lemma 5 *Let F satisfy (2.4)–(2.6) for some $\rho > 0$, let $\{\alpha_k\}$ be given by (2.2) and let $2\|e_0\| < \rho$. If $K_0\|e_0\|$ is suitably small then*

$$x_k \in B_\rho(x^\dagger), \quad \|e_k\| \lesssim \|e_0\| \quad \text{and} \quad \|F'(x^\dagger)e_k\| \lesssim \alpha_k^{1/2}\|e_0\| \tag{4.8}$$

for all integers $k \ge 0$.

Finally we conclude this section by showing that the integer k_δ defined by the discrepancy principle (2.3) must satisfy $k_\delta \le \tilde{k}_\delta$. This together with Lemma 4 shows that the method given by (2.1)–(2.3) is well-defined.

Lemma 6 *Under the conditions in Lemma 4, if $K_0\|e_0\|$ is suitably small, then the integer k_δ determined by the discrepancy principle (2.3) satisfies $k_\delta \le \tilde{k}_\delta$. Consequently $k_\delta = O(1 + |\log\delta|)$.*

Proof We first show $k_\delta \le \tilde{k}_\delta$. If $\tilde{k}_\delta = 0$, then $\alpha_0 \le (\delta/(\gamma_0\|e_0\|))^2$. Thus, by using (1.2), (3.1) and (2.6) we have

$$\begin{aligned}
\|F(x_0) - y^\delta\| &\le \delta + \|F(x_0) - y - F'(x^\dagger)e_0\| + \|F'(x^\dagger)e_0\| \\
&\le \delta + \left(1 + \frac{1}{2}K_0\|e_0\|\right)\|F'(x^\dagger)e_0\| \\
&\le \delta + \left(1 + \frac{1}{2}K_0\|e_0\|\right)\alpha_0^{1/2}\|e_0\| \\
&\le \delta + \left(1 + \frac{1}{2}K_0\|e_0\|\right)\gamma_0^{-1}\delta.
\end{aligned}$$

Since $\gamma_0 > c_0/(\tau - 1)$ and $c_0 > 1$, it is easy to see that if $K_0\|e_0\|$ is suitably small then $\|F(x_0) - y^\delta\| \le \tau\delta$. Thus $k_\delta = 0$.

In the following we will assume $\tilde{k}_\delta > 0$. We will use (4.6). By Lemma 4 we note that for $0 \le k \le \tilde{k}_\delta$ there holds $w_k^\delta = y^\delta - y + \tilde{w}_k^\delta$ with

$$\|\tilde{w}_k^\delta\| \lesssim K_0\|e_0\|^2\alpha_k^{1/2} + K_0\|e_0\|\delta. \tag{4.9}$$

Therefore, it follows from (4.6) that

$$\begin{aligned}
&F'(x^\dagger)e_k^\delta - y^\delta + y \\
&= \prod_{j=0}^{k-1}\alpha_j\left(\alpha_j I + \mathcal{B}\right)^{-1} F'(x^\dagger)e_0 + \sum_{m=0}^{k-1}\alpha_m^{-1}\prod_{j=m}^{k-1}\alpha_j\left(\alpha_j I + \mathcal{B}\right)^{-1}\mathcal{B}\tilde{w}_m^\delta \\
&\quad + \left[I - \sum_{m=0}^{k-1}\alpha_m^{-1}\prod_{j=m}^{k-1}\alpha_j\left(\alpha_j I + \mathcal{B}\right)^{-1}\mathcal{B}\right](y - y^\delta).
\end{aligned}$$

Note that

$$I - \sum_{m=0}^{k-1}\alpha_m^{-1}\prod_{j=m}^{k-1}\alpha_j\left(\alpha_j I + \mathcal{B}\right)^{-1}\mathcal{B} = \prod_{j=0}^{k-1}\alpha_j\left(\alpha_j I + \mathcal{B}\right)^{-1}. \tag{4.10}$$

We may use (3.6), (2.7), (2.2) and (4.9) to obtain

$$\begin{aligned}
\|F'(x^\dagger)e_k^\delta - y^\delta + y\| &\le c_0\alpha_k^{1/2}\|e_0\| + \delta + CK_0\|e_0\|\delta\sum_{m=0}^{k-1}\alpha_m^{-1}\left(\sum_{j=m}^{k-1}\alpha_j^{-1}\right)^{-1} \\
&\quad + CK_0\|e_0\|^2\sum_{m=0}^{k-1}\alpha_m^{-1/2}\left(\sum_{j=m}^{k-1}\alpha_j^{-1}\right)^{-1} \\
&\le c_0\alpha_k^{1/2}\|e_0\| + \delta + C\left(K_0\|e_0\|\delta + K_0\|e_0\|^2\alpha_k^{1/2}\right).
\end{aligned}$$

By taking $k = \tilde{k}_\delta$ in the above inequality, noting from (4.1) that $\alpha_{\tilde{k}_\delta} \le (\delta/(\gamma_0\|e_0\|))^2$ with $\gamma_0 > c_0/(\tau - 1)$, and using Lemma 4, we have

$$\begin{aligned}
\|F(x_{\tilde{k}_\delta}^\delta) - y^\delta\| &\le \|F(x_{\tilde{k}_\delta}^\delta) - y - F'(x^\dagger)e_{\tilde{k}_\delta}^\delta\| + \|F'(x^\dagger)e_{\tilde{k}_\delta}^\delta - y^\delta + y\| \\
&\le \frac{1}{2}K_0\|e_{\tilde{k}_\delta}^\delta\|\|F'(x^\dagger)e_{\tilde{k}_\delta}^\delta\| + c_0\alpha_{\tilde{k}_\delta}^{1/2}\|e_0\| + \delta \\
&\quad + C\left(K_0\|e_0\|^2\alpha_{\tilde{k}_\delta}^{1/2} + K_0\|e_0\|\delta\right) \\
&\le \left(1 + c_0/\gamma_0 + CK_0\|e_0\|\right)\delta \\
&\le \tau\delta
\end{aligned}$$

if $K_0\|e_0\|$ is suitably small. According to the definition of k_δ, this implies that $k_\delta \le \tilde{k}_\delta$.

From the definition (4.1) of $\tilde{k}_\delta$ and (2.2) it follows that

$$\alpha_0 r^{\tilde{k}_\delta - 1} = \alpha_{\tilde{k}_\delta - 1} \ge \left(\frac{\delta}{\gamma_0 \|e_0\|}\right)^2.$$

Note that $0 < r < 1$, we then obtain

$$\tilde{k}_\delta - 1 \le \frac{1}{\log r} \log\left(\frac{\delta^2}{\alpha_0 \gamma_0^2 \|e_0\|^2}\right).$$

Therefore $k_\delta \le \tilde{k}_\delta = O(1 + |\log \delta|)$. □

5 Stability estimates

In this section we will derive the stability estimate on $\|x_k^\delta - x_k\|$ for $0 \le k \le \tilde{k}_\delta$. We will set

$$u_k := F(x_k) - y - F'(x_k)e_k \quad \text{and} \quad u_k^\delta := F(x_k^\delta) - y - F'(x_k^\delta)e_k^\delta.$$

It then follows from (2.1) and (2.12) that

$$\begin{aligned} x_{k+1}^\delta - x_{k+1} = x_k^\delta - x_k &+ \left(\alpha_k I + \mathcal{A}_k^\delta\right)^{-1} F'(x_k^\delta)^*(y^\delta - y) \\ &- \left(\alpha_k I + \mathcal{A}_k^\delta\right)^{-1} F'(x_k^\delta)^* u_k^\delta + (\alpha_k I + \mathcal{A}_k)^{-1} F'(x_k)^* u_k \\ &+ \left[(\alpha_k I + \mathcal{A}_k)^{-1} \mathcal{A}_k e_k - \left(\alpha_k I + \mathcal{A}_k^\delta\right)^{-1} \mathcal{A}_k^\delta e_k^\delta\right]. \end{aligned}$$

