Error estimates of some Newton-type methods for solving nonlinear inverse problems in Hilbert scales

2000 Inverse Problems 16 187

(http://iopscience.iop.org/0266-5611/16/1/315)

Inverse Problems **16** (2000) 187–197. Printed in the UK PII: S0266-5611(00)06993-8

Error estimates of some Newton-type methods for solving nonlinear inverse problems in Hilbert scales

Jin Qi-nian
Institute of Mathematics, Nanjing University, Nanjing 210008, People's Republic of China
E-mail: galgebra@nju.edu.cn

Received 12 August 1999, in final form 17 November 1999

Abstract. In this paper we consider some Newton-type methods in Hilbert scales to solve nonlinear inverse problems. Under certain conditions we obtain the error estimates when the iteration is terminated in an *a posteriori* manner. Finally we present the numerical examples to verify the theoretical results.

1. Introduction

In this paper we consider the nonlinear equation

$$F(x) = y_0 \tag{1.1}$$

arising from the study of nonlinear inverse problems, where F is a 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 the data y_0 is attainable, i.e. $y_0 \in R(F)$. Throughout this paper we assume that F is continuous and Fréchet differentiable; the Fréchet derivative of F at point $x \in D(F)$ and its adjoint are denoted by $F'(x)$ and $F'(x)^*$ respectively. The interest of this paper is confined to the case that the problem (1.1) is ill-posed in the sense that the solution does not depend continuously on the right-hand side; the reader can refer to [3, 4] for a number of important inverse problems in natural science leading to such case.

By assuming y_δ to be the only available approximation of y_0 satisfying

$$\|y_\delta - y_0\| \leqslant \delta \tag{1.2}$$

with a given noise level $\delta > 0$, now the reconstruction of the solution of (1.1) comes into being and several stable methods have been applied to touch on this topic in the literature. Among the methods developed to solve nonlinear inverse problems, Tikhonov regularization is the most well known and has received much attention in recent years. Its convergence analysis has been carried out under quite general conditions with *a priori* or *a posteriori* choice of the regularization parameter (see [3, 9] and references therein). As alternatives to Tikhonov regularization, iterative methods have been also used to overcome the difficulty arising from the ill-posedness of nonlinear inverse problems. Due to their straightforward implementation for the numerical solutions, such methods have attracted more and more attention in recent researches. Landweber iteration was extended to study nonlinear problems in [6] with an elegant convergence analysis; the idea therein was extensively utilized to analyse other iterative methods such as the method of steepest descent [15], the regularizing Levenberg–Marquardt scheme [5] and so on. Newton-type methods were also considered because of their

0266-5611/00/010187+11$30.00

faster convergence and a general framework was given in [10]. In particular, the iteratively regularized Gauss–Newton method proposed in [1] was reconsidered in [2, 8]; it seems that this method is becoming more and more popular in solving nonlinear inverse problems [7].

In this paper we will consider some Newton-type methods for nonlinear inverse problems in Hilbert scales. We formulate our methods in section 2.2 and derive the error estimates in section 2.3 under certain conditions. In section 3, we perform the numerical experiments to verify the theoretical results. Such research has several advantages, which are listed in section 2.2. We note that it was Natterer [12] who first considered the linear Tikhonov regularization in Hilbert scales to prevent the saturation phenomenon; this method was further considered in [13] under the choice of the regularization parameter by Morozov's discrepancy principle; an extension of Tikhonov regularization for linear ill-posed problems in Hilbert scales to a general class of regularization methods was then given in [17]. The extension of Natterer's technique to the study of the nonlinear case has also attracted some attention in recent years; Tikhonov regularization in Hilbert scales for nonlinear inverse problems was analysed in [11,14,16], and then a general regularization scheme in Hilbert scales for nonlinear equations was given in [18].

2. Newton-type methods in Hilbert scales

2.1. Hilbert scales

In this section we give a brief description of Hilbert scales and list some properties used in the subsequent discussion. For more information see [3].

A family of Hilbert spaces $(X_s)_{s\in R}$ is called a Hilbert scale induced by L if, for each $s \in R$, X_s is the completion of $\cap_{k=0}^{\infty} D(L^k)$ with respect to the Hilbert space norm $\|x\|_s := \|L^{s/2}x\|$, where L is a densely defined selfadjoint strictly positive operator in X. It is well known that X_s is densely and continuously embedded in X_r for any $-\infty < r < s < \infty$ and $(X_s)' = X_{-s}$ for any $s \geqslant 0$. Moreover there holds the important interpolation inequality, i.e. for any $-\infty < q < r < s < \infty$ there holds for any $x \in X_s$ that

$$\|x\|_r \leqslant \|x\|_q^{\frac{s-r}{s-q}} \|x\|_s^{\frac{r-q}{s-q}}. \tag{2.1}$$

Let $T : X \mapsto Y$ be a bounded linear operator. Since $L^{-s/2}$ is an isomorphism from X onto X_s, the operator $B := TL^{-s/2} : X \mapsto Y$ is also bounded for $s \geqslant 0$ and the adjoint of B is given by $B^* = L^{-s/2}T^*$, where $T^* : Y \mapsto X$ is the adjoint of T. Let $g : [0, \|B\|^2] \mapsto R$ be a continuous function, then

$$g(B^*B)L^{s/2} = L^{s/2}g(L^{-s}T^*T). \tag{2.2}$$

If we suppose further that there exist constants $M \geqslant m > 0$ and $a \geqslant 0$ such that for all $h \in X$ there holds

$$m\|h\|_{-a} \leqslant \|Th\| \leqslant M\|h\|_{-a},$$

then for all $|\nu| \leqslant 1$ the inequality (see [12])

$$c(\nu)\|h\|_{-\nu(a+s)} \leqslant \|(B^*B)^{\nu/2}h\| \leqslant C(\nu)\|h\|_{-\nu(a+s)} \tag{2.3}$$

holds on $D((B^*B)^{\nu/2})$ with

$$c(\nu) := \min\{m^\nu, M^\nu\}, \qquad C(\nu) := \max\{m^\nu, M^\nu\}. \tag{2.4}$$

