Tikhonov regularization and \textit{a posteriori} rules for solving nonlinear ill posed problems

U Tautenhahn\textsuperscript{1} and Qi-nian Jin\textsuperscript{2}

\textsuperscript{1} Department of Mathematics, University of Applied Sciences Zittau/Görlitz, PO Box 1455, 02755 Zittau, Germany
\textsuperscript{2} Department of Mathematics, Rutgers University, New Brunswick, NJ 08903, USA

E-mail: u.tautenhahn@hs-zigr.de and qjin@math.rutgers.edu

Received 2 April 2002, in final form 24 October 2002
Published 15 November 2002
Online at stacks.iop.org/IP/19/1

Abstract

Besides \textit{a priori} parameter choice we study \textit{a posteriori} rules for choosing the regularization parameter $\alpha$ in the Tikhonov regularization method for solving nonlinear ill posed problems $F(x) = y$, namely a rule 1 of Scherzer \textit{et al} (Scherzer O, Engl H W and Kunisch K 1993 SIAM J. Numer. Anal. \textbf{30} 1796–838) and a new rule 2 which is a generalization of the monotone error rule of Tautenhahn and Hämärik (Tautenhahn U and Hämärik U 1999 Inverse Problems \textbf{15} 1487–505) to the nonlinear case. We suppose that instead of $y$ there are given noisy data $y^\delta$ satisfying $\|y - y^\delta\| \leq \delta$ with known noise level $\delta$ and prove that rule 1 and rule 2 yield order optimal convergence rates $O(\delta^{p/(p+1)})$ for the ranges $p \in (0, 2]$ and $p \in (0, 1]$, respectively. Compared with foregoing papers our order optimal convergence rate results have been obtained under much weaker assumptions which is important in engineering practice. Numerical experiments verify some of the theoretical results.

1. Introduction

In this paper we consider nonlinear ill posed problems

$$F(x) = y \quad (1.1)$$

where $F: D(F) \subset X \rightarrow Y$ is a nonlinear operator with non-closed range $R(F)$ and $X, Y$ are infinite dimensional real Hilbert spaces with corresponding inner products $(\cdot, \cdot)$ and norms $\| \cdot \|$, respectively. We are interested in the $\varpi$-minimum-norm solution $x^\dagger$ of problem (1.1), that is, $x^\dagger$ satisfies

$$F(x^\dagger) = y \quad \text{and} \quad \|x^\dagger - \varpi\| = \min_{x \in D(F)} \{\|x - \varpi\| \mid F(x) = y\}.$$  

We assume throughout this paper that $y^\delta \in Y$ are the available noisy data with

$$\|y - y^\delta\| \leq \delta \quad (1.2)$$
and known noise level $\delta$. We further assume that $F$ possesses a locally uniformly bounded Fréchet derivative $F'(\cdot)$ in a ball $B_r(x^\dagger)$ of radius $r$ around $x^\dagger \in X$.

For the stable numerical solution of nonlinear ill posed problems regularization methods are necessary. While for linear ill posed problems the regularization theory is rather complete (cf [2–4, 10, 12, 15, 22, 33, 35]), there are still many open problems in the regularization theory for nonlinear ill posed problems. For nonlinear ill posed problems, Tikhonov regularization (cf [4, 5, 9, 16, 17, 19–21, 27, 29, 30]) is known as one of the most widely applied methods. In this method a regularized approximation $x^\alpha_\delta$ is obtained by solving the minimization problem

$$
\min_{x \in D(F)} J_\delta(x), \quad J_\delta(x) = \|F(x) - y^\dagger\|^2 + \alpha\|x - \overline{x}\|^2
$$

(1.3)

with an initial guess $\overline{x} \in X$ and a properly chosen regularization parameter $\alpha > 0$. If $x^\alpha_\delta$ is an interior point of $D(F)$, then the regularized approximation $x^\alpha_\delta$ satisfies the Euler equation

$$
F'(x)^* [F(x) - y^\dagger] + \alpha(x - \overline{x}) = 0
$$

(1.4)

of Tikhonov’s functional $J_\delta(x)$. In formula (1.4) $F'(x)^*$ is the adjoint of the Fréchet derivative $F'(x)$.

One important question in the application of the regularization method (1.3) is the proper choice of the regularization parameter $\alpha > 0$. With too little regularization, the regularized approximations $x^\alpha_\delta$ are highly oscillatory due to noise amplification. With too much regularization, the regularized approximations are too close to the initial guess $\overline{x}$ due to the limit relation $\lim_{\alpha \to \infty} x^\alpha_\delta = \overline{x}$. Ideally, one should select the regularization parameter $\alpha$ such that the total error $\|x^\alpha_\delta - x^\dagger\|$ is minimized. Since $x^\dagger$ is unknown one has to choose alternative rules which choose $\alpha$ from quantities that arise during calculations. Although many rules have been proposed, very few of them are used in practice. Among the rules that have found their way into engineering practice, the most common are Morozov’s discrepancy principle [23, 34] which requires the knowledge of a reliable lower bound of the noise level $\delta$, and in the linear case some noise-free rules such as generalized cross validation [8, 36] and the L-curve method [13, 14]. On the other hand, noise-free rules occasionally fail. This can theoretically be justified since due to Bakushinskii [1] there is a negative result which tells us that for noise free rules convergence $x^\alpha_\delta \to x^\dagger$ cannot be guaranteed. Besides Morozov’s discrepancy principle, one prominent a posteriori rule that requires the knowledge of the noise level $\delta$ is the rule of Scherzer, Engl and Kunisch [27], which is an extension of the rule of Raus and Gruber [7, 25] to the nonlinear case (see rule 1 in section 3.1). However, this theoretically well justified method requires very strong assumptions which can seldom be satisfied in engineering problems, see section 3.1 for some more detailed discussion.

In this paper we mainly reconsider rule 1 and give, compared with foregoing papers [17–19, 27], theoretical justifications of this rule under weaker assumptions which are generally met in engineering practice. In addition, we study a second a posteriori rule (see rule 2 in section 3.1) which is a generalization of the monotone error rule (see [31, 32]) to the nonlinear case. Since rule 2 outperforms rule 1 in the case of linear ill posed problems (see [32]) it makes sense to study rule 2 also for nonlinear ill posed problems.

Some important properties of convergent a posteriori rules are order optimal Hölder type error bounds $\|x^\alpha_\delta - x^\dagger\| = O(\delta^{p/(p+1)})$. For the concept of order optimality see, e.g., [5]. The proof of order optimal error bounds requires a source condition and a nonlinearity condition. We will separate our studies into different cases $p = 1$, $p \in [1, 2]$ and $p \in (0, 1]$. In the first two cases $p = 1$ and $p \in [1, 2]$ our analysis requires the following two assumptions.

**Assumption 1.** There exist elements $v \in Y$ and $w \in Y$ such that with $A = F'(x^\dagger)

$$
\overline{x} - x^\dagger = A^*v \quad \text{and} \quad \overline{x} - x^\dagger = A^*(AA^*)^{(p-1)/2}w \quad \text{for} \; p \geq 1.
$$
Assumption 2. The Fréchet derivative $F'(\cdot)$ is locally Lipschitz in a ball $B_r(x^\dagger) \subset D(F)$ of radius $r$ around $x^\dagger \in X$, that is, there exists a Lipschitz constant $L \geq 0$ with

$$\|F'(x) - F'(x_0)\| \leq L \|x - x_0\| \quad \text{for all } x, x_0 \in B_r(x^\dagger).$$

In the rough case $p \in (0, 1]$ we do not know if Lipschitz continuity of the Fréchet derivative $F'(\cdot)$ is sufficient for guaranteeing order optimal error bounds. In this case instead of assumption 1 and assumption 2 we will exploit the following two assumptions.

Assumption 3. There exists an element $w \in X$ such that with $A = F'(x^\dagger)$

$$\xi - x^\dagger = (A^*A)^{p/2} w \quad \text{for } p > 0.$$ 

Assumption 4. There exists a constant $k_0 \geq 0$ such that for all $x, x_0 \in B_r(x^\dagger) \subset D(F)$ and $v \in X$ there exists an element $k(x, x_0, v) \in X$ with property

$$[F'(x) - F'(x_0)]v = F'(x_0)k(x, x_0, v) \quad \text{and} \quad \|k(x, x_0, v)\| \leq k_0\|x - x_0\|\|v\|.$$ 

Note that the converse results in [4] for linear ill posed problems imply that assumption 1 or assumption 3, respectively, is necessary for the convergence rate $\sqrt[2p]{x^\dagger - x^0} = O(\delta^{p/(p+1)})$. However, the verification of assumption 1 or assumption 3, respectively, is very hard or even impossible for the majority of applied nonlinear ill posed problems. Since the operator $F'(x^\dagger)$ is generally smoothing, these assumptions may be considered as abstract smoothness conditions concerning the (unknown) difference element $\xi - x^\dagger$ and become stronger for larger $p$-values.

Further note that from assumption 4 there follows assumption 2 with $L = k_0\|F'(x_0)\|$. Hence, assumption 2 is weaker than assumption 4. An example of a nonlinear ill posed problem satisfying assumption 2 but not assumption 4 is the autoconvolution problem discussed in [9].

Let us collect two well known estimates which will be applied throughout this paper. The first estimate

$$\|F(x) - F(x_0) - F'(x_0)(x - x_0)\| \leq \frac{L}{2}\|x - x_0\|^2 \quad (1.5)$$

follows from assumption 2. The second estimate (see, e.g., [4]) follows from spectral theory and tells us that for linear operators $A \in L(X, Y)$ and $v \in [0, 1]$

$$\|(AA^* + \alpha I)^{-1}(AA^*)^v\| \leq c_\alpha^{v^{-1}} \quad (1.6)$$

with a constant $c_\alpha$ satisfying $c_\alpha \leq v\alpha(1 - \alpha)^{-v-1} \leq 1$.

The paper is organized as follows. In section 2 we study the case of a priori parameter choice. These studies are mainly based on functional techniques and are essential for the study of the a posteriori rule 1 and rule 2 in section 3. Both sections 2 and 3 are divided into subsections in which, for technical reasons, the special cases $p = 1$, $p \in [1, 2]$ and $p \in (0, 1]$, respectively, are treated separately. The order optimal error bounds in sections 2 and 3 reduce to well known error bounds for linear ill posed problems provided $L = 0$ in assumption 2 or $k_0 = 0$ in assumption 4. Finally, in section 4 we provide numerical experiments that illustrate some of the theoretical results.