Note that

$$\begin{aligned} &(\alpha_k I + \mathcal{A}_k)^{-1} \mathcal{A}_k e_k - \left(\alpha_k I + \mathcal{A}_k^\delta\right)^{-1} \mathcal{A}_k^\delta e_k^\delta \\ &\quad = (\alpha_k I + \mathcal{A}_k)^{-1} \mathcal{A}_k \left(x_k - x_k^\delta\right) + \alpha_k \left[\left(\alpha_k I + \mathcal{A}_k^\delta\right)^{-1} - (\alpha_k I + \mathcal{A}_k)^{-1}\right] e_k^\delta \\ &\quad = (\alpha_k I + \mathcal{A})^{-1} \mathcal{A} \left(x_k - x_k^\delta\right) + \alpha_k \left[(\alpha_k I + \mathcal{A}_k)^{-1} - (\alpha_k I + \mathcal{A})^{-1}\right] \left(x_k^\delta - x_k\right) \\ &\qquad + \alpha_k \left[\left(\alpha_k I + \mathcal{A}_k^\delta\right)^{-1} - (\alpha_k I + \mathcal{A}_k)^{-1}\right] e_k^\delta. \end{aligned}$$

We therefore obtain

$$x_{k+1}^\delta - x_{k+1} = \alpha_k (\alpha_k I + \mathcal{A})^{-1} \left(x_k^\delta - x_k\right) + s_k^{(1)} + s_k^{(2)} + s_k^{(3)} + s_k^{(4)} + s_k^{(5)}, \tag{5.1}$$

where

$$
\begin{aligned}
s_k^{(1)} &:= \alpha_k \left[\left(\alpha_k I + \mathcal{A}_k^\delta\right)^{-1} - (\alpha_k I + \mathcal{A}_k)^{-1}\right] e_k^\delta,\\
s_k^{(2)} &:= \alpha_k \left[(\alpha_k I + \mathcal{A}_k)^{-1} - (\alpha_k I + \mathcal{A})^{-1}\right] \left(x_k^\delta - x_k\right),\\
s_k^{(3)} &:= \left[(\alpha_k I + \mathcal{A}_k)^{-1} F'(x_k)^* - \left(\alpha_k I + \mathcal{A}_k^\delta\right)^{-1} F'(x_k^\delta)^*\right] u_k,\\
s_k^{(4)} &:= \left(\alpha_k I + \mathcal{A}_k^\delta\right)^{-1} F'(x_k^\delta)^* \left(u_k - u_k^\delta\right),\\
s_k^{(5)} &:= \left(\alpha_k I + \mathcal{A}_k^\delta\right)^{-1} F'(x_k^\delta)^* \left(y^\delta - y\right).
\end{aligned}
$$

Lemma 7 *Under the conditions in Lemma 4, for all integers* $0 \le k \le \tilde{k}_\delta$ *there hold*

$$
s_k^{(i)} = (\alpha_k I + \mathcal{A})^{-1} F'(x^\dagger)^* w_k^{(i)}, \quad i = 1, \ldots, 5, \tag{5.2}
$$

where

$$
\|w_k^{(i)}\| \lesssim K_0 \|e_0\| \|x_k^\delta - x_k\| \alpha_k^{1/2}, \quad i = 1, \ldots, 4 \tag{5.3}
$$

and

$$
w_k^{(5)} = y^\delta - y + \tilde{w}_k^{(5)}, \quad \|\tilde{w}_k^{(5)}\| \lesssim K_0 \|e_0\| \delta. \tag{5.4}
$$

Proof This result follows essentially from Lemma 1. We first consider $s_k^{(1)}$. With the help of Lemma 1 we have

$$
\begin{aligned}
s_k^{(1)} &= \alpha_k (\alpha_k I + \mathcal{A}_k)^{-1} F'(x_k)^* R_{\alpha_k}(x_k^\delta, x_k) e_k^\delta\\
&= \alpha_k (\alpha_k I + \mathcal{A})^{-1} F'(x^\dagger)^* \left[I + S_{\alpha_k}(x_k, x^\dagger)\right] R_{\alpha_k}(x_k^\delta, x_k) e_k^\delta.
\end{aligned}
$$

Thus $s_k^{(1)}$ has the representation (5.2) with

$$
w_k^{(1)} = \alpha_k \left[I + S_{\alpha_k}(x_k, x^\dagger)\right] R_{\alpha_k}(x_k^\delta, x_k) e_k^\delta.
$$

By using (3.5) and Lemma 4 we can see

$$
\|w_k^{(1)}\| \le (1 + 3K_0\|e_k\|) K_0 \|e_k^\delta\| \|x_k^\delta - x_k\| \alpha_k^{1/2} \lesssim K_0 \|e_0\| \|x_k^\delta - x_k\| \alpha_k^{1/2}.
$$

We thus obtain the conclusion on $s_k^{(1)}$. With an application of Lemma 1 we also see that $s_k^{(2)}$ has the representation (5.2) with $w_k^{(2)} = \alpha_k R_{\alpha_k}(x_k, x^\dagger)(x_k^\delta - x_k)$ which clearly satisfies (5.3). In order to show the conclusion on $s_k^{(3)}$, we use Lemma 1 to write

$$
\begin{aligned}
s_k^{(3)} &= -(\alpha_k I + \mathcal{A}_k)^{-1} F'(x_k)^* S_{\alpha_k}(x_k^\delta, x_k) u_k\\
&= -(\alpha_k I + \mathcal{A})^{-1} F'(x^\dagger)^* \left[I + S_{\alpha_k}(x_k, x^\dagger)\right] S_{\alpha_k}(x_k^\delta, x_k) u_k.
\end{aligned}
$$

Thus $s_k^{(3)}$ is of the form (5.2) with

$$w_k^{(3)} = -\left[I + S_{\alpha_k}(x_k, x^\dagger)\right] S_{\alpha_k}(x_k^\delta, x_k)u_k.$$

It is clear from (3.2) and Lemma 5 that

$$\begin{aligned}
\|w_k^{(3)}\| &\le 3\left(1 + 3K_0\|e_k\|\right) K_0\|x_k^\delta - x_k\|\|u_k\| \\
&\le \frac{9}{2}\left(1 + 3K_0\|e_k\|\right) K_0^2\|x_k^\delta - x_k\|\|e_k\|\|F'(x^\dagger)e_k\| \\
&\lesssim K_0^2\|e_0\|^2\|x_k^\delta - x_k\|\alpha_k^{1/2}.
\end{aligned}$$

For $s_k^{(4)}$ we easily see from Lemma 1 that it has the form (5.2) with

$$w_k^{(4)} = \left[I + S_{\alpha_k}(x_k^\delta, x^\dagger)\right](u_k - u_k^\delta).$$

Note that (2.5), Lemma 4 and Lemma 5 imply

$$\begin{aligned}
\|u_k - u_k^\delta\| &\le \|F(x_k^\delta) - F(x_k) - F'(x_k)(x_k^\delta - x_k)\| + \|(F'(x_k) - F'(x_k^\delta)e_k^\delta\| \\
&\lesssim K_0\|x_k^\delta - x_k\|\|F'(x_k)(x_k^\delta - x_k)\| + K_0\|x_k^\delta - x_k\|\|F'(x_k)e_k^\delta\| \\
&\lesssim K_0\|x_k^\delta - x_k\|\|F'(x^\dagger)(x_k^\delta - x_k)\| + K_0\|x_k^\delta - x_k\|\|F'(x^\dagger)e_k^\delta\| \\
&\lesssim K_0\|e_0\|\|x_k^\delta - x_k\|\alpha_k^{1/2}.
\end{aligned}$$

Therefore

$$\|w_k^{(4)}\| \lesssim K_0\|e_0\|\|x_k^\delta - x_k\|\alpha_k^{1/2}.$$

Finally, applying Lemma 1 to $s_k^{(5)}$ we can conclude that $s_k^{(5)}$ has the representation (5.2) with $w_k^{(5)} = y^\delta - y + \tilde{w}_k^{(5)}$ and $\tilde{w}_k^{(5)} = S_{\alpha_k}(x_k^\delta, x^\dagger)(y^\delta - y)$. It is clear that $\|\tilde{w}_k^{(5)}\| \le 3K_0\|e_k^\delta\|\delta \lesssim K_0\|e_0\|\delta$. □