Moreover,

$$R((B^*B)^{\nu/2}) = X_{\nu(a+s)}. \tag{2.5}$$

If $h \in X_\nu$ with some $0 \leqslant \nu \leqslant a+2s$ then $L^{s/2}h \in X_{\nu-s}$. Since $|\frac{\nu-s}{a+s}| \leqslant 1$ it follows from (2.5) that there is a $v \in X$ such that

$$L^{s/2}h = (B^*B)^{\frac{\mu-s}{2(a+s)}} v, \tag{2.6}$$

and from (2.3) we have

$$\begin{aligned}\|v\| &= \|(B^*B)^{-\frac{\mu-s}{2(a+s)}} L^{s/2}h\| \\ &\leqslant C\left(\frac{s-\mu}{a+s}\right)\|L^{s/2}h\|_{\mu-s} \\ &= C\left(\frac{s-\mu}{a+s}\right)\|h\|_\mu. \end{aligned}\tag{2.7}$$

2.2. *The methods*

Let us formulate some conditions first. We assume that $x^\dagger$ is a solution of (1.1) and x_0 is an available initial guess such that

$$x_0 \in B_\rho(x^\dagger) := \{x \in X : \|x - x^\dagger\| \leqslant \rho\} \subset D(F) \tag{2.8}$$

for a suitable small $\rho > 0$; moreover, we assume that

$$F'(x) = R_x F'(x^\dagger), \qquad \forall x \in B_\rho(x^\dagger), \tag{2.9}$$

where $\{R_x : x \in B_\rho(x^\dagger)\}$ is a family of bounded linear operators $R_x : Y \mapsto X$ such that

$$\|I - R_x\| \leqslant \eta \tag{2.10}$$

with a sufficiently small constant $\eta < 1$. Such an assumption has been verified for several nonlinear inverse problems in [6], and as a consequence of it we have the following estimate:

$$\|F(x) - y_0 - F'(x^\dagger)(x - x^\dagger)\| \leqslant \eta\|F'(x^\dagger)(x - x^\dagger)\|, \qquad x \in B_\rho(x^\dagger). \tag{2.11}$$

In order to formulate our methods in general form, we assume that $s \geqslant 0$ is a given number and the nonlinear operator F is properly scaled in the way that

$$\|F'(x)L^{-s/2}\| \leqslant 1, \qquad x \in B_\rho(x^\dagger). \tag{2.12}$$

We note that (2.9) and (2.10) imply $\|F'(x)L^{-s/2}\| \leqslant (1+\eta)\|F'(x^\dagger)L^{-s/2}\|$ which shows the uniform boundedness of $\|F'(x)L^{-s/2}\|$ in $B_\rho(x^\dagger)$. Therefore the scaling condition (2.12) can always be fulfilled by multiplying both sides of (1.1) by a sufficiently small constant, which then appears as a relaxation parameter in the methods presented below.

Now we suppose that x_n is the current iterate and consider the linearization

$$A_n(x - x_0) = -(F(x_n) - y_\delta - A_n(x_n - x_0)) \tag{2.13}$$

of equation (1.1), where $A_n := F'(x_n)$. Since (2.13) is in general ill-posed, the linear regularization methods should be used to generate an approximation of x which will be used as the new iterate x_{n+1}. To this end, let us choose a parameter-dependent family of real-valued continuous functions $\{g_\alpha : [0,1] \mapsto R\}_{\alpha\in(0,\infty)}$ with the properties that there exist constants τ_1, τ_2 and $\beta_0 \geqslant 1$ such that

$$\sup_{\lambda\in[0,1]} |\lambda^\gamma g_\alpha(\lambda)| \leqslant \tau_1\alpha^{\gamma-1}, \qquad \text{for} \quad 0 \leqslant \gamma \leqslant 1, \tag{2.14}$$

$$\sup_{\lambda\in[0,1]} |\lambda^\beta r_\alpha(\lambda)| \leqslant \tau_2\alpha^{\beta}, \qquad \text{for} \quad 0 \leqslant \beta \leqslant \beta_0, \tag{2.15}$$

where $r_\alpha(\lambda) := 1 - \lambda g_\alpha(\lambda)$. Similar assumptions to (2.14) and (2.15) for g_α have also been used in [10, 17]. Clearly, if $x_n \in B_\rho(x^\dagger)$ then for $B_n := F'(x_n)L^{-s/2}$ we have $\|B_n\| \leqslant 1$, therefore g_α can be used as the filter function and we thus obtain the Newton-type methods in Hilbert scales

$$x_{n+1} = x_0 - L^{-s/2} g_{\alpha_n}(B_n^* B_n) B_n^* (F(x_n) - y_\delta - A_n(x_n - x_0)) \tag{2.16}$$

where $\{\alpha_n\}$ is a given sequence satisfying

$$\alpha_n > 0, 1 \leqslant \frac{\alpha_n}{\alpha_{n+1}} \leqslant p \qquad \text{and} \qquad \lim_{n\to\infty} \alpha_n = 0 \tag{2.17}$$

for some constant $p > 1$. Using the property (2.2) we can rewrite (2.16) as

$$x_{n+1} = x_0 - g_{\alpha_n}(L^{-s} A_n^* A_n) L^{-s} A_n^* (F(x_n) - y_\delta - A_n(x_n - x_0)). \tag{2.18}$$

It is well known in the regularization theory that if x_n is used to approximate $x^\dagger$, the iteration should be terminated properly. Due to the practical applications, the stopping index of iteration should be designated in an *a posteriori* manner. In this paper we will consider the iterative discrepancy principle and choose n_δ as the integer such that

$$\|F(x_{n_\delta}) - y_\delta\| \leqslant \tau\delta < \|F(x_n) - y_\delta\|, \qquad 0 \leqslant n < n_\delta, \tag{2.19}$$

where $\tau \geqslant 1$ is a suitable large number. We will use x_{n_δ} as an approximation of $x^\dagger$.