2. A priori parameter choice

In this section we prove some new error bounds for the method of Tikhonov regularization in the case of a priori parameter choice. These estimates are mainly based on functional techniques and will be exploited in studying a posteriori parameter choice in section 3. Let us start with a well known preliminary result which follows from the minimizing property of the regularized approximation $x^\sigma$ of problem (1.3).
Proposition 2.1. Let \( x^\delta_a \) be a minimizer of Tikhonov’s functional (1.3). Then,
\[
\| x^\delta_a - \overline{x} \| \leq \frac{\delta}{\sqrt{\alpha}} + \| x^\top - \overline{x} \| \quad \text{and} \quad \| x^\delta_a - x^\top \| \leq \frac{\delta}{\sqrt{\alpha}} + 2 \| \overline{x} - x^\top \|. \tag{2.1}
\]

Estimate (2.1) shows that \( x^\delta_a \in B_r (x^\top) \) with radius \( r = \frac{\delta}{\sqrt{\alpha}} + 2 \| \overline{x} - x^\top \| \). In order to guarantee order optimal error bounds of Hölder type \( \| x^\delta_a - x^\top \| = O(\delta^{p/(p+1)}) \) a source condition and a nonlinearity condition are needed. In our first subsection we will consider the special case \( p = 1 \) in assumption 1.

2.1. Error bounds in the case \( p = 1 \)

In this subsection we treat the special case \( p = 1 \) in assumption 1. This case has been studied first in [5]. From [4] we have

**Theorem 2.2.** Let \( x^\delta_a \in D(F) \) be a solution of problem (1.3). Assume assumption 1 with \( p = 1 \) and assumption 2 with radius \( r = 0 \). If \( L \| v \| \leq 1 \), then
\[
\| x^\delta_a - x^\top \| \leq \frac{\delta + \alpha \| v \|}{\sqrt{\alpha} \sqrt{1 - L \| v \|}} \quad \text{and} \quad \| F(x^\delta_a) - y \| \leq \delta + 2\alpha \| v \|. \tag{2.2}
\]

If \( \alpha \) is chosen a priori by \( \alpha \sim \delta \), then \( \| x^\delta_a - x^\top \| = O(\delta^{1/2}) \).

In the next two theorems we treat the two error terms \( \| x_a - x^\top \| \) and \( \| x^\delta_a - x_a \| \) separately. These estimates are useful for a serious study of a posteriori parameter choice in section 3. In addition, our error bound for \( \| x^\delta_a - x_a \| \) in theorem 2.4 improves a corresponding stability estimate which has been derived in [26].

**Theorem 2.3.** Let \( x_a \) be a solution of problem (1.3) with \( y^\delta \) replaced by the exact data \( y \). Assume assumption 1 with \( p = 1 \) and assumption 2 with radius \( r = 2 \| \overline{x} - x^\top \| \). If \( 3L \| v \| \leq 2 \), then
\[
\| x_a - x^\top \| \leq \frac{\sqrt{\| v \|}}{2 \sqrt{1 - L \| v \|}} \quad \text{and} \quad \| F(x_a) - y \| \leq \alpha \| v \|. \tag{2.3}
\]

**Proof.** Let \( A_a = F'(x_a) \). Using the Euler equation \( A^*_a [F(x_a) - y] + \alpha (x_a - \overline{x}) = 0 \) yields
\[
\alpha \| x_a - x^\top \|^2 = \alpha (x_a - x^\top, \overline{x} - x^\top) + \alpha (x_a - x^\top, x_a - \overline{x}) \\
= \alpha (x_a - x^\top, \overline{x} - x^\top) + (A_a (x_a - x^\top), F(x_a) - y).
\]

We add \( \| F(x_a) - y \| \) on both sides, use the representation \( \overline{x} - x^\top = A^* v \) of assumption 1, exploit in addition (1.5) and obtain
\[
\| F(x_a) - y \|^2 + \alpha \| x_a - x^\top \|^2 = \alpha (y + A(x_a - x^\top) - F(x_a), v) + \alpha (F(x_a) - y, v) \\
+ (F(x_a) + A_a (x_a - x_a) - y, F(x_a) - y) \\
\leq \alpha \frac{L \| v \|}{2} \| x_a - x^\top \|^2 + \alpha \| v \| \| F(x_a) - y \| + \frac{L}{2} \| x_a - x^\top \|^2 \| F(x_a) - y \|. \tag{2.4}
\]

From (2.2) with \( \delta = 0 \) we know that \( \| F(x_a) - y \| \leq 2\alpha \| v \| \). Consequently,
\[
\| F(x_a) - y \|^2 + \alpha \| x_a - x^\top \|^2 \leq \alpha \frac{3L \| v \|}{2} \| x_a - x^\top \|^2 + \alpha \| v \| \| F(x_a) - y \|. \tag{2.5}
\]

We use the assumption \( 3L \| v \| \leq 2 \) and obtain \( \| F(x_a) - y \|^2 \leq \alpha \| v \| \| F(x_a) - y \| \) which provides the second estimate of (2.3). To prove the first estimate of (2.3) we use (2.4) and the second estimate of (2.3) as well as the estimate \( 2ab \leq a^2 + b^2 \) and obtain
\[
\| F(x_a) - y \|^2 + \alpha \| x_a - x^\top \|^2 \leq \alpha L \| v \| \| x_a - x^\top \|^2 + \alpha \| v \| \| F(x_a) - y \|^2 \\
\leq \alpha L \| v \| \| x_a - x^\top \|^2 + \frac{\alpha^2 \| v \|^2}{4} + \| F(x_a) - y \|^2
\]

which gives \( \alpha (1 - L \| v \|) \| x_a - x^\top \|^2 \leq \frac{\alpha^2 \| v \|^2}{4} \), and hence the first estimate of (2.3). \( \square \)
In this subsection we shall provide a new order optimal error bound for the total error $\delta$. We have

$$\|x_d^\delta - x_0\| \leq \frac{\delta}{\sqrt{\alpha + 2 \|F - x\|}} \quad \text{and} \quad \|F(x_d^\delta) - y^\delta + y - F(x_0)\| \leq \delta. \quad (2.6)$$

**Proof.** We use the inequality $J_\alpha(x_d^\delta) \leq J_\alpha(x_0)$ and obtain

$$\|F(x_d^\delta) - y^\delta\|^2 + \alpha\|x_d^\delta - x\|^2 \leq \|F(x_0) - y^\delta\|^2 + \alpha\|x_0 - x\|^2.$$

We add on both sides the expression

$$2(F(x_d^\delta) - y^\delta, y - F(x_0)) + \|F(x_0) - y\|^2 + \alpha(F(x_0) - x\|^2 + \|x_0 - x\|^2)$$

and obtain by using the Euler equation $F'(x_0)^* (F(x_0) - y) + \alpha(x_0 - x) = 0$ and (1.5)

$$\|F(x_d^\delta) - y^\delta + y - F(x_0)\|^2 + \alpha\|x_d^\delta - x_0\|^2 \leq \|F(x_0) - y^\delta\|^2 + \|F(x_0) - y\|^2 + 2(F(x_d^\delta) - y^\delta, y - F(x_0)) + 2\alpha(F_\alpha - x, x_0 - x_0)$$

$$= \|y - y^\delta\|^2 + 2(F(x_0) - y, F(x_0) - F(x_d^\delta)) + 2\alpha(x_0 - x, x_0 - x_0)$$

$$= \|y - y^\delta\|^2 + 2(F(x_0) - y, F(x_0) + F'(x_0)(x_d^\delta - x_0) - F(x_d^\delta))$$

$$\leq \delta^2 + L\|F(x_0) - y\|\|x_d^\delta - x_0\|^2. \quad (2.7)$$

From (2.7) and the second estimate of (2.3) we obtain

$$\alpha(1 - L\|v\|)\|x_d^\delta - x_0\|^2 \leq \delta^2$$

which gives the first estimate of (2.6). The second estimate of (2.6) follows from (2.7), the second estimate of (2.3) and $L\|v\| \leq 1$. 

### 2.2. Error bounds in the case $p \in [1, 2]$}

In this subsection we shall provide a new order optimal error bound for the total error $\|x_d^\delta - x^\dagger\|$ in the case of assumption 1 with $p \in [1, 2]$. Error bounds for the range $p \in [1, 2]$ were studied first in [24]. The proof in [24] uses functional techniques and exploits, as theorem 2.2, the inequality $J_\alpha(x_d^\delta) \leq J_\alpha(x^\dagger)$. This in turn requires quite strong assumptions. Our idea of proof requires less strong assumptions and is based on the inequality $J_\alpha(x_d^\delta) \leq J_\alpha(x^\dagger + \alpha Bv)$ with $B = A^*(AA^* + \alpha I)^{-1}$ and $A = F'(x^\dagger)$. The resulting estimate improves a corresponding bound given in [30] and will be exploited for providing error bounds in the case of a posteriori parameter choice in section 3.

**Theorem 2.5.** Let $x_d^\delta \in D(F)$ be a solution of problem (1.3). Assume assumption 1 with fixed $p \in [1, 2]$ and assumption 2 with radius $r = \delta/\sqrt{\alpha + 2 \|F - x\|}$. If $L\|v\| < 1$, then

$$\|x_d^\delta - x^\dagger\| \leq \frac{\delta}{\sqrt{\alpha + 2 \|F - x\|}} + \alpha^{p/2}\|v\| \frac{1 + L\|v\|^2/ \sqrt{1 - L\|v\|^2}}. \quad (2.8)$$

If $\alpha$ is chosen a priori by $\alpha \sim \delta^{2/(p+1)}$, then $\|x_d^\delta - x^\dagger\| = O(\delta^{2/(p+1)}).

**Proof.** Let us introduce the notations $A = F'(x^\dagger)$ and $B = A^*(AA^* + \alpha I)^{-1} = (A^*A + \alpha I)^{-1}A^*$. Since $\|\alpha Bv\| \leq \|\|F - x\|\|\|$, we have $x^\dagger + \alpha Bv \in D(F)$. Consequently, $J_\alpha(x^\dagger + \alpha Bv)$ is well defined and the inequality $J_\alpha(x_d^\delta) \leq J_\alpha(x^\dagger + \alpha Bv)$ provides

$$\alpha\|x_d^\delta - x^\dagger\|^2 \leq \|y^\delta - F(x^\dagger) - \alpha ABv + \alpha v\|^2 + \alpha\|Bv\|^2$$

$$+ 2\alpha(F(x^\dagger) + A(x_d^\delta - x^\dagger) - F(x_d^\delta), v)$$

$$+ \|F(x^\dagger + \alpha Bv) - y^\delta\|^2 - \|F(x^\dagger) + \alpha ABv - y^\delta\|^2. \quad (2.9)$$
(cf estimate (9) in [30]). From the estimates (10) and (11) in [30] we obtain for the first three summands $s_1 + s_2 + s_3$ on the right-hand side of (2.9)

$$s_1 + s_2 + s_3 \leq (\alpha^{(p+1)/2}\|u\| + \delta)^2 + \alpha L\|v\|\|x_\alpha - x^\dagger\|^2. \quad (2.10)$$