Proposition 1 *Under the conditions in Lemma 4, for all* $0 \le k \le \tilde{k}_\delta$ *there hold*

$$\|x_k^\delta - x_k\| \lesssim \frac{\delta}{\sqrt{\alpha_k}} \tag{5.5}$$

and

$$\|F(x_k^\delta) - F(x_k) - y^\delta + y\| \le (1 + CK_0\|e_0\|)\delta \tag{5.6}$$

provided that $K_0\|e_0\|$ *is suitably small.*

Proof We first prove (5.5) by showing that

$$\|x_k^\delta - x_k\| \le 2c_3 \frac{\delta}{\sqrt{\alpha_k}} \tag{5.7}$$

for all $0 \le k \le \tilde{k}_\delta$, where $c_3 > 0$ is a constant such that

$$\sum_{m=0}^{k} \alpha_m^{-1} \left(\sum_{j=m}^{k} \alpha_j^{-1} \right)^{-1/2} \le c_3 \alpha_{k+1}^{-1/2} \tag{5.8}$$

for all $k \ge 0$. The existence of such c_3 is guaranteed by (2.2).

It is clear that (5.7) is trivial for $k = 0$. Now for any fixed $0 \le l < \tilde{k}_\delta$ we assume that (5.7) is true for all $0 \le k \le l$ and we will show that it is also true for $k = l + 1$.

By telescoping (5.1) and using Lemma 7 we have

$$x_{l+1}^\delta - x_{l+1} = \sum_{m=0}^{l} \alpha_m^{-1} \prod_{j=m}^{l} \alpha_j \left(\alpha_j I + \mathcal{A}\right)^{-1} F'(x^\dagger)^* \left(\sum_{i=1}^{5} w_m^{(i)} \right). \tag{5.9}$$

With the help of (3.6), (5.3) and (5.4) we obtain

$$\begin{aligned}
\|x_{l+1}^\delta - x_{l+1}\| &\le \sum_{m=0}^{l} \alpha_m^{-1} \left(\sum_{j=m}^{l} \alpha_j^{-1} \right)^{-1/2} \sum_{i=1}^{5} \|w_m^{(i)}\| \\
&\le C K_0 \|e_0\| \sum_{m=0}^{l} \alpha_m^{-1/2} \left(\sum_{j=m}^{l} \alpha_j^{-1} \right)^{-1/2} \|x_m^\delta - x_m\| \\
&\quad + (1 + C K_0 \|e_0\|)\, \delta \sum_{m=0}^{l} \alpha_m^{-1} \left(\sum_{j=m}^{l} \alpha_j^{-1} \right)^{-1/2}.
\end{aligned}$$

By using (5.8) and the induction hypothesis $\|x_m^\delta - x_m\| \le 2c_3 \delta \alpha_m^{-1/2}$ for $0 \le m \le l$, we have

$$\begin{aligned}
\|x_{l+1}^\delta - x_{l+1}\| &\le C K_0 \|e_0\| \delta \sum_{m=0}^{l} \alpha_m^{-1} \left(\sum_{j=m}^{l} \alpha_j^{-1} \right)^{-1/2} \\
&\quad + (c_3 + C K_0 \|e_0\|)\, \delta \alpha_{l+1}^{-1/2} \\
&\le (c_3 + C K_0 \|e_0\|) \delta \alpha_{l+1}^{-1/2} \\
&\le 2c_3 \delta \alpha_{l+1}^{-1/2}
\end{aligned}$$

if $K_0 \|e_0\|$ is suitably small. This completes the proof of (5.7).

Next we prove (5.6). We first claim that for all $0 \le k \le \tilde{k}_\delta$

$$\|F'(x^\dagger)(x_k^\delta - x_k) - y^\delta + y\| \le (1 + CK_0\|e_0\|)\,\delta. \tag{5.10}$$

This is trivial when $k = 0$. So we may assume $0 < k \le \tilde{k}_\delta$. Then it follows from (5.9) and (5.4) that

$$\begin{aligned} F'(x^\dagger)(x_k^\delta - x_k) - y^\delta + y &= \left[I - \sum_{m=0}^{k-1} \alpha_m^{-1} \prod_{j=m}^{k-1} \alpha_j (\alpha_j I + \mathcal{B})^{-1} \mathcal{B} \right] (y - y^\delta) \\ &\quad + \sum_{m=0}^{k-1} \alpha_m^{-1} \prod_{j=m}^{k-1} \alpha_j (\alpha_j I + \mathcal{B})^{-1} \mathcal{B} \sum_{i=1}^{4} w_m^{(i)} \\ &\quad + \sum_{m=0}^{k-1} \alpha_m^{-1} \prod_{j=m}^{k-1} \alpha_j (\alpha_j I + \mathcal{B})^{-1} \mathcal{B} \tilde{w}_m^{(5)}. \end{aligned}$$

Therefore, we may use (4.10), (3.6), the estimates in Lemma 7, (5.5) and (2.2) to conclude that

$$\begin{aligned} \|F'(x^\dagger)(x_k^\delta - x_k) - y^\delta + y\| &\le \delta + \sum_{m=0}^{k-1} \alpha_m^{-1} \left(\sum_{j=m}^{k-1} \alpha_j^{-1} \right)^{-1} \sum_{i=1}^{4} \|w_m^{(i)}\| \\ &\quad + \sum_{m=0}^{k-1} \alpha_m^{-1} \left(\sum_{j=m}^{k-1} \alpha_j^{-1} \right)^{-1} \|\tilde{w}_m^{(5)}\| \\ &\le \delta + CK_0\|e_0\|\delta \sum_{m=0}^{k-1} \alpha_m^{-1} \left(\sum_{j=m}^{k-1} \alpha_j^{-1} \right)^{-1} \\ &\le (1 + CK_0\|e_0\|)\delta. \end{aligned}$$

We thus obtain (5.10). From (5.10) and (2.5) it follows that

$$\|F'(x^\dagger)(x_k^\delta - x_k)\| \lesssim \delta \quad \text{and} \quad \|F'(x_k)(x_k^\delta - x_k)\| \lesssim \delta.$$

Therefore

$$\begin{aligned} \|F(x_k^\delta) - F(x_k) - y^\delta + y\| &\le \|F(x_k^\delta) - F(x_k) - F'(x_k)(x_k^\delta - x_k)\| \\ &\quad + \|(F'(x_k) - F'(x^\dagger))(x_k^\delta - x_k)\| \\ &\quad + \|F'(x^\dagger)(x_k^\delta - x_k) - y^\delta + y\| \end{aligned}$$

$$\leq \frac{1}{2} K_0 \|x_k^\delta - x_k\| \|F'(x_k)(x_k^\delta - x_k)\|$$
$$+ K_0 \|e_k\| \|F'(x^\dagger)(x_k^\delta - x_k)\|$$
$$+ (1 + C K_0 \|e_0\|)\delta$$
$$\leq (1 + C K_0 \|e_0\|)\delta.$$

Thus we complete the proof of (5.6). □

6 Convergence of the noise free iterations

In this section we will prove the following result which in particular gives the convergence of the sequence $\{x_k\}$ defined by (2.12) to $x^\dagger$ if $e_0 \in \mathcal{N}(F'(x^\dagger))^\perp$ and $K_0\|e_0\|$ is suitably small.