To proceed with our convergence analysis we need the following condition:

$$\begin{aligned} &\exists \text{ constants } M \geqslant m > 0 \text{ and } a \geqslant 0 \text{ such that } \forall h \in X \text{ there holds} \\ &m\|h\|_{-a} \leqslant \|F'(x)h\| \leqslant M\|h\|_{-a}, \forall x \in B_\rho(x^\dagger). \end{aligned} \tag{2.20}$$

Let us give some remarks on this condition. From [14] we know that the following condition has been used to analyse the nonlinear Tikhonov regularization in Hilbert scales:

$$\begin{aligned} &\exists \text{ constants } M \geqslant m > 0 \text{ and } a \geqslant 0 \text{ such that } \forall h \in X \text{ there holds} \\ &m\|h\|_{-a} \leqslant \|F'(x^\dagger)h\| \leqslant M\|h\|_{-a}. \end{aligned} \tag{2.21}$$

Clearly, (2.21) is just a special case of (2.20). However, (2.20) and (2.21) are equivalent under the conditions (2.9) and (2.10) since now we have the inequality

$$(1 - \eta)\|F'(x^\dagger)h\| \leqslant \|F'(x)h\| \leqslant (1 + \eta)\|F'(x^\dagger)h\|. \tag{2.22}$$

In many applications L is chosen to be a differential operator in L^2-space, in which case the number a in (2.21) can be interpreted as the smoothing index of $F'(x^\dagger)$. Therefore (2.20) requires that the smoothing index of $F'(x)$ is invariant around $x^\dagger$.

Now we can state the main result of this paper.

Theorem 2.1. *Let (2.8)–(2.10), (2.12), (2.17) and (2.20) be fulfilled and let g_α satisfy (2.14) and (2.15). Assume n_δ be the integer determined by the discrepancy principle (2.19) with $\tau > 1 + \tau_1\sqrt{p}$ and suppose $x_0 - x^\dagger \in X_\mu$ with some $0 \leqslant \mu \leqslant a + 2s$. If $\|x_0 - x^\dagger\|$ and η are suitable small, then for all $r \in [-a, \mu]$ there holds*

$$\|x_{n_\delta} - x^\dagger\|_r \leqslant C\|x_0 - x^\dagger\|_\mu^{\frac{a+r}{a+\mu}} \delta^{\frac{\mu-r}{a+\mu}} \tag{2.23}$$

with a constant C independent of δ and $\|x_0 - x^\dagger\|_\mu$.

We mention that the result of theorem 2.1 reduces to the well known result in the literature (see [17]) when F is linear. The proof of theorem 2.1 will be given in section 2.3. In the following we discuss the special choice of g_α to generate some regularization methods that fit into our framework.

As the first example, we apply the linear Tikhonov regularization in Hilbert scales to equation (2.13) and obtain the choice $g_\alpha(\lambda) := \frac{1}{\alpha+\lambda}$. A simple exercise shows that g_α satisfies (2.14) and (2.15) with $\tau_1 = \tau_2 = \beta_0 = 1$. Now the method (2.18) has the form

$$x_{n+1} = x_0 - (\alpha_n L^s + A_n^* A_n)^{-1} A_n^* (F(x_n) - y_\delta - A_n(x_n - x_0))$$

which can be rewritten as

$$x_{n+1} = x_n - (\alpha_n L^s + A_n^* A_n)^{-1}(A_n^*(F(x_n) - y_\delta) + \alpha_n L^s(x_n - x_0)). \tag{2.24}$$

This method is called the iteratively regularized Gauss–Newton method in Hilbert scales which is well known for $s = 0$ (see [1, 2, 8]).

In contrast to the above example, we may apply the Landweber iteration in Hilbert scales for linear ill-posed problems to the equation (2.13) to obtain the so-called Newton–Landweber method in Hilbert scales. By choosing $g_\alpha(\lambda) := \sum_{k=0}^{[\frac{1}{\alpha}]}(1-\lambda)^k$ which satisfies (2.14) and (2.15) with $\tau_1 = \tau_2 = 1$ and $\beta_0 = \infty$, (2.18) then becomes

$$x_{n+1} = x_0 - \sum_{k=0}^{[\frac{1}{\alpha_n}]}(I - L^{-s}A_n^*A_n)^k L^{-s}A_n^*(F(x_n) - y_\delta - A_n(x_n - x_0))$$

which is equivalent to the form

$$\begin{aligned} &x_{n,0} = x_0, \\ &x_{n,k+1} = x_{n,k} - L^{-s}A_n^*(F(x_n) + A_n(x_{n,k} - x_n) - y_\delta), \qquad 1 \leqslant k < [1/\alpha_n], \\ &x_{n+1} = x_{n,[1/\alpha_n]}. \end{aligned} \tag{2.25}$$

Such a method with $s = 0$ has been studied in [10] recently.

Let us conclude this section with several remarks.

Remark 2.1. Theorem 2.1 requires $x_0 - x^\dagger \in X_\mu$ with some $0 \leqslant \mu \leqslant 2s + a$ which can be viewed as a smoothness condition on $x_0 - x^\dagger$. A similar assumption was used in [11, 14] where Tikhonov regularization was considered in Hilbert scales to solve nonlinear inverse problems, but there it needed $\mu \geqslant a$ which seems to be a restrictive requirement.