We apply (1.5) with $x = x^\dagger + \alpha Bu$, the inequality $\|u\|^2 - \|v\|^2 \leq \|u + v\||u - v||$, the triangle inequality and (1.2) as well as assumption 1 and obtain for the final two summands $s_4 + s_5$ on the right-hand side of (2.9) the estimate

$$s_4 + s_5 \leq \|F(x^\dagger + \alpha Bu) + \alpha ABv + F(x^\dagger) - 2y^\delta\||F(x^\dagger + \alpha Bu) - \alpha ABv - F(x^\dagger)||$$

$$\leq (\|F(x^\dagger + \alpha Bu) - \alpha ABv - F(x^\dagger)\| + 2\|\alpha ABv + F(x^\dagger) - y^\delta\|) \frac{L}{2}\|\alpha Bu\|^2$$

$$\leq \left\{ \frac{L}{2}\|\alpha Bu\|^2 + 2\alpha\|v\| + 2\delta \right\} \frac{L}{2}\|\alpha Bu\|^2$$

where we have used the estimate $\|\alpha ABv\| \leq \|v\|$. Now we apply the two valid estimates $\|\alpha Be\|^2 \leq \alpha^{(p+1)/2}\|v\||w||$ and $\|\alpha Be\| \leq \alpha^{p/2}\|w\|$ that follow from assumption 1, (1.6) and obtain

$$s_4 + s_5 \leq \frac{L^2}{4} \alpha^{p+1}\|v\|^2\|w\|^2 + L\alpha^{p+1}\|v\|\|w\|^2 + L\delta\alpha^{(p+1)/2}\|v||w||. \quad (2.11)$$

Substituting (2.10), (2.11) into (2.9) provides

$$\alpha(1 - L\|v\|)\|x_\alpha - x^\dagger\|^2 \leq (\delta + \alpha^{(p+1)/2}\|w\|(1 + L\|v\|/2))^2.$$

From this estimate we obtain (2.8). The convergence rate result follows from (2.8) together with the a priori parameter choice of $\alpha$.

\[\square\]

2.3. Error bounds in the case $p \in (0, 1)$

The theorems in sections 2.1 and 2.2 treat the smooth case $p \in [1, 2]$ in assumption 1. In the rough case $p \in (0, 1)$ in assumption 1 or assumption 3, respectively, we do not know if order optimal error bounds hold true with the nonlinearity assumption 2. Nonlinearity assumptions of a different kind which are stronger than assumption 2 have been exploited in [16, 18, 20]. Here we exploit a nonlinearity assumption 4 from [18] which allows us to treat not only the case $p \in (0, 1]$, but even the general case $p \in (0, 2]$. The resulting error bounds will be exploited in studying the case of a posteriori parameter choice in section 3.4. Our first theorem provides an error bound for $\|x_\alpha - x^\dagger\|$ and improves a corresponding result in [18].

**Theorem 2.6.** Let $x_\alpha$ be a solution of the regularized problem (1.3) with $y^\delta$ replaced by the exact data $y$. Assume assumption 3 with $p \in (0, 2]$ and assumption 4 with radius $r = 2\|\alpha - x^\dagger\|$. If $2k_\alpha\|\alpha - x^\dagger\| < 1$, then

$$\|x_\alpha - x^\dagger\| \leq \frac{\alpha^{p/2}\|w\|}{1 - 2k_\alpha\|\alpha - x^\dagger\|}. \quad (2.12)$$

**Proof.** Let us use the notations $A = F'(x^\dagger)$ and $A_\alpha = F'(x_\alpha)$. We apply the Euler equation $A_\alpha^* [F(x_\alpha) - y] + \alpha(x_\alpha - x^\dagger) = 0$ and obtain

$$(A^* A + \alpha I)(x_\alpha - x^\dagger) = A^* A(x_\alpha - x^\dagger) + A_\alpha^* [y - F(x_\alpha)] + \alpha(x^\dagger - x^\alpha)$$

$$= \alpha(x^\dagger - x^\alpha) + (A^* A - A_\alpha^*)[y - F(x_\alpha)] + A^*[y + A(x_\alpha - x^\dagger) - F(x_\alpha)].$$

Multiplying both sides by $(A^* A + \alpha I)^{-1}$ yields

$$x_\alpha - x^\dagger = \alpha(A^* A + \alpha I)^{-1}(x^\dagger - x^\alpha) + (A^* A + \alpha I)^{-1}(A_\alpha^* - A^*)[y - F(x_\alpha)] + (A^* A + \alpha I)^{-1}A^*[y + A(x_\alpha - x^\dagger) - F(x_\alpha)]. \quad (2.13)$$
To estimate the first summand $s_1$ on the right-hand side of (2.13) we use assumption 3 as well as estimate (1.6) and obtain
\[ \|s_1\| = \|\alpha (A^*A + \alpha I)^{-1} (A^*A)^{1/2} w\| \leq \alpha^{1/2} \|w\|. \]  
(2.14)

To estimate the second summand $s_2$ on the right-hand side of (2.13) we use assumption 4 and the Euler equation $A^*w[F(x^\delta) - y] + \alpha (x^\delta - \overline{x}) = 0$ as well as the estimate $\|\overline{x} - x^\delta\| \leq \|\overline{x} - x^d\|$ which follows from proposition 2.1 with $\delta = 0$ and obtain along the lines of estimate (A.7) in [19]
\[ \|s_2\| \leq k_0\|\overline{x} - x^d\|\|x^\delta - x^\dagger\|. \]  
(2.15)

From the mean value theorem and assumption 4 we obtain
\[ F(x^\delta) - F(x_1) = \int_0^1 [F'(x_1 + t(x_2 - x_1)) - F'(x_1)](x_2 - x_1) \, dt \]
\[ = F'(x_1) \int_0^1 \delta(x_2 - x_1, x_2 - x_1) \, dt. \]  
(2.16)

We use the representation (2.16) with $x_1 = x^d$, $x_2 = x^\delta$ and obtain for the third summand on the right-hand side of (2.13)
\[ \|s_3\| \leq \|(A^*A + \alpha I)^{-1} A^*A\| \int_0^1 \|\delta(x^d + t(x^\delta - x^d), x^\delta - x^d)\| \, dt \]
\[ \leq \frac{k_0}{2}\|x^\delta - x^d\|^2 \leq k_0\|\overline{x} - x^d\|\|x^\delta - x^d\|. \]  
(2.17)

Now (2.12) follows from (2.13)–(2.15) and (2.17).

Our next theorem provides an error bound for the error $\|x^\delta - x^\alpha\|$. In contrast to the error bounds given in theorem 2.4 no source condition is required. This theorem improves stability results which have been obtained in the papers [27] and [19].

**Theorem 2.7.** Let $x^\delta$ be a solution of problem (1.3) and $x^\alpha$ a solution of (1.3) with $y^\delta$ replaced by the exact data $y$. Assume assumption 4 with radius $r = \delta/\sqrt{\alpha + 2\|\overline{x} - x^d\|}$. If $k_0\|x^d - \overline{x}\| < 1$, then
\[ \|x^\delta - x^\alpha\| \leq \frac{\delta}{\sqrt{\alpha} \sqrt{1 - k_0\|x^d - \overline{x}\|}} \quad \text{and} \quad \|F(x^\delta) - y^\delta + y - F(x^\alpha)\| \leq \delta. \]  
(2.18)

**Proof.** Using the first part of (2.7), the representation (2.16), the Euler equation (1.4) with $\delta = 0$ and assumption 4 as well as the estimate $\|x^\delta - \overline{x}\| \leq \|x^d - \overline{x}\|$ yields
\[ \|F(x^\delta) - y^\delta + y - F(x^\alpha)\|^2 \leq \delta^2 + 2\|F(x^\alpha) - y, F(x^\alpha) + F(x^\delta)\|^2 \]
\[ = \delta^2 + 2\left(\|F(x^\alpha) - y, F(x^\alpha)\| \int_0^1 \|k(x^\alpha + t(x^\delta - x^\alpha), x^\delta - x^\alpha)\| \, dt\right) \]
\[ = \delta^2 + 2\left(\alpha\|\overline{x} - x^\alpha\|, \int_0^1 \|k(x^\alpha + t(x^\delta - x^\alpha), x^\delta - x^\alpha)\| \, dt\right) \]
\[ \leq \delta^2 + 2\alpha\|x^\delta - x^\alpha\| \int_0^1 \|k(x^\alpha + t(x^\delta - x^\alpha), x^\delta - x^\alpha)\| \, dt \]
\[ \leq \delta^2 + 2\alpha k_0\|x^d - \overline{x}\|\|x^\delta - x^\alpha\|^2. \]  
(2.19)

We neglect the first summand on the left-hand side, rearrange terms and obtain the estimate $\alpha(1 - k_0\|x^d - \overline{x}\|)\|x^\delta - x^\alpha\|^2 \leq \delta^2$ which gives the first inequality of (2.18). The second inequality of (2.18) follows from (2.19) and $k_0\|x^d - \overline{x}\| < 1$. 
\[ \square \]
3. A posteriori parameter choice

3.1. A posteriori rules

Throughout this section we will consistently use the notations

\[
\begin{align*}
A & = F'(x^\dagger), & R &= \alpha(AA^* + \alpha I)^{-1}, & \hat{R} &= \alpha(A^*A + \alpha I)^{-1}, \\
A_a & = F'(x_a), & R_a &= \alpha(A_aA_a^* + \alpha I)^{-1}, & \hat{R}_a &= \alpha(A_a^*A_a + \alpha I)^{-1}, \\
A_0^\delta & = F'(x_0^\delta), & R_0^\delta &= \alpha((A_0^\delta)^* + \alpha I)^{-1}, & \hat{R}_0^\delta &= \alpha(((A_0^\delta)^*)^* + \alpha I)^{-1}.
\end{align*}
\]

A priori parameter choice is not suitable in practice since a good regularization parameter \(\alpha\) requires the knowledge of the norm \(\|w\|\) and the smoothness parameter \(p\) of assumption 1 or assumption 3, respectively. This knowledge is not necessary for a posteriori parameter choice. For the method of Tikhonov regularization one well known a posteriori rule is Morozov’s discrepancy principle (see [4, 23, 34]) in which the regularization parameter \(\alpha\) is chosen as the solution of the nonlinear scalar equation \(\|F(x_0^\delta) - y^\delta\| = C\delta\) with a constant \(C > 1\). Morozov’s discrepancy principle possesses some nice order optimality properties discussed in [34]. However, the best possible convergence rate is known to be \(\|x_0^\delta - x^\dagger\| = O(\sqrt{\delta})\). An implementable a posteriori rule for choosing the regularization parameter \(\alpha\) in the method of Tikhonov regularization (1.3) for which order optimal convergence rates up to the best possible order \(O(\delta^{2/3})\) can be guaranteed has been studied in [17, 19, 27].

**Rule 1.** Choose the regularization parameter \(\alpha\) as the solution of the equation

\[
d_1(\alpha) := \|(R_0^\delta)^{1/2}[F(x_0^\delta) - y^\delta]\| = C\delta. \tag{3.1}
\]

A similar a posteriori rule which is based on monotonicity arguments has been proposed in [32].

**Rule 2.** Choose the regularization parameter \(\alpha\) as the solution of the equation

\[
d_2(\alpha) := \frac{\|(R_0^\delta)^{1/2}[F(x_0^\delta) - y^\delta]\|^2}{\|R_0^\delta[F(x_0^\delta) - y^\delta]\|} = C\delta. \tag{3.2}
\]

For ill posed problems with linear operators \(F\) it is well known that rule 2 always provides a more accurate regularized solution than rule 1 (see [31]). Therefore it also makes sense to study this rule in the case of nonlinear ill posed problems.