Proposition 2 *Assume that the conditions in Lemma 5 hold and $e_0 \in \mathcal{N}(F'(x^\dagger))^\perp$. If $K_0\|e_0\|$ is suitably small, then*

$$\lim_{k\to\infty} \|e_k\| = 0 \quad \textit{and} \quad \lim_{k\to\infty} \frac{\|F'(x^\dagger)e_k\|}{\sqrt{\alpha_k}} = 0. \tag{6.1}$$

The proof of this result is a little involved. We first show that if $e_0 \in \mathcal{R}(F'(x^\dagger)^*)$ then $x_k \to x^\dagger$ as $k \to \infty$. We then perturb the initial guess x_0 to be $\hat{x}_0$ such that $\hat{x}_0 - x^\dagger \in \mathcal{R}(F'(x^\dagger)^*)$ and define $\{\hat{x}_k\}$ by

$$\hat{x}_{k+1} = \hat{x}_k - \left(\alpha_k I + F'(\hat{x}_k)^* F'(\hat{x}_k)\right)^{-1} F'(\hat{x}_k)^* \left(F(\hat{x}_k) - y\right). \tag{6.2}$$

Since $\mathcal{N}(F'(x^\dagger))^\perp = \overline{\mathcal{R}(F'(x^\dagger)^*)}$, such $\hat{x}_0$ can be chosen as close to x_0 as we want. By comparing $\{x_k\}$ with $\{\hat{x}_k\}$ we then get the convergence of $\{x_k\}$.

Lemma 8 *Assume that the conditions in Lemma 5 hold and that $K_0\|e_0\|$ is suitably small. If $e_0 = \mathcal{A}^\nu \omega$ for some $0 < \nu \leq 1/2$ and $\omega \in X$, and if $K_0\|\omega\| \leq \eta_1$ for some $\eta_1 > 0$ depending only on α_0 and r, then*

$$\|e_k\| \leq 2c_0\alpha_k^\nu \|\omega\| \quad \textit{and} \quad \|F'(x^\dagger)e_k\| \leq 2c_0^2 \alpha_k^{\nu+1/2} \|\omega\| \tag{6.3}$$

for all $k \geq 0$.

Proof We will use induction for the proof. From (2.6) it follows easily that (6.3) is true for $k = 0$. Now we assume that (6.3) is true for every $0 \leq k \leq l$, and we will show that it is also true for $k = l + 1$.

We may use (2.12) and follow the similar derivation of (4.5) and (4.6) to obtain

$$e_{l+1} = \prod_{j=0}^{l} \alpha_j \left(\alpha_j I + \mathcal{A}\right)^{-1} e_0 + \sum_{m=0}^{l} \alpha_m^{-1} \prod_{j=m}^{l} \alpha_j \left(\alpha_j I + \mathcal{A}\right)^{-1} F'(x^\dagger)^* w_m \tag{6.4}$$

and

$$F'(x^\dagger)e_{l+1} = \prod_{j=0}^{l} \alpha_j \left(\alpha_j I + \mathcal{B}\right)^{-1} F'(x^\dagger)e_0 + \sum_{m=0}^{l} \alpha_m^{-1} \prod_{j=m}^{l} \alpha_j \left(\alpha_j I + \mathcal{B}\right)^{-1} \mathcal{B} w_m \tag{6.5}$$

where

$$w_m = \alpha_m R_{\alpha_m}(x_m, x^\dagger)e_m - u_m - S_{\alpha_m}(x_m, x^\dagger)u_m.$$

By using (3.1), (3.5) and Lemma 5 it is easy to see

$$\|w_m\| \lesssim K_0 \|e_m\|^2 \alpha_m^{1/2} + K_0 \|e_m\| \|F'(x^\dagger)e_m\|. \tag{6.6}$$

We then use the induction hypothesis to obtain

$$\|w_m\| \lesssim K_0 \|\omega\|^2 \alpha_m^{2\nu+1/2}.$$

Thus, by using $e_0 = \mathcal{A}^\nu \omega$ with $0 < \nu \le 1/2$, (3.6), (2.7) and Lemma 3, we obtain from (6.4) and (6.5) that

$$\begin{aligned}\|e_{l+1}\| &\le \left(\sum_{j=0}^{l} \alpha_j^{-1}\right)^{-\nu} \|\omega\| + C K_0 \|\omega\|^2 \sum_{m=0}^{l} \alpha_m^{2\nu-1/2} \left(\sum_{j=m}^{l} \alpha_j^{-1}\right)^{-1/2} \\ &\le (c_0 + C K_0 \|\omega\|) \alpha_{l+1}^{\nu} \|\omega\|\end{aligned}$$

and

$$\begin{aligned}\|F'(x^\dagger)e_{l+1}\| &\le \left(\sum_{j=0}^{l} \alpha_j^{-1}\right)^{-\nu-1/2} \|\omega\| + C K_0 \|\omega\|^2 \sum_{m=0}^{l} \alpha_m^{2\nu-1/2} \left(\sum_{j=m}^{l} \alpha_j^{-1}\right)^{-1} \\ &\le \left(c_0^2 + C K_0 \|\omega\|\right) \alpha_{l+1}^{\nu+1/2} \|\omega\|,\end{aligned}$$

where $C > 0$ is a constant depending only on α_0 and r. Thus if $K_0\|\omega\| \le \eta_1$ for some $\eta_1 > 0$ depending only on α_0 and r then $\|e_{l+1}\| \le 2c_0\alpha_{l+1}^{\nu}\|\omega\|$ and $\|F'(x^\dagger)e_{l+1}\| \le 2c_0^2\alpha_{l+1}^{\nu+1/2}\|\omega\|$. We thus obtain (6.3). □

Remark 2 Since the results in Lemma 8 require $K_0\|\omega\|$ to be small, they cannot be used to prove Proposition 2 directly. However, by modifying the argument we can show that if $K_0\|e_0\|$ is suitably small and if $e_0 \in \mathcal{R}(F'(x^\dagger)^*)$ then

$$\|e_k\| \to 0 \quad \text{and} \quad \frac{\|F'(x^\dagger)e_k\|}{\sqrt{\alpha_k}} \to 0 \tag{6.7}$$

as $k \to \infty$. To see this, by noting that $\mathcal{R}(F'(x^\dagger)^*) = \mathcal{R}(\mathcal{A}^{1/2}) \subset \mathcal{R}(\mathcal{A}^{1/4})$, we have $e_0 = \mathcal{A}^{1/4}\omega$ for some $\omega \in X$. We will establish (6.7) by showing that if $K_0\|e_0\|$ is suitably small then

$$\|e_k\| \le 2c_0^{1/2}\alpha_k^{1/4}\|\omega\| \quad \text{and} \quad \|F'(x^\dagger)e_k\| \le 2c_0^{3/2}\alpha_k^{3/4}\|\omega\| \tag{6.8}$$

for all $k \ge 0$. In fact, it is obvious for $k = 0$. Suppose that it is true for all $0 \le k \le l$. Then from (6.6) and Lemma 5 we have

$$\|w_m\| \lesssim K_0\|e_0\|\|\omega\|\alpha_m^{3/4}, \quad 0 \le m \le l.$$

Thus, using $e_0 = \mathcal{A}^{1/4}\omega$, it follows from (6.4), (6.5), (3.6), (2.7) and (2.2) that

$$\begin{aligned}\|e_{l+1}\| &\le \left(\sum_{j=0}^{l}\alpha_j^{-1}\right)^{-1/4}\|\omega\| + CK_0\|e_0\|\|\omega\|\sum_{m=0}^{l}\alpha_m^{-1/4}\left(\sum_{j=m}^{l}\alpha_j^{-1}\right)^{-1/2}\\ &\le \left(c_0^{1/2} + CK_0\|e_0\|\right)\alpha_{l+1}^{1/4}\|\omega\| \le 2c_0^{1/2}\alpha_{l+1}^{1/4}\|\omega\|\end{aligned}$$

and

$$\begin{aligned}\|F'(x^\dagger)e_{l+1}\| &\le \left(\sum_{j=0}^{l}\alpha_j^{-1}\right)^{-3/4}\|\omega\| + CK_0\|e_0\|\|\omega\|\sum_{m=0}^{l}\alpha_m^{-1/4}\left(\sum_{j=m}^{l}\alpha_j^{-1}\right)^{-1}\\ &\le \left(c_0^{3/2} + CK_0\|e_0\|\right)\alpha_{l+1}^{3/4}\|\omega\| \le 2c_0^{3/2}\alpha_{l+1}^{3/4}\|\omega\|\end{aligned}$$

if $K_0\|e_0\|$ is suitably small.