Remark 2.2. It should be mentioned that the restriction $\mu \leqslant a+2s$ which leads to a saturation in the convergence rates is not required if L and $F'(x)^*F'(x)$ commute for each $x \in B_\rho(x^\dagger)$ which includes the case $L := I$. In this case, the inequality (2.3) holds with $B := F'(x)L^{-s/2}$ for all real ν, and the result of theorem 2.1 then holds for all $\mu \leqslant 2\beta_0(a + s) - a$. Hence for the regularization methods with qualification $\beta_0 = \infty$ the error bound (2.23) holds then for all $\mu < \infty$. This fact was first observed in [17] for linear regularization methods in Hilbert scales.

Remark 2.3. The *ordinary* iteratively regularized Gauss–Newton method (i.e. the method (2.24) with $s = 0$) has been considered in [2, 8]. It has been pointed out in [2] that the best possible convergence rate of x_{n_δ} to $x^\dagger$ is at most $O(\delta^{1/2})$ if n_δ is chosen by the discrepancy principle (2.19). Moreover, even using the n_δ chosen by the strategy in [8] we have, at most, the error bound $O(\delta^{2/3})$. However, if we consider this method in Hilbert scales, such a saturation phenomenon can be prevented if we choose s in a suitable way.

Remark 2.4. It has been noted in [10] that it seems quite difficult to work out the convergence if we use (2.19) to choose the stopping index of iteration for some Newton-type methods including Newton–Landweber iteration with $s = 0$. However, it is interesting to find that the case is quite different if we consider these methods in Hilbert scales.

Remark 2.5. The requirement $\tau > 1 + \tau_1\sqrt{p}$ given in theorem 2.1 mainly arises for technical reasons (see section 2.3). Additional effort makes it possible to drop this restriction. If one can show that the discrepancy principle (2.19) is well-defined for smaller τ, one should use this smaller τ in the numerical computation since the absolute error increases with τ.

Remark 2.6. Tautenhahn [18] suggested a general regularization scheme in Hilbert scales for (1.1) which is closely related to our methods and established the order optimal convergence rates under certain conditions. His method, however, is indeed very different from ours since the regularizer in [18] is given implicitly by a nonlinear well-posed equation which in practice has to be solved approximately by an iteration and this additional approximation was not incorporated into the convergence analysis.

2.3. Error estimates

In this section we give the proof of theorem 2.1. We first show that x_n is well-defined for all $0 \leqslant n \leqslant n_\delta$ with n_δ determined by (2.19).

Suppose $x_n \in B_\rho(x^\dagger)$ for some n and set $e_n = x_n - x^\dagger$, then the definition (2.16) of x_{n+1} gives

$$\begin{aligned} e_{n+1} = x_0 - x^\dagger - L^{-s/2} g_{\alpha_n}(B_n^* B_n) B_n^* (F(x_n) - y_\delta - A_n(x_n - x_0)) \\ = L^{-s/2} r_{\alpha_n}(B_n^* B_n) L^{s/2}(x_0 - x^\dagger) \\ - L^{-s/2} g_{\alpha_n}(B_n^* B_n) B_n^* (y_0 - y_\delta) \\ - L^{-s/2} g_{\alpha_n}(B_n^* B_n) B_n^* (F(x_n) - y_0 - A_n(x_n - x^\dagger)) \\ := I_1 + I_2 + I_3. \end{aligned} \tag{2.26}$$

Multiplying both sides of (2.26) by $A := F'(x^\dagger)$ and noting that $A = R_{x_n}^{-1} A_n$ we obtain

$$\begin{aligned} Ae_{n+1} = R_{x_n}^{-1} B_n r_{\alpha_n}(B_n^* B_n) L^{s/2}(x_0 - x^\dagger) \\ - R_{x_n}^{-1} B_n g_{\alpha_n}(B_n^* B_n) B_n^* (y_0 - y_\delta) \\ - R_{x_n}^{-1} B_n g_{\alpha_n}(B_n^* B_n) B_n^* (F(x_n) - y_0 - A_n(x_n - x^\dagger)). \end{aligned} \tag{2.27}$$

Based on (2.26) and (2.27) we can obtain

Lemma 2.1. *Let all the assumptions in theorem 2.1 be fulfilled. If $x_n \in B_\rho(x^\dagger)$ for some n, then for all $r \in [-a, \mu]$ there hold*

$$\|e_{n+1}\|_r \leqslant b_0(r) \alpha_n^{\frac{\mu - r}{2(a+s)}} \|x_0 - x^\dagger\|_\mu + b_1(r) \alpha_n^{-\frac{a+r}{2(a+s)}} (\delta + 2\eta \|Ae_n\|), \tag{2.28}$$

$$\|Ae_{n+1}\| \leqslant b_2 \alpha_n^{\frac{a+\mu}{2(a+s)}} \|x_0 - x^\dagger\|_\mu + \frac{\tau_1}{1-\eta}\delta + \frac{2\eta\tau_1}{1-\eta} \|Ae_n\|, \tag{2.29}$$

where $b_0(r) := c(\frac{s-r}{a+s})^{-1} C(\frac{s-\mu}{a+s})\tau_2$, $b_1(r) := c(\frac{s-r}{a+s})^{-1}\tau_1$ *and* $b_2 := \frac{\tau_2}{1-\eta} C(\frac{s-\mu}{a+s})$.