Rule 1 has been studied in different papers. In [27] it has been shown that under assumption 3 and some further strong nonlinearity conditions concerning \(F\) (compare the assumptions (10)–(14), (93)–(98) and (151) in [27]) order optimal error bounds \(\|x_0^\delta - x^\dagger\| = O(\delta^{p/(p+1)})\) can be guaranteed for the maximal range \(p \in (0, 2]\). The strong nonlinearity conditions in [27] concerning \(F\) have been weakened in [19] for the range \(p \in (0, 2]\) where assumption 4 is exploited, and in [17] for \(p = 2\) where assumption 2 is exploited.

In our forthcoming sections we consider the range \(p \in (0, 2]\) and divide our studies into the special cases \(p = 1\), \(p \in [1, 2]\) and \(p \in (0, 1]\). Compared with foregoing papers [17, 19, 27] we obtain order optimal error bounds under much weaker assumptions. We found the most important improvements for the special cases \(p = 1\) and \(p \in (1, 2]\). In these cases the nonlinearity assumption 4 is not necessary and assumption 2 is sufficient. Moreover, also for the special case \(p \in (0, 1]\), in which the nonlinearity assumption 4 is exploited, we have found improvements of former results. Our order optimal error bounds are sharper and are valid for arbitrary constants \(C > 1\) in both rule 1 and rule 2 (provided \(k_0\|x^\dagger - x^\|\) is sufficiently small), while in foregoing papers order optimal error bounds for rule 1 could only be established for large \(C\)-values.
Let us start our study with the justification of rule 1 and rule 2. In [17] conditions are given that guarantee that equation (3.1) has a solution \( \alpha = \alpha(\delta) \). In an analogous way it can be shown that under the same conditions equation (3.2) also has a solution \( \alpha = \alpha(\delta) \).

**Proposition 3.1.** Let \( C > 1 \) and \( \|F(\mathbf{x}) - y^\delta\| > C\delta \). If the regularized problem (1.3) has a unique solution \( x^\delta \) for each \( \alpha \geq \alpha_0 := (C - 1)^2 \frac{\|\mathbf{x} - x^\dagger\|^2}{\delta^2} \), then both equations (3.1) and (3.2) possess a solution \( \alpha = \alpha_1(\delta) \) and \( \alpha = \alpha_2(\delta) \), respectively, and both solutions satisfy (3.3).

The proof of this proposition for rule 1 may be found in [17] and is based on the facts that \( d_1(\alpha) \) is continuous for \( \alpha \in [\alpha_0, \infty) \), that \( \lim_{\alpha \to \infty} d_1(\alpha) = \|F(\mathbf{x}) - y^\delta\| \) and that

\[
d_1(\alpha_0) = \|F(x^\delta_0) - y^\delta\| \leq \delta + \sqrt{\alpha_0\|\mathbf{x} - x^\dagger\|^2} = C\delta
\]

which follows from the estimates \( \|(R^\delta_0)^{1/2}\| \leq 1 \), \( \|F(x^\delta_0) - y^\delta\|^2 \leq \delta^2 + \alpha\|\mathbf{x} - x^\dagger\|^2 \) and the definition of \( \alpha_0 \). For rule 2 the proof of proposition 3.1 can be performed analogously.

In our forthcoming considerations we always assume that rule 1 and rule 2 are well defined and do not state the conditions explicitly.

### 3.2. Error bounds in the case \( p = 1 \)

In order to prove order optimal error bounds for \( \|x^\delta - x^\dagger\| \) with \( \alpha \) chosen from rule 1 or rule 2, respectively, a preparatory proposition is required. This proposition gives a lower bound for the regularization parameters \( \alpha = \alpha(\delta) \) obtained by rule 1 and rule 2 which is sharper than the bound (3.3) in proposition 3.1.

**Proposition 3.2.** Let \( \alpha \) be chosen by rule 1 or rule 2 with \( C > 1 \). Let assumption 1 with \( p = 1 \) and assumption 2 with radius \( r = \delta/\sqrt{\alpha} + 2\|\mathbf{x} - x^\dagger\| \) hold. If \( L\|v\| < 1 \), then

\[
\alpha \geq \frac{C - 1}{\|v\|} \delta.
\]  

**Proof.** From rule 1, \( \|(R^\delta_0)^{1/2}\| \leq 1 \), the second inequality of (2.6) and the second inequality of (2.3) we have

\[
C\delta \geq \|(R^\delta_0)^{1/2}[F(x^\delta_0) - y^\delta]\| \leq \|F(x^\delta_0) - y^\delta\| \leq \delta + \alpha\|v\|
\]

which gives (3.4) for rule 1. In order to prove (3.4) for rule 2 we use the estimate

\[
C\delta \geq \frac{\|(R^\delta_0)^{1/2}[F(x^\delta_0) - y^\delta]\|^2}{\|R^\delta_0[F(x^\delta_0) - y^\delta]\|} \leq \|F(x^\delta_0) - y^\delta\|
\]

and proceed along the lines of the proof for rule 1. \( \square \)

The next theorem shows that order optimal error bounds for \( \|x^\delta - x^\dagger\| \) are valid under assumption 1 with \( p = 1 \) and assumption 2 provided the regularization parameter \( \alpha \) is chosen \textit{a posteriori} by rule 1 or rule 2 with \( C > 1 \), respectively. The proof is based on some ideas in [32] where order optimal error bounds have been established for rule 1 and rule 2 in the case of linear ill posed operator equations.
Theorem 3.3. Let α be chosen by rule 1 or rule 2 with C > 1. Let assumption 1 with p = 1 and assumption 2 with radius r = δ/√α + 2∥x − x∥ hold. Suppose L∥v∥ is sufficiently small that 3L∥v∥ ≤ 2 and

\[ \varepsilon_1 := \sqrt{\frac{L∥v∥}{2}} + L∩v∥ + \frac{L∥v∥}{8\sqrt{1−L∩v∥}} C + 1 \leq 1 \] (3.5)

hold. Then,

\[ \|x_x - x_x^p\| \leq \frac{1}{1−\varepsilon_1} \left((C + 1)^{1/2} + \frac{1}{2}(C − 1)^{-1/2}\right)∥v∥^{1/2}δ^{1/2}. \] (3.6)

Proof. First, let us prove the theorem for rule 1. We apply the Euler equation (1.4) with \( x = x_a^δ \) and obtain

\[ ((A_a^δ)^* A_a^δ + αI)(x_a^δ − x) = (A_a^δ)^* A_a^δ(x_a^δ − x) + α(x_a^δ − x) \]

\[ = (A_a^δ)^* A_a^δ(x_a^δ − x) + (A_a^δ)^*[y^δ - F(x_a^δ)] + α(x_a^δ − x) \]

Multiplying both sides by \(((A_a^δ)^* A_a^δ + αI)^{-1}\) yields the error representation

\[ x_a^δ − x^p = (A_a^δ)^* A_a^δ(x_a^δ − x) + (A_a^δ)^*[y^δ - F(x_a^δ)] − (A_a^δ)^*[y^δ - F(x_a^δ)]. \] (3.7)

Due to (3.7), (1.6), (1.2) and (1.5), the first estimates of (2.3), (2.6) and the estimate \( δ/α ≤ 1/∥v∥/(C − 1) \) which follows from (3.4) we obtain

\[ \|x_a^δ - x^p\| ≤ \|\hat{R}_a^δ(x_a^δ - x)\| + \frac{1}{2\sqrt{α}}\left(\frac{δ + L ∥v∥}{2}∥x_a^δ - x^p∥^2\right) \]

\[ ≤ \|\hat{R}_a^δ(x_a^δ - x)\| + \frac{1}{2\sqrt{α}}\left(\frac{δ + L ∥v∥}{\sqrt{1−L∩v∥}}∥x_a^δ - x^p∥\right) \]

\[ ≤ \|\hat{R}_a^δ(x_a^δ - x)\| + \frac{1}{2\sqrt{C − 1}}\left(\frac{L ∥v∥}{C + 1}\frac{C + 1}{C + 1}∥x_a^δ - x^p∥\right). \] (3.8)

From assumption 1 with p = 1, \( \|\hat{R}_a^δ\| ≤ 1 \) and assumption 2 we obtain

\[ \|\hat{R}_a^δ(x_a^δ - x)\| = \|\hat{R}_a^δ(x_a^δ - x)\| \leq \|\hat{R}_a^δ(x_a^δ - x)\| + ∥\hat{R}_a^δ(x_a^δ - x)\|L∩v∥∥x_a^δ - x^p∥ \]

which gives, using the implication \( c^2 ≤ a + b \) \( c ≤ a + b \) and \( ∥R_a^δ∥^{1/2} ≤ 1, \)

\[ \|\hat{R}_a^δ(x_a^δ - x)\| ≤ \sqrt{(\|\hat{R}_a^δ(x_a^δ - x)\|)^2 + (∥\hat{R}_a^δ(x_a^δ - x)∥L∩v∥∥x_a^δ - x^p∥} \]

\[ ≤ (∥\hat{R}_a^δ∥^2 A_a^δ(x_a^δ - x)∥L∩v∥∥x_a^δ - x^p∥ + ∥\hat{R}_a^δ∥L∩v∥∥x_a^δ - x^p∥). \] (3.9)

We multiply the Euler equation \((A_a^δ)^*[F(x_a^δ) − y] + α(x_a^δ − x) = 0 \) by \( A_a^δ \) and obtain the equation \((A_a^δ)^*[F(x_a^δ) − y] = αA_a^δ(x_a^δ − x)\). Adding \( α[F(x_a^δ) − y] \) on both sides and multiplying by \((A_a^δ)^*[A_a^δ + αI]^{-1}\) yields

\[ F(x_a^δ) − y = \hat{R}_a^δ(F(x_a^δ) + A_a^δ(x_a^δ − x) − y) \]

\[ \hat{R}_a^δ(F) + A_a^δ(x_a^δ − x) − y = \hat{R}_a^δ(F + A_a^δ(x_a^δ − x) − y) + R_a^δ A_a^δ(x_a^δ − x). \]

Multiplying by \((R_a^δ)^{1/2}\) provides

\[ (R_a^δ)^{1/2} A_a^δ(x_a^δ − x) = (R_a^δ)^{1/2}[F(x_a^δ) − y] − (R_a^δ)^{1/2}[F(x_a^δ) + A_a^δ(x_a^δ − x) − y]. \] (3.10)

From (3.10) we obtain due to rule 1

\[ ∥(R_a^δ)^{1/2} A_a^δ(x_a^δ − x)∥^{1/2} ≤ \sqrt{C^2 + δ + \frac{L}{2}∥x_a^δ - x^p∥} ≤ \sqrt{C + 1} + δ + \sqrt{\frac{L}{2}∥x_a^δ - x^p∥}. \] (3.11)
Now the proof of the theorem for rule 1 follows from (3.8), (3.9) and (3.11). In order to prove the theorem for rule 2 we note that rule 1 has been exploited in the estimates (3.8) and (3.11). However, due to proposition 3.2 and \( d_1(\alpha) \leq d_2(\alpha) \), the estimates (3.8) and (3.11) are also valid for rule 2. Hence, the proof is complete.