In the next result, we take $\hat{x}_0$ to be a perturbation of x_0 and let $\{\hat{x}_k\}$ be defined by (6.2). We set

$$\hat{e}_k := \hat{x}_k - x^\dagger \quad \text{and} \quad \hat{\mathcal{A}}_k := F'(\hat{x}_k)^*F'(\hat{x}_k).$$

According to Lemma 5, it follows that

$$\|\hat{e}_k\| \lesssim \|\hat{e}_0\| \quad \text{and} \quad \|F'(x^\dagger)\hat{e}_k\| \lesssim \alpha_k^{1/2}\|\hat{e}_0\|$$

for all $k \ge 0$ provided $K_0\|\hat{e}_0\|$ is suitably small.

Lemma 9 *Assume that the conditions in Lemma 5 hold. If both $K_0\|e_0\|$ and $K_0\|\hat{e}_0\|$ are suitably small, then*

$$\|x_k - \hat{x}_k\| \le 2\|x_0 - \hat{x}_0\| \tag{6.9}$$

and

$$\|F'(x^\dagger)(x_k - \hat{x}_k)\| \le 2c_0\alpha_k^{1/2}\|x_0 - \hat{x}_0\| \tag{6.10}$$

for all $k \ge 0$.

Proof We will use the essential idea in the proof of Proposition 1. Let

$$u_k := F(x_k) - y - F'(x_k)e_k, \quad \hat{u}_k := F(\hat{x}_k) - y - F'(\hat{x}_k)\hat{e}_k.$$

Similar to the derivation of (5.1), we can obtain from (2.12) and (6.2) that

$$x_{k+1} - \hat{x}_{k+1} = \alpha_k\left(\alpha_k I + \mathcal{A}\right)^{-1}\left(x_k - \hat{x}_k\right) + t_k^{(1)} + t_k^{(2)} + t_k^{(3)} + t_k^{(4)}, \tag{6.11}$$

where

$$\begin{aligned}
t_k^{(1)} &:= \alpha_k\left[(\alpha_k I + \mathcal{A}_k)^{-1} - \left(\alpha_k I + \hat{\mathcal{A}}_k\right)^{-1}\right]\hat{e}_k,\\
t_k^{(2)} &:= \alpha_k\left[(\alpha_k I + \mathcal{A}_k)^{-1} - (\alpha_k I + \mathcal{A})^{-1}\right]\left(x_k - \hat{x}_k\right),\\
t_k^{(3)} &:= \left[\left(\alpha_k I + \hat{\mathcal{A}}_k\right)^{-1} F'(\hat{x}_k)^* - (\alpha_k I + \mathcal{A}_k)^{-1} F'(x_k)^*\right]\hat{u}_k,\\
t_k^{(4)} &:= (\alpha_k I + \mathcal{A}_k)^{-1} F'(x_k)^*\left(\hat{u}_k - u_k\right).
\end{aligned}$$

The analogous argument in Lemma 7 gives

$$t_k^{(i)} = (\alpha_k I + \mathcal{A})^{-1} F'(x^\dagger)^* h_k^{(i)}, \quad i = 1, \ldots, 4$$

with

$$\|h_k^{(i)}\| \lesssim K_0\left(\|e_0\| + \|\hat{e}_0\|\right)\|x_k - \hat{x}_k\|\alpha_k^{1/2}, \quad i = 1, \ldots, 4.$$

Consequently

$$x_{k+1} - \hat{x}_{k+1} = \alpha_k\left(\alpha_k I + \mathcal{A}\right)^{-1}\left(x_k - \hat{x}_k\right) + (\alpha_k I + \mathcal{A})^{-1} F'(x^\dagger)^*\xi_k, \tag{6.12}$$

where $\xi_k := h_k^{(1)} + h_k^{(2)} + h_k^{(3)} + h_k^{(4)}$ satisfies

$$\|\xi_k\| \lesssim K_0\left(\|e_0\| + \|\hat{e}_0\|\right)\|x_k - \hat{x}_k\|\alpha_k^{1/2}. \tag{6.13}$$

Now we prove (6.9) by induction. It is obvious for $k = 0$. Suppose it is true for all $0 \le k \le l$. By telescoping (6.12) we have

$$x_{l+1} - \hat{x}_{l+1} = \prod_{j=0}^{l} \alpha_j \left(\alpha_j I + \mathcal{A}\right)^{-1} \left(x_0 - \hat{x}_0\right) + \sum_{m=0}^{l} \alpha_m^{-1} \prod_{j=m}^{l} \alpha_j \left(\alpha_j I + \mathcal{A}\right)^{-1} F'(x^\dagger)^* \xi_m. \tag{6.14}$$

Thus, by using (6.13), the induction hypothesis, and (2.8), we have

$$\begin{aligned}
&\|x_{l+1} - \hat{x}_{l+1}\| \\
&\le \|x_0 - \hat{x}_0\| + C K_0 \left(\|e_0\| + \|\hat{e}_0\|\right) \|x_0 - \hat{x}_0\| \sum_{m=0}^{l} \alpha_m^{-1/2} \left(\sum_{j=m}^{l} \alpha_j^{-1}\right)^{-1/2} \\
&\le \left(1 + C K_0 \left(\|e_0\| + \|\hat{e}_0\|\right)\right) \|x_0 - \hat{x}_0\| \\
&\le 2\|x_0 - \hat{x}_0\|
\end{aligned}$$

if both $K_0\|e_0\|$ and $K_0\|\hat{e}_0\|$ are suitably small. This shows (6.9).

Next we prove (6.10). By (2.6) it is obvious for $k = 0$. In the following we will consider the case $k > 0$. It follows from (6.14) that

$$F'(x^\dagger)(x_k - \hat{x}_k) = \prod_{j=0}^{k-1} \alpha_j (\alpha_j I + \mathcal{B})^{-1} F'(x^\dagger)(x_0 - \hat{x}_0) + \sum_{m=0}^{k-1} \alpha_m^{-1} \prod_{j=m}^{k-1} \alpha_j (\alpha_j I + \mathcal{B})^{-1} \mathcal{B} \xi_m.$$

Thus, by using (3.6), (2.7), (4.3), (6.13) and (6.9), we have

$$\begin{aligned}
\|F'(x^\dagger)(x_k - \hat{x}_k)\| &\le \left(\sum_{j=0}^{k-1} \alpha_j^{-1}\right)^{-1/2} \|x_0 - \hat{x}_0\| \\
&\quad + C K_0 \left(\|e_0\| + \|\hat{e}_0\|\right) \|x_0 - \hat{x}_0\| \sum_{m=0}^{k-1} \alpha_m^{-1/2} \left(\sum_{j=m}^{k-1} \alpha_j^{-1}\right)^{-1} \\
&\le \left(c_0 + C K_0 \left(\|e_0\| + \|\hat{e}_0\|\right)\right) \|x_0 - \hat{x}_0\| \alpha_k^{1/2} \\
&\le 2 c_0 \|x_0 - \hat{x}_0\| \alpha_k^{1/2}
\end{aligned}$$

if both $K_0\|e_0\|$ and $K_0\|\hat{e}_0\|$ are suitably small. We therefore obtain (6.10). □