Proof. Since $x_0 - x^\dagger \in X_\mu$ with $0 \leqslant \mu \leqslant a + 2s$, from (2.6) and (2.7) it follows that there exists a $v_n \in X$ such that $L^{s/2}(x_0 - x^\dagger) = (B_n^* B_n)^{\frac{\mu - s}{2(a+s)}} v_n$ and

$$\|v_n\| \leqslant C\left(\frac{s-\mu}{a+s}\right) \|x_0 - x^\dagger\|_\mu. \tag{2.30}$$

Thus, by noting that $|\frac{s-r}{a+s}| \leqslant 1$ and $0 \leqslant \frac{\mu - r}{2(a+s)} \leqslant 1$ we can use (2.3), (2.15) and (2.30) to obtain

$$\begin{aligned} \|I_1\|_r &= \|L^{(r-s)/2} r_{\alpha_n}(B_n^* B_n) L^{s/2}(x_0 - x^\dagger)\| \\ &= \|r_{\alpha_n}(B_n^* B_n)(B_n^* B_n)^{\frac{\mu - s}{2(a+s)}} v_n\|_{r-s} \\ &\leqslant c\left(\frac{s-r}{a+s}\right)^{-1} \|(B_n^* B_n)^{\frac{\mu - r}{2(a+s)}} r_{\alpha_n}(B_n^* B_n) v_n\| \\ &\leqslant b_0(r) \alpha_n^{\frac{\mu - r}{2(a+s)}} \|x_0 - x^\dagger\|_\mu. \end{aligned}$$

Applying (2.3) again to I_2 we have from (2.14) that

$$\begin{aligned}\|I_2\|_r &= \|g_{\alpha_n}(B_n^* B_n) B_n^*(y_0 - y_\delta)\|_{r-s} \\ &\leqslant c\left(\frac{s-r}{a+s}\right)^{-1} \|(B_n^* B_n)^{\frac{s-r}{2(a+s)}} g_{\alpha_n}(B_n^* B_n) B_n(y_0 - y_\delta)\| \\ &= c\left(\frac{s-r}{a+s}\right)^{-1} \|(B_n B_n^*)^{\frac{a+2s-r}{2(a+s)}} g_{\alpha_n}(B_n B_n^*)(y_0 - y_\delta)\| \\ &\leqslant b_1(r)\alpha_n^{-\frac{a+r}{2(a+s)}} \delta.\end{aligned}$$

Similar to the estimate of I_2 it follows that

$$\|I_3\|_r \leqslant b_1(r)\alpha_n^{-\frac{a+r}{2(a+s)}} \|F(x_n) - y_0 - A_n(x_n - x^\dagger)\|.$$

By exploiting (2.10) and (2.11) we have

$$\|F(x_n) - y_0 - A_n(x_n - x^\dagger)\| \leqslant 2\eta \|Ae_n\|. \tag{2.31}$$

Therefore $\|I_3\|_r \leqslant 2\eta b_1(r)\alpha_n^{-\frac{a+r}{2(a+s)}} \|Ae_n\|$. Combining the above we obtain (2.28).

Let us prove (2.29) now. By noting that $\|R_{x_n}^{-1}\| \leqslant \frac{1}{1-\eta}$, it is easy to obtain

$$\begin{aligned}\|Ae_{n+1}\| \leqslant \frac{1}{1-\eta}\{ &\sup_{\lambda\in[0,1]} \lambda^{\frac{a+\mu}{2(a+s)}} |r_{\alpha_n}(\lambda)| \|v_n\| \\ &+ \sup_{\lambda\in[0,1]} |\lambda g_{\alpha_n}(\lambda)| (\delta + \|F(x_n) - y_0 - A_n(x_n - x^\dagger)\|\}.\end{aligned}$$

Using (2.14), (2.15), (2.30) and (2.31) we get (2.29). □

With lemma 2.1 at hand, let us show that $x_n \in B_\rho(x^\dagger)$ for all $0 \leqslant n \leqslant n_\delta$ if η and $\|x_0 - x^\dagger\|$ are suitably small. Indeed, if $x_n \in B_\rho(x^\dagger)$ for some $n < n_\delta$, then the definition of n_δ suggests that $\|F(x_n) - y_\delta\| > \tau\delta$. Since (1.2) and (2.11) imply $\|F(x_n) - y_\delta\| \leqslant \delta + (1 + \eta)\|Ae_n\|$, we therefore obtain

$$\delta < \frac{1+\eta}{\tau - 1}\|Ae_n\|. \tag{2.32}$$

Substituting (2.32) into (2.30) and (2.31) (with $\mu = r = 0$) we have

$$\|e_{n+1}\| \leqslant d_1\|x_0 - x^\dagger\| + \frac{(1-\eta+2\eta\tau)d_2}{\tau - 1}\alpha_n^{-\frac{a}{2(a+s)}}\|Ae_n\|, \tag{2.33}$$

$$\|Ae_{n+1}\| \leqslant d_3\alpha_n^{\frac{a}{2(a+s)}}\|x_0 - x^\dagger\| + \left(\frac{\tau_1}{\tau - 1} + \frac{2\eta\tau_1\tau}{(1-\eta)(\tau-1)}\right)\|Ae_n\|, \tag{2.34}$$

where $d_1 := \tau_2 C(\frac{s}{a+s})c(\frac{s}{a+s})^{-1}$, $d_2 := \tau_1 c(\frac{s}{a+s})^{-1}$ and $d_3 := \frac{\tau_2}{1-\eta}C(\frac{s}{a+s})$. Based on these observations, by induction we can obtain the following lemma.