In the second part of this subsection we shall prove \( \|x_\alpha - x^*\| = O(\sqrt{\delta}) \) provided \( \alpha \) is chosen by rule 1. This result will especially be useful for studying the smooth case \( p \in [1, 2] \) in section 3.3. We start our examinations with an auxiliary result whose proof is along the lines of lemma 4.2 in [17].

**Proposition 3.4.** Let assumption 2 hold and let \( x, z \in B_r(x^+) \). Define
\[
A_x := F'(x), \quad A_z := F'(z),
\]
\[
R_x := \alpha(A_x A_x^* + \alpha I)^{-1} \quad \text{and} \quad R_z := \alpha(A_z A_z^* + \alpha I)^{-1}.
\]
Then for \( a = \|R_x^{1/2}y\|^2 \) and \( b = \|R_x^{1/2}y\|^2 \) with arbitrary \( y \in Y \) there holds
\[
|a - b| \leq \frac{L\|x - z\|}{\sqrt{\alpha}}(a + b).
\]

**Theorem 3.5.** Let \( \alpha \) be chosen by rule 1 with \( C > 1 \). Suppose assumption 1 with \( p = 1 \) and assumption 2 with radius \( r = 2\|X - x^*\| \). If \( L\|v\| \) is sufficiently small that \( 2L\|v\| \leq 3 \),
\[
\varepsilon_3 := \frac{L\|v\|}{(C - 1)\sqrt{1 - L\|v\|}} < 1 \quad \text{and} \quad \varepsilon_4 := \sqrt{\frac{L\|v\|}{2} + L\|v\| + \frac{L\|v\|}{8\sqrt{1 - L\|v\|}}} < 1,
\]
then
\[
\|x_\alpha - x^*\| \leq k_4\sqrt{\delta} \quad \text{with} \quad k_4 = \left[ \frac{\|v\|^{1/2}}{1 - \varepsilon_4} \right]^{1/2}
\]
From (3.18), (3.19), \( \| R^{1/2}_a \| \leq 1 \), (1.5) and the estimate \( \sqrt{a} + b \leq \sqrt{a} + \sqrt{b} \) we obtain
\[
\| R^{1/2}_a (x^* - x^t) \|^{1/2} \leq \left[ 1 + C \sqrt{\frac{1 + \varepsilon_3}{1 - \varepsilon_3}} \right]^{1/2} \sqrt{\| x_a - x^t \|}.
\] (3.20)

Now (3.14) follows from (3.16), (3.17) and (3.20). \( \square \)

### 3.3. Error bounds in the case \( p \in [1, 2] \)

In this subsection we will show that it is possible to derive order optimal error bounds for
\( \| x^d - x^t \| \) under assumption 1 and assumption 2 for the special case \( p \in [1, 2] \) provided the regularization parameter \( \alpha \) is chosen from rule 1. Unfortunately, analogous results for rule 2 are unknown so far. Since the proof for \( p = 1 \) in section 3.2 fails for the general case \( p \in [1, 2] \), we will use another method which combines some new ideas with those from [19] where the total error \( \| x^d - x^t \| \) has been decomposed into the sum \( \| x_a - x_a^d \| + \| x_a^d - x^t \| + \| x_a^d - x_a \| \) with a properly chosen \( \alpha_0 \). In order to maintain the right order \( \mathcal{O}(\delta^{p/(p+1)}) \), the nonlinear terms appearing in the estimates have to be handled carefully. We start our study with a proposition which gives a lower bound for the regularization parameter \( \alpha = \alpha(\delta) \) obtained by rule 1 which, for \( p > 1 \), is sharper than the bound (3.4).

**Proposition 3.6.** Let \( \alpha = \alpha(\delta) \) be chosen by rule 1 with \( C > 1 \). Assume assumption 1 with \( p \in [1, 2] \) and assumption 2 with radius \( r = \delta/\sqrt{\alpha} + 2 \| x - y \| \). If \( L \| v \| \) is sufficiently small such that \( 3L \| v \| \leq 2 \) and \( \varepsilon_1 < 1 \) with \( \varepsilon_3 \) given in (3.13), then
\[
\alpha \geq \left( \frac{C - 1}{\| v \|} \sqrt{\frac{1 - \varepsilon_3}{1 + \varepsilon_3}} \right)^{2/(p+1)} \delta^{2/(p+1)}. \tag{3.21}
\]

**Proof.** From rule 1, \( \| R^{1/2}_a \| \leq 1 \) and the second inequality of (2.6) we have
\[
C \delta = \|(R^{1/2}_a) F(x_a) - y^d\| \leq \|(R^{1/2}_a) F(x_a) - y\| + \| F(x^d) - y^d + y - F(x_a) \|
\leq \|(R^{1/2}_a) F(x_a) - y\| + \delta
\]
which gives
\[
(C - 1) \delta \leq \|(R^{1/2}_a) F(x_a) - y\|.
\] (3.22)

Changing the roles of \( x_a \) and \( x^d \) we obtain in analogy to the first part of (3.19) that
\[
\|(R^{1/2}_a) F(x_a) - y\| \leq \left[ \frac{1 + \varepsilon_3}{1 - \varepsilon_3} \right] \|(R^{1/2}_a) F(x_a) - y\|.
\] (3.23)

To estimate \( \| R^{1/2}_a [F(x_a) - y] \| \) in terms of \( \alpha \) we use the triangle inequality, apply the estimates \( \| R^{1/2}_a \| \leq 1 \), \( \| R^{1/2}_a \| \leq \sqrt{\alpha} \) and the two estimates (2.8) with \( \delta = 0 \) and (2.3), and obtain
\[
\| R^{1/2}_a [F(x_a) - y] \| \leq \| R^{1/2}_a [F(x_a) - y - x_a + x^t] + R^{1/2}_a [x_a - x^t] \|
\leq \left\{ \frac{L}{\sqrt{1 - \| v \|}} \| x_a - x^t \| + \sqrt{\alpha} \right\} \| x_a - x^t \|
\leq \left\{ \frac{L}{2 \sqrt{1 - \| v \|}} + 1 \right\} \alpha^{(p+1)/2} \| v \|^{1/2} \frac{1 + \| v \|/2}{\sqrt{1 - \| v \|}} \leq \alpha^{(p+1)/2} \| v \|^2 \left( \frac{1 + \| v \|/2}{\sqrt{1 - \| v \|}} \right)^2.
\] (3.24)

Now estimate (3.21) follows from (3.22)–(3.24). \( \square \)

Our second proposition in this subsection shows that under appropriate conditions the norm \( \| x_a - x_a^d \| \) with \( 0 < \alpha_0 \leq \alpha \) and \( \alpha \) chosen from rule 1 can be estimated properly.
Proposition 3.7. Let $\alpha$ be chosen by rule 1 with $C > 1$. Assume assumption 1 with $p \in [1, 2]$ and assumption 2 with radius $\delta = \frac{1}{2\sqrt{\alpha} + 2\|\kappa - x^\dagger\|}$. If $\alpha \geq \alpha_0 := c_0 \delta^{2/p+1}$ with a constant $c_0$ independent of $\delta$ and if $L\|u\|$ is sufficiently small that $3L\|v\| \leq 2$, (3.13) and

$$\varepsilon_s := L\|v\| + \frac{L\|v\|}{8\sqrt{1 - L\|v\|}} + \frac{Lk_4}{4\sqrt{c_0}} \delta^{(p-1)/(2p+2)} < 1$$

(3.25)

with $k_4$ given in (3.14), then

$$\|x_\alpha - x_{\alpha_0}\| \leq \frac{\|R_0^{3/2}[F(x_\alpha) - y]\|}{(1 - \varepsilon_s)\sqrt{\alpha_0}}$$

(3.26)

Proof. From the Euler equation $A^*\alpha[F(x_\alpha) - y] + \alpha(x_\alpha - \kappa) = 0$ we have

$$\alpha(\kappa - x_\alpha) = A^*_a[F(x_\alpha) - y] \quad \text{and} \quad \alpha_0(\kappa - x_{\alpha_0}) = A^*_a[F(x_{\alpha_0}) - y].$$

(3.27)

Due to the identity $\alpha_0(x_\alpha - x_{\alpha_0}) = (\alpha - \alpha_0)(\kappa - x_\alpha) + \alpha_0(\kappa - x_{\alpha_0}) - \alpha(\kappa - x_\alpha)$ we obtain by using (3.27) that

$$\alpha_0(x_\alpha - x_{\alpha_0}) = \frac{\alpha - \alpha_0}{\alpha} A^*_a[F(x_\alpha) - y] + A^*_a[F(x_{\alpha_0}) - y] - A^*_a[F(x_\alpha) - y].$$

We add on both sides $A^*_aA_\alpha(x_\alpha - x_{\alpha_0})$, multiply by $(A^*_aA_\alpha + \alpha_0 I)^{-1}$ and obtain

$$x_\alpha - x_{\alpha_0} = (A^*_aA_\alpha + \alpha_0 I)^{-1} \left\{ \frac{\alpha - \alpha_0}{\alpha} A^*_a[F(x_\alpha) - y] + A^*_aA_\alpha(x_\alpha - x_{\alpha_0}) \right\}
+ A^*_a[F(x_{\alpha_0}) - y] - A^*_a[F(x_\alpha) - y] - (A^*_aA_\alpha + \alpha_0 I)[y - F(x_{\alpha_0})]
+ A^*_a[F(x_{\alpha_0}) - F(x_\alpha) - A_\alpha(x_\alpha - x_{\alpha_0})] =: s_1 + s_2 + s_3.$$

(3.28)

From [19] we know that due to the monotonicity of $\alpha/(\lambda + \alpha)$ as a function of the regularization parameter $\alpha$ there holds

$$\left\|\frac{\alpha}{\alpha_0}(AA^* + \alpha_0 I)^{-1} \frac{1}{\alpha}(AA^* + \alpha I)\right\| \leq 1 \quad \text{for} \quad \alpha_0 \leq \alpha.$$ 

(3.29)

Using the estimates $|\alpha - \alpha_0|/\alpha_0 \leq 1$, $\|A^*_a(A_\alpha A^*_a + \alpha_0 I)^{-1/2}\| \leq 1$ and (3.29) we obtain for the first summand $s_1$ on the right-hand side of (3.28)

$$\|s_1\| \leq \|A^*_aA_\alpha A^*_a + \alpha_0 I\|^{1/2} \|\sqrt{\alpha_0}(A^*_aA_\alpha A^*_a + \alpha_0 I)^{-1/2} \frac{1}{\sqrt{\alpha}}(A^*_aA_\alpha + \alpha I)^{1/2}\| \times \|\sqrt{\alpha_0}[A_\alpha A^*_a + \alpha I]^{-1/2} [F(x_{\alpha_0}) - y]\| \leq \frac{R_0^{1/2}[F(x_{\alpha_0}) - y]}{\sqrt{\alpha_0}}.$$ 

(3.30)