Proof of Proposition 2 Let $0 < \varepsilon < \|e_0\|$ be an arbitrarily small number. Since $x_0 - x^\dagger \in \mathcal{N}(F'(x^\dagger))^\perp = \overline{\mathcal{R}(F'(x^\dagger)^*)}$, there is an $\hat{x}_0 \in X$ such that $\hat{x}_0 - x^\dagger \in \mathcal{R}(F'(x^\dagger)^*)$ and $\|x_0 - \hat{x}_0\| < \varepsilon$. Note that $\hat{x}_0 \in B_\rho(x^\dagger)$ and $K_0\|\hat{e}_0\| \le 2K_0\|e_0\|$. Thus, if $K_0\|e_0\|$ is suitably small, then for the sequence $\{\hat{x}_k\}$ defined by (6.2), it follows from Lemma 9 that

$$\|x_k - \hat{x}_k\| \le 2\|x_0 - \hat{x}_0\| < 2\varepsilon$$

and

$$\alpha_k^{-1/2}\|F'(x^\dagger)(x_k - \hat{x}_k)\| \le 2c_0\|x_0 - \hat{x}_0\| < 2c_0\varepsilon$$

for all $k \ge 0$, while by Remark 2 we have $\|\hat{e}_k\| \to 0$ and $\alpha_k^{-1/2}\|F'(x^\dagger)\hat{e}_k\| \to 0$ as $k \to \infty$. Thus, there is a k_0 such that $\|\hat{e}_k\| < \varepsilon$ and $\alpha_k^{-1/2}\|F'(x^\dagger)\hat{e}_k\| < c_0\varepsilon$ for all $k \ge k_0$. Consequently

$$\|e_k\| \le \|x_k - \hat{x}_k\| + \|\hat{e}_k\| < 3\varepsilon$$

and

$$\alpha_k^{-1/2}\|F'(x^\dagger)e_k\| \le \alpha_k^{-1/2}\|F'(x^\dagger)(x_k - \hat{x}_k)\| + \alpha_k^{-1/2}\|F'(x^\dagger)\hat{e}_k\| < 3c_0\varepsilon$$

for all $k \ge k_0$. Since $\varepsilon > 0$ is arbitrarily small, we therefore obtain (6.1). □

7 A crucial inequality

In this section we will establish the following important inequality.

Proposition 3 *Let the conditions in Lemma 5 be fulfilled. If $K_0\|e_0\|$ is suitably small, then*

$$\|x_k - x_l\| \lesssim \frac{1}{\sqrt{\alpha_l}}\|F(x_k) - y\|$$

for all integers $0 \le k \le l < \infty$.

The proof is based on the following result which shows that $\{\|F(x_k) - y\|\}$ is monotonically decreasing with respect to k.

Lemma 10 *Let the conditions in Lemma 5 be fulfilled. If $K_0\|e_0\|$ is suitably small, then*

$$\|F(x_{k+1}) - y\| \le \|F(x_k) - y\|$$

for all $k \ge 0$.

Proof We first have

$$\begin{aligned}
&\|F(x_{k+1}) - y\|^2 - \|F(x_k) - y\|^2 \\
&\quad = \|F(x_{k+1}) - F(x_k)\|^2 + 2\,(F(x_{k+1}) - F(x_k), F(x_k) - y) \\
&\quad = \|F(x_{k+1}) - F(x_k)\|^2 + 2\left(x_{k+1} - x_k, F'(x_k)^*(F(x_k) - y)\right) \\
&\qquad + 2\left(F(x_{k+1}) - F(x_k) - F'(x_k)(x_{k+1} - x_k), F(x_k) - y\right). \qquad (7.1)
\end{aligned}$$

By using (2.12) we have for $s := 2\left(x_{k+1} - x_k, F'(x_k)^*(F(x_k) - y)\right)$ that

$$\begin{aligned}
s &= -2\left(x_{k+1} - x_k, (\alpha_k I + F'(x_k)^* F'(x_k))(x_{k+1} - x_k)\right) \\
&= -2\alpha_k \|x_{k+1} - x_k\|^2 - 2\|F'(x_k)(x_{k+1} - x_k)\|^2.
\end{aligned}$$

Note that the elementary inequality $a^2 \ge \frac{1}{1+\mu}(a+b)^2 - \frac{1}{\mu}b^2$ for any $\mu > 0$, we obtain

$$\begin{aligned}
s \le -2\alpha_k \|x_{k+1} - x_k\|^2 &- \frac{2}{1+\mu}\|F(x_{k+1}) - F(x_k)\|^2 \\
&+ \frac{2}{\mu}\|F(x_{k+1}) - F(x_k) - F'(x_k)(x_{k+1} - x_k)\|^2.
\end{aligned}$$

Combining this estimate with (7.1) yields

$$\begin{aligned}
&\|F(x_{k+1}) - y\|^2 - \|F(x_k) - y\|^2 \\
&\quad \le -2\alpha_k \|x_{k+1} - x_k\|^2 - \left(\frac{2}{1+\mu} - 1\right)\|F(x_{k+1}) - F(x_k)\|^2 \\
&\qquad + \frac{2}{\mu}\|F(x_{k+1}) - F(x_k) - F'(x_k)(x_{k+1} - x_k)\|^2 \\
&\qquad + 2\left(F(x_{k+1}) - F(x_k) - F'(x_k)(x_{k+1} - x_k), F(x_k) - y\right). \qquad (7.2)
\end{aligned}$$

Note that (3.1) implies

$$\begin{aligned}
&\|F(x_{k+1}) - F(x_k) - F'(x_k)(x_{k+1} - x_k)\| \\
&\quad \le \frac{1}{2}K_0 \|x_{k+1} - x_k\| \|F'(x_k)(x_{k+1} - x_k)\|. \qquad (7.3)
\end{aligned}$$

With the help of Lemma 5 we then obtain

$$\|F(x_{k+1}) - F(x_k) - F'(x_k)(x_{k+1} - x_k)\| \lesssim K_0 \|e_0\| \|F'(x_k)(x_{k+1} - x_k)\|$$

which implies that if $K_0\|e_0\|$ is suitably small then

$$\|F'(x_k)(x_{k+1} - x_k)\| \le 2\|F(x_{k+1}) - F(x_k)\|.$$

Consequently, it follows from (7.3) that

$$\|F(x_{k+1}) - F(x_k) - F'(x_k)(x_{k+1} - x_k)\| \le K_0\|x_{k+1} - x_k\|\|F(x_{k+1}) - F(x_k)\|.$$

Applying this inequality to (7.2) yields

$$\begin{aligned}
&\|F(x_{k+1}) - y\|^2 - \|F(x_k) - y\|^2 \\
&\quad\le -2\alpha_k\|x_{k+1} - x_k\|^2 - \left(\frac{2}{1+\mu} - 1\right)\|F(x_{k+1}) - F(x_k)\|^2 \\
&\qquad + \frac{2}{\mu}K_0^2\|x_{k+1} - x_k\|^2\|F(x_{k+1}) - F(x_k)\|^2 \\
&\qquad + 2K_0\|F(x_k) - y\|\|x_{k+1} - x_k\|\|F(x_{k+1}) - F(x_k)\|.
\end{aligned}$$

Using the inequality

$$\begin{aligned}
&2K_0\|F(x_k) - y\|\|x_{k+1} - x_k\|\|F(x_{k+1}) - F(x_k)\| \\
&\quad\le \varepsilon\|F(x_{k+1}) - F(x_k)\|^2 + \frac{1}{\varepsilon}K_0^2\|F(x_k) - y\|^2\|x_{k+1} - x_k\|^2,
\end{aligned}$$

where $\varepsilon > 0$ is an arbitrary number, we then derive that

$$\begin{aligned}
&\|F(x_{k+1}) - y\|^2 - \|F(x_k) - y\|^2 \\
&\quad\le -\left(\frac{2}{1+\mu} - 1 - \varepsilon - \frac{2}{\mu}K_0^2\|x_{k+1} - x_k\|^2\right)\|F(x_{k+1}) - F(x_k)\|^2 \\
&\qquad - \left(2\alpha_k - \frac{1}{\varepsilon}K_0^2\|F(x_k) - y\|^2\right)\|x_{k+1} - x_k\|^2.
\end{aligned}$$

With the help of Lemma 5 we have $\|x_{k+1} - x_k\| \le C\|e_0\|$ and $\|F(x_k) - y\| \le C\|F'(x^\dagger)e_k\| \le C\alpha_k^{1/2}\|e_0\|$. Thus

$$\begin{aligned}
&\|F(x_{k+1}) - y\|^2 - \|F(x_k) - y\|^2 \\
&\quad\le -\left(\frac{2}{1+\mu} - 1 - \varepsilon - \frac{2}{\mu}C^2K_0^2\|e_0\|^2\right)\|F(x_{k+1}) - F(x_k)\|^2 \\
&\qquad - \alpha_k\left(2 - \frac{1}{\varepsilon}C^2K_0^2\|e_0\|^2\right)\|x_{k+1} - x_k\|^2.
\end{aligned}$$