Lemma 2.2. *Let all the assumptions in theorem 2.1 hold and $\tau > 1 + \tau_1\sqrt{p}$. If*

$$\eta < 1 - \frac{1}{1 + \frac{\tau - 1 - \tau_1\sqrt{p}}{2\tau(\tau-1)}} \qquad \text{and} \qquad D_1\|x_0 - x^\dagger\| \leqslant \rho \tag{2.35}$$

with

$$D_1 := \max\left\{1, d_1 + \frac{(1-\eta+2\eta\tau_1)d_2}{\tau - 1}D_2\right\}$$

$$D_2 := \frac{\sqrt{p}}{1 - (\frac{\tau_1}{\tau-1} + \frac{2\eta\tau_1\tau}{(\tau-1)(1-\eta)})\sqrt{p}} \max\{d_3, \|A\|\alpha_0^{-\frac{a}{2(a+s)}}\}$$

then $x_n \in B_\rho(x^\dagger)$ for all $0 \leqslant n \leqslant n_\delta$ and

$$\|e_n\| \leqslant D_1\|x_0 - x^\dagger\| \qquad \text{and} \qquad \|Ae_n\| \leqslant D_2\alpha_n^{\frac{a}{2(a+s)}}\|x_0 - x^\dagger\|. \tag{2.36}$$

Proof. Clearly the assertions are trivial for $n = 0$. Suppose that the assertions are valid for some $n < n_\delta$, then $x_n \in B_\rho(x^\dagger)$ and from (2.33), (2.34), the inductive hypothesis and (2.17) we have

$$\|e_{n+1}\| \leqslant \left(d_1 + \frac{(1-\eta+2\eta\tau)d_2}{\tau-1}D_2\right)\|x_0 - x^\dagger\| \leqslant D_1\|x_0 - x^\dagger\| \leqslant \rho$$

and by noting that $(\alpha_n/\alpha_{n+1})^{\frac{a}{2(a+s)}} \leqslant p^{\frac{a}{2(a+s)}} \leqslant \sqrt{p}$,

$$\begin{aligned}\|Ae_{n+1}\| &\leqslant \left(d_3 + \left(\frac{\tau_1}{\tau-1} + \frac{2\eta\tau_1\tau}{(\tau-1)(1-\eta)}\right)D_2\right)\alpha_n^{\frac{a}{2(a+s)}}\|x_0 - x^\dagger\| \\ &\leqslant D_2\alpha_{n+1}^{\frac{a}{2(a+s)}}\|x_0 - x^\dagger\|.\end{aligned}$$

Thus the proof is complete. □

Now we can show the justification of rule (2.19) if $\tau > 1 + \tau_1\sqrt{p}$. Suppose the contrary: we then have $n_\delta = \infty$ which together with lemma 2.2 suggests that $x_n \in B_\rho(x^\dagger)$ and (2.32) is valid for all $0 \leqslant n < \infty$. The combination of (2.32) and (2.36) immediately gives

$$\alpha_n^{\frac{a}{2(a+s)}} > \frac{(\tau-1)\delta}{(1+\eta)D_2\|x_0 - x^\dagger\|}$$

for all $0 \leqslant n < \infty$ which is a contradiction to (2.11) since $\lim_{n\to\infty}\alpha_n = 0$. Therefore the integer n_δ defined by the discrepancy principle (2.19) always exists and is finite.

Finally, we are in a position to give the proof of the main result.

Proof of theorem 2.1. Let $q > 0$ be the constant such that $\|x\| \leqslant q\|x\|_\mu$ for all $x \in X_\mu$ and set

$$d := \max\left\{\|A\|\alpha_0^{-\frac{a+\mu}{2(a+s)}}q, 2\left(b_2 + \frac{b_2\tau_1(1+\eta)}{(\tau-1)(1-\eta)-\tau_1(1+\eta)}\right)p^{\frac{a+\mu}{2(a+s)}}\right\}.$$

Then, under the conditions in theorem 2.1 we first show by induction that

$$\|Ae_n\| \leqslant d\alpha_n^{\frac{a+\mu}{2(a+s)}}\|x_0 - x^\dagger\|_\mu \tag{2.37}$$

for all $0 \leqslant n < n_\delta$ if η and $\|x_0 - x^\dagger\|$ are suitable small such that (2.35) holds. In fact, (2.37) is obvious for $n = 0$. Assume it is true for some $n - 1 < n_\delta - 1$. Then by noting that $n < n_\delta$ it follows from (2.29) and (2.32) that

$$\frac{\tau-1}{1+\eta}\delta \leqslant b_2\alpha_{n-1}^{\frac{a+\mu}{2(a+s)}}\|x_0 - x^\dagger\|_\mu + \frac{\tau_1}{1-\eta}\delta + \frac{2\eta\tau_1}{1-\eta}\|Ae_{n-1}\|.$$

Therefore we have

$$\delta \leqslant \frac{1-\eta^2}{(\tau-1)(1-\eta)-\tau_1(1+\eta)}\left(b_2\alpha_{n-1}^{\frac{a+\mu}{2(a+s)}}\|x_0 - x^\dagger\|_\mu + \frac{2\eta\tau_1}{1-\eta}\|Ae_{n-1}\|\right).$$

Substituting this into (2.29) then yields

$$\begin{aligned}\|Ae_n\| \leqslant{}& \left(b_2 + \frac{b_2\tau_1(1+\eta)}{(\tau-1)(1-\eta)-\tau_1(1+\eta)}\right)\alpha_{n-1}^{\frac{a+\mu}{2(a+s)}}\|x_0 - x^\dagger\|_\mu \\ &+ \frac{2\eta\tau_1(\tau-1)}{(\tau-1)(1-\eta)-\tau_1(1+\eta)}\|Ae_{n-1}\|.\end{aligned}$$

Thus by utilizing the inductive hypothesis and (2.17) it follows that

$$\begin{aligned}\|Ae_n\| &\leqslant \left\{\left(b_2 + \frac{b_2\tau_1(1+\eta)}{(\tau-1)(1-\eta)-\tau_1(1+\eta)}\right) + \frac{2\eta\tau_1(\tau-1)d}{(\tau-1)(1-\eta)-\tau_1(1+\eta)}\right\} \\ &\quad\times p^{\frac{a+\mu}{2(a+s)}}\alpha_n^{\frac{a+\mu}{2(a+s)}}\|x_0 - x^\dagger\|_\mu \\ &\leqslant d\alpha_n^{\frac{a+\mu}{2(a+s)}}\|x_0 - x^\dagger\|_\mu\end{aligned}$$

which completes the proof of (2.37).