To estimate the second summand $s_2$ we use assumption 2 as well as (2.3) and obtain

$$\|s_2\| \leq \frac{L}{\alpha_0} \|x_\alpha - x_{\alpha_0}\| \|F(x_{\alpha_0}) - y\| \leq L\|v\| \|x_\alpha - x_{\alpha_0}\|.$$ 

(3.31)

In order to estimate the third summand $s_3$ on the right-hand side of (3.28) we first give an upper bound for $\|x_\alpha - x_{\alpha_0}\|/\alpha_0$ with $\alpha$ from rule 1 and $\alpha \geq \alpha_0$. This bound can be obtained by using theorems 3.5 and 2.3, which provide the estimate

$$\|x_\alpha - x_{\alpha_0}\| \leq \frac{1}{\sqrt{\alpha_0}} (\|x_\alpha - x^\dagger\| + \|x_{\alpha_0} - x^\dagger\|) \leq \frac{k_4}{\sqrt{c_0}} \delta^{(p-1)/(2p+2)} + \frac{\|v\|}{2\sqrt{1 - L\|v\|}}$$

(3.32)
Proof. Consider a fixed regularization parameter $\alpha = \alpha_0$ of the form

$$\alpha_0 = C_0 \delta^{2/(p+1)}$$

and distinguish two cases. In the first case we assume that the solution $\alpha = \alpha(\delta)$ of rule 1 satisfies $\alpha \leq \alpha_0$. In this case we obtain from theorem 2.5 that

$$\|x_\alpha^\delta - x^\dagger\| \leq \alpha_0^{p/2} \|w\| \frac{1 + \sqrt{1 - L\|v\|}}{\sqrt{1 - L\|v\|}} + \frac{\delta}{\sqrt{1 - L\|v\|}}$$

and in the second case we assume that the solution $\alpha = \alpha(\delta)$ of rule 1 satisfies $\alpha \geq \alpha_0$. In this second case we use the triangle inequality and obtain from proposition 3.7, theorem 2.5 with $\delta = 0$ and theorem 2.4 that

$$\|x_\alpha^\delta - x^\dagger\| \leq \|x_\delta - x_\alpha\| + \|x_\alpha - x_\alpha^\delta\| + \|x_\alpha^\delta - x^\dagger\|$$

with $\varepsilon_\delta$ given in (3.25). Since the error bound (3.36) of the first case is smaller than the error bound (3.37) of the second case we conclude that the error bound for $\|x_\alpha^\delta - x^\dagger\|$ with $\alpha$ chosen from rule 1 is given by (3.37). To estimate (3.37) in terms of $\delta$ we use (3.19), substitute the special value $\alpha_0$ from (3.35) into the first two summands on the right-hand side of (3.37), use inequality (3.21) to estimate the third summand on the right-hand side of (3.37) and obtain

$$\|x_\alpha^\delta - x^\dagger\| \leq \left[ 1 + C \frac{1 + \varepsilon_3}{1 - \varepsilon_3} \frac{\delta^{p/(p+1)}}{(1 - \varepsilon_3)^2 \sqrt{\varepsilon_0}} + C_0 \alpha_0^{p/2} \|w\| \frac{1 + \sqrt{1 - L\|v\|}}{\sqrt{1 - L\|v\|}} \right]$$

$$+ \left( \frac{\|w\|}{C - 1} \frac{1 + \varepsilon_3}{1 - \varepsilon_3} \frac{1 + \sqrt{1 - L\|v\|}}{\sqrt{1 - L\|v\|}} \right)^{1/(p+1)} \frac{\delta^{p/(p+1)}}{\sqrt{1 - L\|v\|}}$$

with $\varepsilon_3$ given by (3.13). Hence, the proof of (3.34) is complete. □

Since the constant $C_0 > 0$ in the error bound (3.38) is arbitrary, a sharp constant $C_0$ in the error estimate (3.34) can be found by minimizing (3.38) with respect to $c_0$. 

with $k_4$ given in (3.14). Due to (1.6), (1.5) and (3.32) we obtain that the third summand $s_3$ on the right-hand side of (3.28) can be estimated by

$$\|s_3\| \leq \frac{L}{4\sqrt{c_0}} \|x_\alpha^\delta - x_\alpha\|^2 \leq \left( \frac{L k_4}{4\sqrt{c_0}} \frac{\delta^{p/2}}{\sqrt{1 - L\|v\|}} \right) \|x_\alpha^\delta - x_\alpha\|.$$ (3.33)

From (3.28), (3.30), (3.31) and (3.33) we obtain (3.26).

**Theorem 3.8.** Let $\alpha$ be chosen by rule 1 with $C > 1$. Assume assumption 1 with $p \in [1, 2]$ and assumption 2 with radius $r = \delta/\sqrt{\alpha} + 2\|T - x^\dagger\|$. Suppose $L\|v\|$ is sufficiently small that $3L\|v\| \leq 2$, (3.13) and (3.25) hold. Then there exists a constant $C_0$ independent of $\delta$ with

$$\|x_\alpha^\delta - x^\dagger\| \leq C_0 \delta^{p/(p+1)}.$$ (3.44)
3.4. Error bounds in the case \( p \in (0, 1] \)

In this subsection we will show that in the case \( p \in (0, 1] \) order optimal error bounds are valid under assumption 3 and assumption 4 provided \( \alpha \) is chosen from rule 1 or rule 2, respectively.

Let us start with a preliminary proposition which gives a lower bound for the regularization parameter \( \alpha \) chosen by rule 1 or rule 2, respectively.

**Proposition 3.9.** Let \( \alpha \) be chosen by rule 1 or rule 2 with \( C > 1 \). Assume assumption 3 with \( p \in (0, 1] \) and assumption 4 with radius \( r = \delta/\sqrt{\alpha} + 2\| \mathbf{x} - \mathbf{x}^\dagger \| \). If \( \varepsilon_0 := k_0\| \mathbf{x} - \mathbf{x}^\dagger \| \leq 1/2 \), then

\[
\alpha \geq \left( \frac{C - 1}{\| w \| \sqrt{1 + \varepsilon_0^2}} \right)^{2/(p+1)} \delta^{2/(p+1)}. \tag{3.39}
\]

**Proof.** First, let us estimate \( \| F(x_\alpha) - y \| \) in terms of \( \alpha \). We use (2.16) with \( x_1 = x_\alpha \), \( x_2 = x^\dagger \), the Euler equation \( A_\alpha^*[F(x_\alpha) - y] + \alpha(x_\alpha - \mathbf{x}) = 0 \) as well as the first estimate of (2.1) with \( \delta = 0 \) and obtain

\[
(F(x_\alpha) + A_\alpha(x^\dagger - x_\alpha) - y, F(x_\alpha) - y) \leq \int_0^1 \| k(x_\alpha + t(x^\dagger - x_\alpha), x_\alpha, x^\dagger - x_\alpha) \| A_\alpha^*[F(x_\alpha) - y]\| \leq k_0 \frac{1}{2} \| x_\alpha - x^\dagger \| \| x_\alpha - \mathbf{x} \| \leq \frac{\alpha k_0}{2} \| \mathbf{x} - \mathbf{x}^\dagger \| \| x_\alpha - x^\dagger \| ^2. \tag{3.40}
\]

From (2.13) we obtain

\[
\alpha(x_\alpha - x^\dagger, \mathbf{x} - x^\dagger) = \alpha^2((A^* A + \alpha I)^{-1}(\mathbf{x} - x^\dagger), \mathbf{x} - x^\dagger) + \alpha((A^* A + \alpha I)^{-1}(A^* A - A^\dagger)[y - F(x_\alpha)], \mathbf{x} - x^\dagger) + \alpha((A^* A + \alpha I)^{-1}A^*[y + A(x_\alpha - x^\dagger)] - F(x_\alpha)], \mathbf{x} - x^\dagger) =: s_1 + s_2 + s_3. \tag{3.41}
\]

To estimate \( s_1 \) we use assumption 3 with \( p \in (0, 1] \) and obtain due to (1.6)

\[
s_1 = \alpha^2((A^* A + \alpha I)^{-1}(A^* A)^p w, w) \leq \alpha^{p+1} \| w \|^2. \tag{3.42}
\]

To estimate \( s_2 \) we use assumption 3 and assumption 4, the Euler equation \( A_\alpha^*[F(x_\alpha) - y] + \alpha(x_\alpha - \mathbf{x}) = 0 \), the first estimate of (2.1) with \( \delta = 0 \) as well as the estimate \( 2ab \leq a^2 + b^2 \) and obtain

\[
s_2 = \alpha(F(x_\alpha) - y, (A - A_\alpha)(A^* A + \alpha I)^{-1}(A^* A)^{p/2} w) = \alpha^2(\mathbf{x} - x_\alpha, k(x^\dagger, x^\dagger, (A^* A + \alpha I)^{-1}(A^* A)^{p/2} w) \leq \alpha^{p/2+1}k_0 \| \mathbf{x} - \mathbf{x}^\dagger \| \| x_\alpha - x^\dagger \| \| w \| \leq \alpha k_0 \| \mathbf{x} - \mathbf{x}^\dagger \| \| x_\alpha - x^\dagger \| ^2 + \frac{1}{2} \alpha^{p+1} k_0 \| \mathbf{x} - \mathbf{x}^\dagger \| \| w \| ^2. \tag{3.43}
\]

To estimate \( s_3 \) we use (2.16) and obtain due to assumption 4

\[
s_3 = \alpha(y + A(x_\alpha - x^\dagger) - F(x_\alpha), A(A^* A + \alpha I)^{-1} \mathbf{x} - x^\dagger) \leq \alpha \| \mathbf{x} - \mathbf{x}^\dagger \| \int_0^1 \| k(x^\dagger + t(x_\alpha - x^\dagger), x^\dagger, x_\alpha - x^\dagger) \| \| \mathbf{x} - \mathbf{x}^\dagger \| \| x_\alpha - x^\dagger \| \| w \| \leq \frac{\alpha k_0}{2} \| \mathbf{x} - \mathbf{x}^\dagger \| \| x_\alpha - x^\dagger \| ^2. \tag{3.44}
\]

From the first part of (2.4) and (3.40) – (3.44) we obtain

\[
\| F(x_\alpha) - y \|^2 + \alpha \| x_\alpha - x^\dagger \|^2 \leq (1 + \frac{1}{2}k_0 \| \mathbf{x} - \mathbf{x}^\dagger \|) \alpha^{p+1} \| w \| ^2 + 2\alpha k_0 \| \mathbf{x} - \mathbf{x}^\dagger \| \| x_\alpha - x^\dagger \| ^2.
\]
We use the assumption $2k_0 \| \bar{x} - x^\dagger \| \leq 1$ and obtain

$$
\| F(x_a) - y \| \leq \sqrt{1 + k_0 \| \bar{x} - x^\dagger \| / 4} \| w \| (p^{(p+1)/2}).
$$

(3.45)

Since $d_1(\alpha) \leq \| F(x_a^\dagger) - y \|$ and $d_2(\alpha) \leq \| F(x_a^\dagger) - y \|$, we conclude from the second estimate of (2.18) that for both rule 1 and rule 2, respectively, we have $C \delta \leq \delta + \| F(x_a) - y \|$. This estimate and (3.45) provide (3.39).