This inequality suggests that if we take $\mu > 0$ and $\varepsilon > 0$ so small that $\frac{2}{1+\mu} - 1 - \varepsilon > 0$ and if $K_0\|e_0\|$ is small enough, then there holds

$$\|F(x_{k+1}) - y\|^2 - \|F(x_k) - y\|^2 \le 0.$$

This is exactly what we want to prove. □

Proof of Proposition 3 From (2.12) it follows that for any $j = 0, 1, \ldots$

$$\|x_{j+1} - x_j\| = \|(\alpha_j I + \mathcal{A}_j)^{-1} F'(x_j)^*(F(x_j) - y)\| \le \frac{1}{2\sqrt{\alpha_j}} \|F(x_j) - y\|.$$

Therefore

$$\|x_l - x_k\| \le \sum_{j=k}^{l-1} \|x_{j+1} - x_j\| \le \sum_{j=k}^{l-1} \frac{1}{2\sqrt{\alpha_j}} \|F(x_j) - y\|.$$

By using the monotonicity of $\{\|F(x_j) - y\|\}$ proved in Lemma 10 we then have

$$\|x_l - x_k\| \le \left(\sum_{j=k}^{l-1} \frac{1}{2\sqrt{\alpha_j}} \right) \|F(x_k) - y\|.$$

Since $\{\alpha_j\}$ is given by (2.2), it is easy to see that $\sum_{j=k}^{l-1} \alpha_j^{-1/2} \lesssim \alpha_l^{-1/2}$. We thus complete the proof. □

8 Proof of Theorem 1

The conclusion (2.9) is proved in Lemma 6, where we also obtain $k_\delta \le \tilde{k}_\delta$. In the following we will prove (2.10) and (2.11). From the definition of k_δ and Proposition 1 we have

$$\|F(x_{k_\delta}) - y\| \le \|F(x_{k_\delta}^\delta) - y^\delta\| + \|F(x_{k_\delta}^\delta) - F(x_{k_\delta}) - y^\delta + y\| \lesssim \delta \tag{8.1}$$

and for $0 \le k < k_\delta$

$$\begin{aligned}
\tau\delta &\le \|F(x_k^\delta) - y^\delta\| \\
&\le \|F(x_k^\delta) - F(x_k) - y^\delta + y\| + \|F(x_k) - y\| \\
&\le (1 + C K_0 \|e_0\|)\, \delta + \|F(x_k) - y\|.
\end{aligned}$$

Since $\tau > 1$, if $K_0\|e_0\|$ is suitably small, then

$$\delta \lesssim \|F(x_k) - y\| \lesssim \|F'(x^\dagger) e_k\|, \quad 0 \le k < \tilde{k}_\delta. \tag{8.2}$$

We now prove (2.10). Assume first that there is a sequence $\delta_n \searrow 0$ such that $k_n := k_{\delta_n} \to k$ as $n \to \infty$ for some finite integer k. Without loss of generality, we can assume that $k_n = k$ for all n. It then follows from (8.1) that $F(x_k) = y$. Thus, from (2.12) we can conclude that $x_j = x_k$ for all $j \ge k$. Since Proposition 2 implies $x_j \to x^\dagger$ as $j \to \infty$, we must have $x_k = x^\dagger$, which together with Proposition 1 implies $x_{k_n}^{\delta_n} \to x^\dagger$ as $n \to \infty$.

Assume next that there is a sequence $\delta_n \searrow 0$ such that $k_n := k_{\delta_n} \to \infty$ as $n \to \infty$. Then Proposition 2 and (8.2) imply that $\|e_{k_n}\| \to 0$ and $\delta_n/\sqrt{\alpha_{k_n}} \to 0$ as $n \to \infty$. Consequently, by Proposition 1 we again obtain $x_{k_n}^{\delta_n} \to x^\dagger$ as $n \to \infty$.

We next prove (2.11). By using Lemma 8 and (8.2) there is a positive constant $C_0 > 0$ such that

$$\delta < C_0 \alpha_k^{\nu+1/2} \|\omega\|, \quad 0 \le k < k_\delta.$$

Therefore

$$\alpha_k > \left(\frac{\delta}{C_0\|\omega\|}\right)^{2/(1+2\nu)}, \quad 0 \le k < k_\delta.$$

Now we define the integer $\bar{k}_\delta$ by

$$\alpha_{\bar{k}_\delta} \le \left(\frac{\delta}{C_0\|\omega\|}\right)^{2/(1+2\nu)} < \alpha_k, \quad 0 \le k < \bar{k}_\delta.$$

Then it is clear that $k_\delta \le \bar{k}_\delta$. Thus we may use Proposition 1, Proposition 3, (8.1) and Lemma 8 to obtain

$$\begin{aligned} \|e_{k_\delta}^\delta\| &\lesssim \|e_{k_\delta}\| + \frac{\delta}{\sqrt{\alpha_{k_\delta}}} \lesssim \|e_{\bar{k}_\delta}\| + \frac{\|F(x_{k_\delta}) - y^\delta\|}{\sqrt{\alpha_{\bar{k}_\delta}}} + \frac{\delta}{\sqrt{\alpha_{\bar{k}_\delta}}} \\ &\lesssim \alpha_{\bar{k}_\delta}^\nu \|\omega\| + \frac{\delta}{\sqrt{\alpha_{\bar{k}_\delta}}}. \end{aligned}$$

With the help of the definition of $\bar{k}_\delta$ and (2.2), it is now easy to obtain the desired estimate (2.11).

9 Numerical examples

In this section we present some numerical examples to test the convergence result in Theorem 1 on the method (2.1)–(2.3) by considering the estimation of the coefficient a in the two-point boundary value problem

$$\begin{cases} -u'' + au = f, & t \in (0, 1) \\ u(0) = g_0, & u(1) = g_1 \end{cases} \tag{9.1}$$

from the L^2 measurement u^δ of the state variable u, where g_0, g_1 and $f \in L^2[0, 1]$ are given. This inverse problem reduces to solving Eq. (1.1) with the nonlinear operator $F : D(F) \subset L^2[0, 1] \mapsto L^2[0, 1]$ defined as the parameter-to-solution mapping $F(a) := u(a)$, where $u(a)$ denotes the unique solution of (9.1). It is well known that F is well-defined on

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

Table 1 Numerical results for Example 1 with $\alpha_k = 1.0 \times (1.5)^{-k}$ and three different values of τ, where k_δ denotes the integer determined by (2.3), error $:= \|a^\delta_{k_\delta} - a^\dagger\|_{L^2}$, and ratio $:=$ error$/\delta^{1/2}$

δ	$\tau = 1.1$			$\tau = 2.5$			$\tau = 5.0$		
	k_δ	Error	Ratio	k_δ	Error	Ratio	k_δ	Error	Ratio
10^{-1}	10	1.69e − 1	0.53	1	7.20e − 1	2.28	1	7.20e − 1	2.28
10^{-2}	13	7.16e − 2	0.72	11	1.11e − 1	1.11	9	2.56e − 1	2.56
10^{-3}	19	1.79e − 2	0.57	15	4.77e − 2	1.51	13	7.16e − 2	2.26
10^{-4}	24	5.40e − 3	0.54	21	1.09e − 2	1.09	19	1.79e − 2	1.79
10^{-5}	29	1.60e − 3	0.50	26	3.30e − 3	1.04	24	5.40e − 3	1.70
10^{-6}	34	4.53e − 4	0.45	31	9.60e − 4	0.96	29	1.60e − 3	1.58

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

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

where $A(a) : H^2 \cap H^1_0 \mapsto L^2$ is defined by $A(a)u = -u'' + au$. It has been shown in [6] that if, for the sought solution $a^\dagger$, $|u(a^\dagger)(t)| \ge \kappa > 0$ for all $t \in [0, 1]$, then (2.5) is satisfied in a neighborhood of $a^\dagger$. In the following we report some numerical results on the method (2.1)–(2.3). During the computation, all differential equations are solved approximately by finite difference method by dividing the interval $[0, 1]$ into $n + 1$ subintervals with equal length $h = 1/(n + 1)$; we take $n = 200$ in our actual computation.