Now we turn to the proof of (2.23). From (2.20), (2.11), the definition of n_δ, (2.28) and (2.37) we immediately get

$$\begin{aligned}\|e_{n_\delta}\|_{-a} &\leqslant \frac{1}{m}\|Ae_{n_\delta}\| \\ &\leqslant \frac{1}{m(1-\eta)}(\|F(x_{n_\delta}) - y_\delta\| + \delta) \\ &\leqslant \frac{1+\tau}{m(1-\eta)}\delta\end{aligned}$$

and

$$\begin{aligned}\|e_{n_\delta}\|_\mu &\leqslant b_0(\mu)\|x_0 - x^\dagger\|_\mu + b_1(\mu)\alpha_{n_\delta-1}^{-\frac{a+\mu}{2(a+s)}}(\delta + 2\eta\|Ae_{n_\delta-1}\|) \\ &\leqslant (b_0(\mu) + 2\eta b_1(\mu)d)\|x_0 - x^\dagger\|_\mu + b_1(\mu)\delta\alpha_{n_\delta-1}^{-\frac{a+\mu}{2(a+s)}}.\end{aligned}$$

Noting that (2.37) and (2.32) imply $\delta\alpha_{n_\delta-1}^{-\frac{a+\mu}{2(a+s)}} \leqslant \frac{(1+\eta)d}{\tau-1}\|x_0 - x^\dagger\|_\mu$ we thus have

$$\|e_{n_\delta}\|_\mu \leqslant \left(b_0(\mu) + 2\eta b_1(\mu)d + \frac{(1+\eta)d}{\tau-1}\right)\|x_0 - x^\dagger\|_\mu.$$

Therefore (2.23) follows by a simple application of the interpolation inequality (2.1). □

3. Numerical tests

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

$$\begin{aligned}&-u'' + cu = f, \qquad t \in (0,1)\\ &u(0) = h_0, \qquad u(1) = h_1\end{aligned} \tag{3.1}$$

from the measurement data u_δ of the state variable u, where h_0, h_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 $F(c) = u(c)$ with $u(c)$ being the unique solution of (3.1). F is well-defined on

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

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

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

where $A(c) : H^2 \cap H_0^1 \mapsto L^2$ is defined by $A(c)u = -u'' + cu$.

Let us define L to be the linear operator in $L^2[0,1]$ as follows:

$$L : H^2 \cap H_0^1[0,1] \subset L^2[0,1] \mapsto L^2[0,1], \qquad Lc = -c''.$$

It is easy to verify that L is densely defined, self-adjoint and positive definite, and the Hilbert scale $\{X_s\}$ induced by L is

$$X_s = \left\{c \in H^s[0,1] : c^{(2l)}(0) = c^{(2l)}(1) = 0,\ l = 0, 1, \ldots, \left[\frac{s}{2} - \frac{1}{4}\right]\right\} \tag{3.2}$$

for any $s \in R$, where $H^s[0,1]$ is the usual Sobolev space. Moreover, $\|c\|_s = \int_0^1 |c^{(s)}(t)|^2 \mathrm{d}t$ for all $s = 0, 1, 2, \ldots$.

Table 1. Numerical results for example 3.1, where Error $:= \|c_{n_\delta} - c^\dagger\|_{L^2}$.

	$s = 1$			$s = 0$		
δ	n_δ	Error	Error$/\delta^{5/9}$	n_δ	Error	Error$/\delta^{1/2}$
0.10e − 1	8	0.21e + 0	0.27e + 1	5	0.28e + 0	0.28e + 1
0.10e − 2	12	0.25e − 1	0.12e + 1	9	0.45e − 1	0.14e + 1
0.10e − 3	16	0.21e − 2	0.35e + 0	12	0.13e − 1	0.13e + 1
0.10e − 4	19	0.47e − 3	0.28e + 0	15	0.37e − 2	0.12e + 1
0.10e − 5	22	0.93e − 4	0.20e + 0	17	0.68e − 3	0.68e + 0
0.10e − 6	26	0.14e − 4	0.11e + 0	20	0.17e − 3	0.54e + 0
0.10e − 7	29	0.21e − 5	0.59e − 1	24	0.39e − 4	0.39e + 0

Table 2. Numerical results for example 3.2, where $e_1 := \|c_{n_\delta} - c^\dagger\|_{L^2}$ and $e_2 := \|c_{n_\delta} - c^\dagger\|_{H^1}$.

δ	n_δ	e_1	$e_1/\delta^{5/9}$	e_2	$e_2/\delta^{1/3}$
0.10e − 1	5	0.39e − 1	0.50e + 0	0.14e + 0	0.66e + 0
0.10e − 2	7	0.71e − 2	0.33e + 0	0.66e − 1	0.66e + 0
0.10e − 3	9	0.33e − 2	0.56e + 0	0.39e − 1	0.84e + 0
0.10e − 4	11	0.10e − 2	0.60e + 0	0.20e − 1	0.89e + 0
0.10e − 5	13	0.32e − 3	0.70e + 0	0.97e − 2	0.97e + 1
0.10e − 6	16	0.63e − 4	0.48e + 0	0.36e − 2	0.78e + 0
0.10e − 7	18	0.17e − 4	0.48e + 0	0.16e − 2	0.73e + 0

By assuming that the exact solution $c^\dagger$ has the property that $u_0 := u(c^\dagger) > 0$ on $[0, 1]$, we can verify (2.9) and (2.10) in a neighbourhood $B_\rho(c^\dagger)$ around $c^\dagger$, and following the lines in [11] we have (2.20) in $B_\rho(c^\dagger)$.