In the next proposition we estimate $\| x_a - x_{a_0} \|$ for $0 < a_0 \leq \alpha$.

**Proposition 3.10.** Let assumption 4 with $\varepsilon_6 := k_0 \| x^\dagger - \bar{x} \| < 1/2$ and radius $r = 2 \| \bar{x} - x^\dagger \|$ be satisfied. Then, for all $0 < a_0 \leq \alpha$,

$$
\| x_a - x_{a_0} \| \leq \frac{\| R_{a_0}^2 \| F(x_a) - y \|}{(1 - 2\varepsilon_6))\sqrt{C_0}}.
$$

(3.46)

**Proof.** We consider the error representation (3.28) and estimate the three summands $s_1$, $s_2$ and $s_3$ separately. To estimate the first summand $s_1$ we use (3.30). In order to estimate the second summand $s_2$ on the right-hand side of (3.28) we proceed along the lines of the proof of (2.15) and obtain

$$
\| s_2 \| \leq k_0 \| x^\dagger - \bar{x} \| \| x_a - x_{a_0} \|.
$$

(3.47)

To estimate the third summand $s_3$ on the right-hand side of (3.28) we proceed according to the proof of (2.17) and obtain

$$
\| s_3 \| \leq k_0 \| x^\dagger - \bar{x} \| \| x_a - x_{a_0} \|.
$$

(3.48)

Now the proof of (3.46) follows from (3.28), (3.30), (3.47) and (3.48).

Before we provide order optimal error bounds for $\| x_a^\delta - x^\dagger \|$ with $\alpha$ chosen from rule 1 or rule 2, respectively, let us formulate an auxiliary result whose proof is based on lemma 3.6 in [27].

**Proposition 3.11.** Let assumption 4 hold and let $x, z \in B_r(x^\dagger)$. Define

$$
A_x := F'(x), \quad A_z := F'(z),
$$

$$
R_x := \alpha(A_x A_x^\dagger + \alpha I)^{-1} \quad \text{and} \quad R_z := \alpha(A_z A_z^\dagger + \alpha I)^{-1}.
$$

Then for $a = \| R_x^{1/2} y \|$, $b = \| R_z^{1/2} y \|$ with arbitrary $y \in Y$ there holds

$$
| a - b | \leq k_0 \| x - z \| | a + b |.
$$

(3.49)

Now we are ready to provide order optimal error bounds for $\| x_a^\delta - x^\dagger \|$ with $\alpha$ chosen from rule 1 or rule 2, respectively.

**Theorem 3.12.** Let $\alpha = \alpha(\delta)$ be chosen by rule 1 or rule 2 with $C > 1$. Assume assumption 3 with fixed $p \in (0, 1]$ and assumption 4 with radius $r = 2 \| \bar{x} - x^\dagger \| + \delta / \sqrt{\alpha}$. If $k_0 \| \bar{x} - x^\dagger \|$ is sufficiently small such that $\varepsilon_6 := k_0 \| \bar{x} - x^\dagger \| < 1/2$ then

$$
\varepsilon_7 := \frac{k_0 \| \bar{x} - x^\dagger \|}{(C - 1) \sqrt{1 - k_0 \| \bar{x} - x^\dagger \|}} < 1,
$$

(3.50)

then

$$
\| x_a^\delta - x^\dagger \| \leq c_p \| w \| (1 + \varepsilon_7 \varepsilon_7^{p/(p+1)})^p
$$

(3.51)

with a constant $c_p$ independent of $\delta$ and $\| w \|$ of the form

$$
c_p = \frac{2}{1 - 2\varepsilon_6} \left[ 1 + C \sqrt{\frac{1 + \varepsilon_7}{1 - \varepsilon_7}} \right]^{p/(p+1)} + \frac{1}{\sqrt{1 - \varepsilon_6} \sqrt{C - 1}} \left[ \frac{\sqrt{1 + \varepsilon_6 / 4}}{C - 1} \right]^{1/(p+1)}.
$$

(3.52)
Proof. Consider a fixed regularization parameter \( \alpha = \alpha_0 \) of the form
\[
\alpha_0 = c_0 \left( \frac{\delta}{\|w\|} \right)^{2/(p+1)}
\]
with \( c_0 = \left[ 1 + C \sqrt{\frac{1 + \varepsilon_7}{1 - \varepsilon_7}} \right]^{2/(p+1)} \quad (3.53)
\]
and distinguish two cases. In the first case we assume that the solution \( \alpha = \alpha(\delta) \) of rule 1 or rule 2, respectively, satisfies \( \alpha \leq \alpha_0 \). In this first case we obtain from (2.12) and the first inequality of (2.18) that
\[
\|x_a^\delta - x^\dagger\| \leq \|x_a - x^\dagger\| + \|x_a^\delta - x_a\| \leq \|w\| \frac{\alpha_0^{p/2}}{1 - 2\varepsilon_6} + \frac{\delta}{\sqrt{\alpha} \sqrt{1 - \varepsilon_6}}.
\]
In the second case we assume that the solution \( \alpha = \alpha(\delta) \) of rule 1 or rule 2 satisfies \( \alpha \geq \alpha_0 \). In this second case we obtain from proposition 3.10, (2.12) and the first inequality of (2.18) that
\[
\|x_a^\delta - x^\dagger\| \leq \|x_a - x^\dagger\| + \|x_a^\delta - x_a\| \leq \frac{R_\alpha^{1/2}[F(x_a^\delta) - y] + \alpha_0^{p/2}\|w\|}{1 - 2\varepsilon_6} + \frac{\delta}{\sqrt{\alpha} \sqrt{1 - \varepsilon_6}}.
\]
Since the error bound (3.54) of the first case is smaller than the error bound (3.55) of the second case we conclude that the error bound for \( \|x_a^\delta - x^\dagger\| \) with \( \alpha \) chosen from rule 1 or rule 2, respectively, is given by (3.55). To estimate \( \|R_\alpha^{1/2}[F(x_a^\delta) - y]\| \) in terms of \( \delta \) we use proposition 3.11 with \( a = \|R_\alpha^{1/2}[F(x_a^\delta) - y]\| \) and \( b = \|R_\alpha^{1/2}[F(x_a^\delta) - y]\|^2 \) and obtain due to (2.18) and (3.3) the estimate
\[
|a - b| \leq k_0 \|x_a^\delta - x_a\| (a + b) \leq \frac{k_0 \delta}{\sqrt{\alpha} \sqrt{1 - \varepsilon_6}} (a + b) \leq 7 \varepsilon_\gamma (a + b)
\]
with \( \varepsilon_\gamma \) given in (3.50). This estimate provides \( a \leq (1 + \varepsilon_\gamma)(1 - \varepsilon_\gamma)^{-1} b \). We apply the triangle inequality, the estimate \( \|R_\alpha^{1/2}\| \leq 1 \), the second estimate of (2.18) and rule 1 or rule 2, respectively, use in the case of rule 2 the valid estimate \( d_1(\alpha) \leq d_2(\alpha) \) and obtain
\[
\|R_\alpha^{1/2}[F(x_a^\delta) - y]\| \leq \|R_\alpha^{1/2}[F(x_a^\delta) - y^\delta + y - F(x_a)]\| + \|R_\alpha^{1/2}[F(x_a^\delta) - y]\|
\]
\[
\leq \delta + \sqrt{\frac{1 + \varepsilon_7}{1 - \varepsilon_6}} \|R_\alpha^{1/2}[F(x_a^\delta) - y]\| \leq \left[ 1 + C \sqrt{\frac{1 + \varepsilon_7}{1 - \varepsilon_7}} \right] \delta.
\]
Substituting (3.56) into the first summand on the right-hand side of (3.55), substituting the special value \( \alpha_0 \) from (3.53) into the first two summands on the right-hand side of (3.55) and using inequality (3.39) to estimate the third summand on the right-hand side of (3.55) provides (3.51) and (3.52).

Finally we remark that for rule 1 the results of theorem 3.12 can be extended to the maximal range \( p \in (0, 2] \). For the proof of this result we show along the lines of proposition 3.6 that for \( \alpha \) from rule 1 we have \( \alpha \geq c \delta^{2/(p+1)} \) with a constant \( c > 0 \), proceed according to the proof of theorem 3.12 and obtain

**Theorem 3.13.** Let \( \alpha = \alpha(\delta) \) be chosen by rule 1 with \( C > 1 \). Assume assumption 3 with fixed \( p \in (0, 2] \) and assumption 4 with radius \( r = 2 \|x - x^\dagger\| + \delta/\sqrt{\alpha} \). If \( k_0 \|x - x^\dagger\| \) is sufficiently small that \( \varepsilon_6 := k_0 \|x - x^\dagger\| < 1/2 \) and (3.50) hold, then
\[
\|x_a^\delta - x^\dagger\| \leq c_p \|w\| \frac{\delta}{\sqrt{\alpha}}
\]
with a constant \( c_p \), independent of \( \delta \) and \( \|w\| \) of the form
\[
c_p = \frac{2}{1 - 2\varepsilon_6} \left( 1 + C \frac{1 + \varepsilon_7}{1 - \varepsilon_7} \right)^{p/(p+1)} + \frac{1}{\sqrt{1 - \varepsilon_6}} \left( 1 + \varepsilon_7 \right) \frac{1 + \varepsilon_6}{(C - 1)(1 - 2\varepsilon_6)} \right]^{1/(p+1)}.
\]
4. Illustration

Nonlinear ill posed problems (1.1) arise in a number of applications. In this section we will shortly discuss a parameter identification problem which has been considered in different papers, see, e.g., [4, 5, 20]. This nonlinear ill posed problem consists of identifying the coefficient \( c(x), x \in (0, 1) \), in the elliptic boundary value problem

\[-u_{xx} + cu = f, \quad u(0) = g_0, \quad u(1) = g_1 \quad (4.1)\]

from noisy data \( u^\delta(x) \in Y = L^2(0, 1) \) where \( f \in L^2(0, 1) \), \( g_0 \) and \( g_1 \) are given. We define the nonlinear operator \( F : D(F) \subset X \to Y \) as the solution operator \( c \to u \) of the boundary value problem (4.1) according to

\[ F(c) = u \quad \text{with} \quad D(F) = \{ c \in X | c \geq 0 \text{ a.e.} \} \quad (4.2) \]

where \( u = u(x; c) \) is the solution of (4.1) for given \( c \in D(F) \). We consider (4.2) in an \( L^2 \)-space setting with \( X = Y = L^2(0, 1) \). Our test example is given by

Example 1. \( u(x) = 100 + x(1 - x), \quad c^\delta(x) = 1 + \frac{31}{18} x - \frac{49}{18} x^3 + \frac{2}{5} x^5 - \frac{1}{5} x^7, \quad \overline{c}(x) = 1. \)