Example 1 We estimate a in (9.1) by assuming $f(t) = 1 + t$, $g_0 = 1$ and $g_1 = 2$. If $u(a^\dagger) = 1 + t$, then $a^\dagger = 1$ is the sought solution. In our computation, instead of $u(a^\dagger)$, we use the special perturbation

$$u^\delta = 1 + t + \delta\sqrt{2}\sin(\pi t/\delta)$$

with high frequency. Clearly $\|u^\delta - u(a^\dagger)\|_{L^2[0,1]} \le \delta$. As a first guess we choose $a_0 = 1 + 4t(1 - t)$. One can show that $a_0 - a^\dagger \in \mathcal{R}(F'(a^\dagger)^*)$. Thus, according to Theorem 1, the expected rate of convergence should be $O(\delta^{1/2})$.

In Table 1 we summarized the numerical results obtained by the method (2.1)–(2.3) with $\alpha_k = 1.0 \times (1.5)^{-k}$. In order to see the effect of τ in the discrepancy principle (2.3), we consider the three values $\tau = 1.1, 2.5$ and 5. In order to indicate the dependence of the convergence rates on the noise level, different values of δ are selected. The rates in Table 1 coincides with Theorem 1 very well. Table 1 indicates also that the absolute error increases with respect to τ. Thus, in numerical computation, one should use smaller τ if possible.

Table 2 Numerical results for Example 2, where error $:= \|a_{k_\delta}^\delta - a^\dagger\|_{L^2}$

δ	$a_0 = 1 + t^2 - 2t(1-t)(1+t-t^2)$			$a_0 = 1.2$	
	k_δ	Error	Error$/\delta^{1/2}$	k_δ	Error
10^{-2}	12	6.98e − 2	0.70	1	3.24e − 1
10^{-3}	16	4.50e − 3	0.14	20	1.86e − 1
10^{-4}	17	3.80e − 3	0.38	30	1.04e − 1
10^{-5}	23	6.68e − 4	0.21	38	6.71e − 2
10^{-6}	27	1.84e − 4	0.18	47	3.15e − 2
10^{-7}	31	6.04e − 5	0.19	55	7.50e − 3

Table 3 Numerical results on the effect of $\{\alpha_k\}$, where error $:= \|a_{k_\delta}^\delta - a^\dagger\|_{L^2}$

δ	$\alpha_k = 1.0 \times 2^{-k}$		$\alpha_k = 1.0 \times 4^{-k}$		$\alpha_k = 1.0 \times 8^{-k}$	
	k_δ	Error	k_δ	Error	k_δ	Error
10^{-2}	9	2.77e − 2	6	1.70e − 2	5	4.77e − 2
10^{-3}	11	4.50e − 3	7	6.30e − 3	6	3.68e − 2
10^{-4}	12	3.10e − 3	8	2.40e − 3	7	2.38e − 2
10^{-5}	15	5.80e − 4	9	8.66e − 4	7	2.40e − 3
10^{-6}	17	1.92e − 4	11	6.45e − 4	8	7.50e − 4

Example 2 In this example we consider the estimation of a in (9.1) with $f = 1 + t^2$ and $g_0 = g_1 = 1$. If $u(a^\dagger) = 1$, then $a^\dagger = 1 + t^2$ is the sought solution. When applying the method (2.1)–(2.3), we use the special noise data $u^\delta = 1 + \sqrt{2}\delta \sin(10\pi t)$, $\alpha_k = 1.0 \times (1.5)^{-k}$ and $\tau = 1.5$. In Table 2 we summarize the numerical results corresponding to two different choices of the initial guess

$$a_0 = 1 + t^2 - 2t(1-t)(1+t-t^2) \tag{9.2}$$

and

$$a_0 = 1.2. \tag{9.3}$$

For the a_0 given by (9.2) one can check $a_0 - a^\dagger \in \mathcal{R}(F'(a^\dagger)^* F'(a^\dagger))$. However, Table 2 indicates that the convergence rate is only $O(\delta^{1/2})$. This perhaps suggests that the best possible rate that the method (2.1)–(2.3) can provide is $O(\delta^{1/2})$ even if $a_0 - a^\dagger$ has higher regularity, which sharply contrasts to the linear situation. On the other hand, for the a_0 given by (9.3), $a_0 - a^\dagger \notin \mathcal{R}(F'(a^\dagger)^*)$, and in fact $a_0 - a^\dagger$ has no source-wise representation $a_0 - a^\dagger \in \mathcal{R}((F'(a^\dagger)^* F'(a^\dagger))^\nu)$ with a good $\nu > 0$. However, Table 2 clearly indicates the convergence of the method although it could be slow.

Example 3 In this example we give numerical results to indicate the effect of $\{\alpha_k\}$. We consider the estimation of a in (9.1) with the same data as in Example 2, i.e., $f = 1 + t^2$, $g_0 = g_1 = 1$, $u(a^\dagger) = 1$, $a^\dagger = 1 + t^2$ and $u^\delta = 1 + \sqrt{2}\delta \sin(10\pi t)$. Table 3 summarizes the numerical results obtained by the method (2.1)–(2.3) with $\tau = 2.0$ and the first guess a_0 given by (9.2) but with several different choices of $\{\alpha_k\}$. It can be seen that the choices of $\{\alpha_k\}$ affects the convergence speed. A faster decreasing sequence $\{\alpha_k\}$ could give the final results using fewer iterations, but the risk of worse convergence could arise. In general it is hard to tell how to choose the sequence $\{\alpha_k\}$ to achieve the best efficiency; an ideal approximation, however, should be able to obtain if $\{\alpha_k\}$ does not decrease too fast.

References

1. Bakushinskii, A.B.: The problems of the convergence of the iteratively regularized Gauss–Newton method. Comput. Math. Math. Phys. **32**, 1353–1359 (1992)
2. Groetsch, C.W.: Stable Approximate Evaluation of Unbounded Operators. Lecture Notes in Mathematics, vol. 1894. Springer, Berlin (2007)
3. Hanke, M.: A regularizing Levenberg–Marquardt scheme with applications to inverse groundwater filtration problems. Inverse Probl. **13**, 79–95 (1997)
4. Hanke, M.: The regularizing Levenberg–Marquardt scheme is of optimal order (2009, preprint)
5. Hanke, M., Groetsch, C.W.: Nonstationary iterated Tikhonov regularization. J. Optim. Theory Appl. **97**(1), 37–53 (1998)
6. Hanke, M., Neubauer, A., Scherzer, O.: A convergence analysis of Landweber iteration of nonlinear ill-posed problems. Numer. Math. **72**, 21–37 (1995)
7. Jin, Q.N.: On the iteratively regularized Gauss–Newton method for solving nonlinear ill-posed problems. Math. Comput. **69**(232), 1603–1623 (2000)
8. Jin, Q.N., Hou, Z.Y.: On an a posteriori parameter choice strategy for Tikhonov regularization of nonlinear ill-posed problems. Numer. Math. **83**(1), 139–159 (1999)
9. Jin, Q.N., Tautenhahn, U.: On the discrepancy principle for some Newton type methods for solving nonlinear inverse problems. Numer. Math. **111**, 509–558 (2009)
10. Kaltenbacher, B., Neubauer, A., Scherzer, O.: Iterative regularization methods for nonlinear ill-posed problems. de Gruyter, Berlin (2008)
11. Rieder, A.: On the regularization of nonlinear ill-posed problems via inexact Newton iterations. Inverse Probl. **15**, 309–327 (1999)
12. Rieder, A.: On convergence rates of inexact Newton regularizations. Numer. Math. **88**, 347–365 (2001)