In the following we just do the numerical experiments for the iteratively regularized Gauss–Newton method in Hilbert scales. We remark that the differential equation problems we meet during computation are always solved by finite element method on the subspace of piecewise linear splines on a uniform grid with subinterval length $\frac{1}{128}$; the iterative solution c_{n_δ} is also approximated by the function from the finite-dimensional subspace of piecewise linear splines on a uniform grid with subinterval length $\frac{1}{128}$. All computations are performed by Matlab software package.

Example 3.1. Here we estimate c in (3.1) by assuming $f = 1 + t^2$ and $h_0 = h_1 = 1$. If $u(c^\dagger) = 1$ then the true solution is $c^\dagger = 1 + t^2$. In our computation, we use the first guess as $c_0 = 1 + t^2 - 2t(1 - t)(1 + t - t^2)$, and instead of $u(c^\dagger)$ we use the special perturbation $u_\delta = 1 + \delta\sqrt{2}\sin(10\pi t/\delta^2)$. Clearly $\|u_\delta - u(c^\dagger)\|_{L^2} \leqslant \delta$. It is easy to know that $c_0 - c^\dagger \in X_\mu$ for all $\mu < 2.5$.

When we use the iteratively regularized Gauss–Newton method in Hilbert scales, the choice of s plays certain roles. From theorem 2.1 it is easy to see that the best possible error bound to be expected is $\|c_{n_\delta} - c^\dagger\|_{L^2} \leqslant \mathrm{O}(\delta^\kappa)$ with $\kappa < \frac{5}{9}$ if we choose $s = 1$, while for $s = 0$ we have at most $\|c_{n_\delta} - c^\dagger\|_{L^2} \leqslant \mathrm{O}(\delta^{1/2})$ due to the saturation property. Table 1 reports the related numerical results by choosing $\alpha_n := 0.1 \times 0.5^{n-1}$ and $\tau = 2.5$ which satisfies the requirement in theorem 2.1.

Example 3.2. Here we again estimate the parameter c in (3.1) but with $f = 2 + (1 + t(1 - t))t(1 - t)$ and $h_0 = 1 = h_1$. If $u(c^\dagger) = 1 + t(1 - t)$, then the true solution is $c^\dagger = t(1 - t)$. In our calculation we used the special perturbation $u_\delta = 1 + t(1 - t) + \delta\sqrt{2}\sin(10\pi t/\delta^2)$. As the first guess we choose $c_0 = 0$. It is easy to see that $c_0 - c^\dagger \in X_\mu$ for all $\mu < 2.5$. Table 2 reports the numerical results by using the iteratively regularized Gauss–Newton method in Hilbert scales with $s = 1$, $\alpha_n = 0.1 \times 0.25^{n-1}$ and $\tau = 1$. According to theorem 2.1 we can obtain the error estimate in the norm $\|\cdot\|_r$ for all $r \in [-2, 2.5)$. Table 2 summarizes the related results in L^2-norm and H^1-norm. This example also shows that the discrepancy principle (2.19) works well for $\tau = 1$ even if it does not satisfy the requirement in theorem 2.1.

Acknowledgments

The author is grateful to the referees for their constructive suggestions which led to this substantially improved version. This work is supported by National Natural Science Foundation of China under grant 19801018.

References

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

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

[3] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht: Kluwer)

[4] Groetsch C W 1993 *Inverse Problems in Mathematical Sciences* (Wiesbaden: Viewig)

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

[6] Hanke M, Neubauer A and Scherzer O 1995 A convergence analysis of Landweber iteration of nonlinear ill-posed problems *Numer. Math.* **72** 21–37

[7] Hohage T 1999 Convergence rates of a regularized Newton method in sound-hard inverse scattering *SIAM J. Numer. Anal.* **36** 125–42

[8] Jin Q N 1999 On the iteratively regularized Gauss–Newton method for solving nonlinear ill-posed problems *Math. Comput.* at press

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

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

[11] Köhler J and Tautenhahn U 1995 Error bounds for regularized solutions of nonlinear ill-posed problems *J. Inv. Ill-Posed Problems* **3** 47–74

[12] Natterer F 1984 Error bounds for Tikhonov regularization in Hilbert scales *Appl. Anal.* **18** 29–37

[13] Neubauer A 1988 An *a posteriori* parameter choice for Tikhonov regularization in Hilbert scales leading to optimal convergence rates *SIAM J. Numer. Anal.* **25** 1313–26

[14] Neubauer A 1992 Tikhonov regularization of nonlinear ill-posed problems in Hilbert scales *Appl. Anal.* **46** 59–72

[15] Scherzer O 1996 A convergence analysis of a method of steepest descent and a two-step algorithm for nonlinear ill-posed problems *Numer. Funct. Anal. Optim.* **17** 197–214

[16] Tautenhahn U 1994 Error estimates for the regularized solutions of nonlinear ill-posed problems *Inverse Problems* **10** 485–500

[17] Tautenhahn U 1996 Error estimates for regularization methods in Hilbert scales *SIAM J. Numer. Anal.* **33** 2120–30

[18] Tautenhahn U 1998 On a general regularization scheme for nonlinear ill-posed problems II: Regularization in Hilbert scales *Inverse Problems* **14** 1607–16