In example 1, the function \( f \) of problem (4.1) has the form \( f(x) = 2 + c^\delta(x)u(x) \). For this test example assumption 1 is satisfied with \( p = 2 \), see [20]. To discretize problem (4.1) we decompose the interval \([0, 1]\) into \( n + 1 \) equidistant subintervals \([x_j, x_{j+1}]\) \((j = 0, \ldots, n)\) of length \( h = x_{j+1} - x_j = 1/(n + 1) \) and approximate the boundary value problem (4.1) by the finite-dimensional system equation

\[ T_u u = b, \quad T : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \]

of the form \([h^{-2}\text{tridiag}(-1, 2, -1) + \text{diag}(c_i)]u = b\) with \( u = (u_i)_{i=1}^n, b = (b_i)_{i=1}^n, u_i = u(x_i), c_i = c(x_i) \) for \( i = 1, \ldots, n \), \( b_i = f(x_i) \) for \( i = 2, \ldots, n - 1 \), \( b_1 = f(x_1) + g_0/h^2 \) and \( b_n = f(x_n) + g_1/h^2 \). The partial derivatives \( T_u \) and \( T_c \) of the bilinear operator \( T \) are given by

\[ T_u = h^{-2} \text{tridiag}(-1, 2, -1) + \text{diag}(c_i), \quad T_c = \text{diag}(u_i). \]

Due to the implicit function theorem the Fréchet derivative \( J := F'(c) \) can be computed according to the formula \( J = -T_u^{-1}T_c \) with \( u = F(c) = T_u^{-1}b \). To compute discretized regularized approximations \( c^\delta \in \mathbb{R}^n \), instead of exact data vectors \( u \in \mathbb{R}^n \) normally distributed randomly perturbed data vectors \( u^\delta \in \mathbb{R}^n \) with \( \|u - u^\delta\| \leq \delta \) have been used where here and in the rest of this subsection \( \| \cdot \| \) denotes the discretized \( L^2 \)-norm. To solve the discretized Euler equation (1.4) by the Gauss–Newton iteration

\[ c := c - (J^T(J + \alpha I)^{-1}J^T(F(c) - u^\delta) + \alpha(c - \overline{c})) \quad (4.3) \]

due to the implicit function theorem the Fréchet derivative \( J = -T_u^{-1}T_c \) has to be generated and a dense linear system with the symmetric positive definite system matrix \( J^T J + \alpha I \) has to be solved. However, the partial derivative matrices \( T_u \) and \( T_c \) are sparse. Hence, we replace the Gauss–Newton iteration (4.3) by an equivalent iteration with sparse matrix operations:

\[ c := \overline{c} - J^T(JJ^T + \alpha I)^{-1}[F(c) - u^\delta] \]
\[ = \overline{c} + T^\delta_c(T_c(T_c^T + \alpha T_u T_u^T)^{-1}[b - T_u u^\delta + T_c(c - \overline{c})]. \quad (4.4) \]

Due to \( T_c = T_c^T \) and \( T_u = T_u^T \), this equivalent iteration makes it possible to perform one iteration step by

(i) solving the tridiagonal system \( T_u u = b \) in order to find \( u \),
(ii) solving the five-diagonal system \( (T_c^T + \alpha T_u^T)p = b - T_u u^\delta + T_c(c - \overline{c}) \) to find \( p \) and
(iii) computing the new iterate \( c \) according to \( c := \overline{c} + T_c p. \)
Tikhonov regularization and a posteriori rules.

The results in table 2 show that the regularization parameters follows.

In our numerical experiments the starting vector $c_0$ chosen according to rules 1, 2 and 3.

Table 1. Errors for example 1 with $\alpha$ chosen according to rules 1, 2 and 3.

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$e_1$</th>
<th>$e_2$</th>
<th>$e_3$</th>
<th>$e_{1/2}$</th>
<th>$e_{2/3}$</th>
<th>$e_{1/3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-2}$</td>
<td>$4.87 \times 10^{-3}$</td>
<td>$3.83 \times 10^{-3}$</td>
<td>$1.90 \times 10^{-3}$</td>
<td>$1.05 \times 10^{-1}$</td>
<td>$8.24 \times 10^{-2}$</td>
<td>$1.90 \times 10^{-2}$</td>
</tr>
<tr>
<td>$10^{-3}$</td>
<td>$1.06 \times 10^{-3}$</td>
<td>$8.30 \times 10^{-4}$</td>
<td>$7.68 \times 10^{-4}$</td>
<td>$1.06 \times 10^{-1}$</td>
<td>$8.30 \times 10^{-2}$</td>
<td>$2.43 \times 10^{-2}$</td>
</tr>
<tr>
<td>$10^{-4}$</td>
<td>$2.30 \times 10^{-4}$</td>
<td>$1.81 \times 10^{-4}$</td>
<td>$3.20 \times 10^{-4}$</td>
<td>$1.07 \times 10^{-1}$</td>
<td>$8.41 \times 10^{-2}$</td>
<td>$3.20 \times 10^{-2}$</td>
</tr>
<tr>
<td>$10^{-5}$</td>
<td>$5.14 \times 10^{-5}$</td>
<td>$4.07 \times 10^{-5}$</td>
<td>$1.29 \times 10^{-4}$</td>
<td>$1.11 \times 10^{-1}$</td>
<td>$8.77 \times 10^{-2}$</td>
<td>$4.08 \times 10^{-2}$</td>
</tr>
<tr>
<td>$10^{-6}$</td>
<td>$1.14 \times 10^{-5}$</td>
<td>$9.09 \times 10^{-6}$</td>
<td>$5.08 \times 10^{-5}$</td>
<td>$1.14 \times 10^{-1}$</td>
<td>$9.09 \times 10^{-2}$</td>
<td>$5.08 \times 10^{-2}$</td>
</tr>
<tr>
<td>$10^{-7}$</td>
<td>$2.56 \times 10^{-6}$</td>
<td>$2.04 \times 10^{-6}$</td>
<td>$2.08 \times 10^{-5}$</td>
<td>$1.19 \times 10^{-1}$</td>
<td>$9.47 \times 10^{-2}$</td>
<td>$6.59 \times 10^{-2}$</td>
</tr>
</tbody>
</table>

Table 2. Regularization parameters for example 1.

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$\alpha_3$</th>
<th>$\alpha_{1/2}$</th>
<th>$\alpha_{2/3}$</th>
<th>$\alpha_{1/3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-2}$</td>
<td>$9.69 \times 10^{-1}$</td>
<td>$7.54 \times 10^{-1}$</td>
<td>$1.08 \times 10^{-1}$</td>
<td>$2.09e+1$</td>
<td>$1.62e+1$</td>
<td>$1.08e+1$</td>
</tr>
<tr>
<td>$10^{-3}$</td>
<td>$2.07 \times 10^{-1}$</td>
<td>$1.61 \times 10^{-1}$</td>
<td>$1.10 \times 10^{-2}$</td>
<td>$2.07e+1$</td>
<td>$1.61e+1$</td>
<td>$1.10e+1$</td>
</tr>
<tr>
<td>$10^{-4}$</td>
<td>$4.50 \times 10^{-2}$</td>
<td>$3.50 \times 10^{-2}$</td>
<td>$1.16 \times 10^{-3}$</td>
<td>$2.09e+1$</td>
<td>$1.62e+1$</td>
<td>$1.16e+1$</td>
</tr>
<tr>
<td>$10^{-5}$</td>
<td>$9.97 \times 10^{-3}$</td>
<td>$7.74 \times 10^{-3}$</td>
<td>$1.27 \times 10^{-4}$</td>
<td>$2.15e+1$</td>
<td>$1.67e+1$</td>
<td>$1.27e+1$</td>
</tr>
<tr>
<td>$10^{-6}$</td>
<td>$2.20 \times 10^{-3}$</td>
<td>$1.70 \times 10^{-3}$</td>
<td>$1.42 \times 10^{-5}$</td>
<td>$2.20e+1$</td>
<td>$1.70e+1$</td>
<td>$1.42e+1$</td>
</tr>
<tr>
<td>$10^{-7}$</td>
<td>$4.81 \times 10^{-4}$</td>
<td>$3.71 \times 10^{-4}$</td>
<td>$1.66 \times 10^{-6}$</td>
<td>$2.23e+1$</td>
<td>$1.72e+1$</td>
<td>$1.66e+1$</td>
</tr>
</tbody>
</table>

The method (4.4) of performing one Gauss–Newton step has been applied in different parameter identification problems for differential equations and can also be formulated for the corresponding infinite-dimensional problems (see [6, 28]). After having computed the regularized approximation $c^\delta$, the computation of function values $d_i(c) (i = 1, 2, 3)$ can also be done by sparse matrix computations according to

$$d_1(c) = \sqrt{c(b_2, b_3)}, \quad d_2(c) = \frac{(b_2, b_3)}{\|b_3\|}, \quad d_3(c) = \|b_3\|,$$

with $b_1 = T^{-1}u - u^d, b_2 = T(aTu)^{-1}b, b_3 = [T^2 + \alpha Tu]^{-1}b$, and $b_4 = T_u b_3$.

In our numerical experiments the starting vector $c_0 := \xi = (\xi(x_j))_{j=1}^n \in \mathbb{R}^n$ has been used and the regularization parameters $\alpha_1, \alpha_2$ and $\alpha_3$ have been chosen according to rule 1 and rule 2 and Morozov’s discrepancy principle (rule 3), respectively, where rule 3 reads as follows.

**Rule 3.** Choose the regularization parameter $\alpha$ as the solution of the equation $d_3(c) := \|F(x^\delta) - y^d\| = C\delta$.

The corresponding nonlinear equations $d_i(c) = C\delta$ ($i = 1, 2, 3$) have been solved by the secant method and for $C$ the constant $C = 1.1$ has been used. For the dimension number $n$ we used in all experiments $n = 400$. Furthermore, for the error values $e_1 := \|c_{\alpha_1} - c\|, \quad e_2 := \|c_{\alpha_2} - c\|, \quad e_3 := \|c_{\alpha_3} - c\|$ we performed 20 experiments (with 20 different noisy data vectors $u^d$) and the given error values as well as the given regularization parameters in tables 1 and 2 are mean values.

The results in table 1 verify the theoretical results of the estimate (3.34) of theorem 3.8 which tells us that for $\delta = r^2$ holds $e_1 = O(\delta^{2/3})$ and shows the well known fact that $e_3 = O(\delta^{1/2})$. In addition we observed that $e_1 > e_2$ always holds true which is in agreement with the linear case (see [32]). Note that for $\delta = 10^{-7}$ the accuracy for rule 2 is ten times better than the accuracy for Morozov’s discrepancy principle. The results in table 2 show that the regularization parameters $\alpha_1, \alpha_2$
and $\alpha_3$ of the rules 1–3 are related by $\alpha_3 < \alpha_2 < \alpha_1$, that $\alpha_1$ is of the order $O(\delta^{2/3})$ and that $\alpha_3$ is of the order $O(\delta)$. Table 2 also shows that $\alpha_2$ is of the order $O(\delta^{2/3})$ which, however, could not be supported theoretically. For some further numerical experiments for comparing rule 1 and rule 2 see [27].

References

[22] Louis A K 1989 Inverse and Schlecht Gestellte Probleme (Stuttgart: Teubner)
Tikhonov regularization and a posteriori rules


