Georg-August-Universität Göttingen

A Lepskij-type stopping rule for regularized Newton methods

Frank Bauer, Thorsten Hohage

Preprint Nr. 2005-10

Preprint-Serie des
Instituts für Numerische und Angewandte Mathematik
Lotzestr. 16-18
D - 37083 Göttingen

A Lepskij-type stopping rule for regularized Newton methods

Frank Bauer†‡, Thorsten Hohage‡

†University of Göttingen, Institute for Mathematical Stochastics,
Maschmühlenweg 8-10, 37073 Göttingen, Germany
‡University of Göttingen, Institute for Numerical and Applied Mathematics,
Lotzestr. 16-18, 37083 Göttingen, Germany

E-mail: `bauer@mathematik.uni-kl.de`, `hohage@math.uni-goettingen.de`

Abstract. We investigate an a-posteriori stopping rule of Lepskij-type for a class of regularized Newton methods and show that it leads to order optimal convergence rates for Hölder and logarithmic source conditions without a-priori knowledge of the smoothness of the solution. Numerical experiments show that this stopping rule yields at least as good, and in some situations significantly better results than Morozov's discrepancy principle.

AMS classification scheme numbers: 49N45 65J20

1. Introduction

Many inverse problems in partial differential equations and other fields can be formulated as nonlinear operator equations

$$F(x^\dagger) = y \tag{1}$$

with an injective operator $F : D(F) \subset \mathcal{X} \to \mathcal{Y}$ between Hilbert spaces $\mathcal{X}, \mathcal{Y}$. We assume that F is Fréchet differentiable on its domain $D(F)$ and that the measured data y^δ are perturbed by noise with known noise level δ, i.e.

$$\|y - y^\delta\| \leq \delta. \tag{2}$$

Often such problems are ill-posed in the sense that the solution does not depend continuously on the data, i.e. F^{-1} is not continuous. One of the most successful methods to solve such problem given a sufficiently good initial guess x_0 is the *iteratively regularized Gauß-Newton method* (IRGNM) suggested by Bakushinskii [1]. The n-th step of this methods consists in applying Tikhonov regularization to the linearized operator equation $F'[x_n^\delta](x_{n+1}^\delta - x_n^\delta) = y^\delta - F(x_n^\delta)$ with the initial guess x_0:

$$x_{n+1}^\delta = \operatorname{argmin}_{x \in \mathcal{X}} \left(\|F'[x_n^\delta](x_{n+1}^\delta - x_n^\delta) - y^\delta + F(x_n^\delta)\|^2 + \alpha_n \|x_{n+1}^\delta - x_0\|^2 \right). \tag{3}$$

If the penalty term $\|x_{n+1}^\delta - x_0\|^2$ is replaced by $\|x_{n+1}^\delta - x_n^\delta\|^2$, this is called the *Levenberg-Marquardt* algorithm. For a discussion of this and other iterative regularization methods we refer to the recent monograph by Kaltenbacher, Neubauer & Scherzer [11]. For simplicity we only consider the following a-priori choice of the regularization parameters α_n:

$$\alpha_n = \alpha_0 q^n, \qquad q \in (0,1), \alpha_0 > 0 \tag{4}$$

As usual, the regularity of the solution relative to the smoothing properties of the operator is measured in terms of source conditions of the form

$$x_0 - x^\dagger = \Lambda(F'[x^\dagger]^* F'[x^\dagger])w, \qquad \|w\| \leq \varrho \tag{5}$$

where $\Lambda : [0, \|F'[x^\dagger]\|^2] \to \mathbb{R}$ is a monotonic increasing function satisfying $\Lambda(0) = 0$. Later we will assume that ϱ is sufficiently small. In particular, we will consider Hölder-type source conditions

$$x_0 - x^\dagger = (F'[x^\dagger]^* F'[x^\dagger])^\mu w, \qquad \|w\| \leq \varrho, \mu > 0 \tag{6}$$

and logarithmic source conditions

$$x_0 - x^\dagger = \ln(F'[x^\dagger]^* F'[x^\dagger])^{-\mu} w, \qquad \|w\| \leq \varrho, \mu > 0. \tag{7}$$

It is always assumed that the scaling condition

$$\|F'[x^\dagger]\| \leq C_{\mathrm{s}} < 1 \tag{8}$$

is satisfied, which can always be achieved by rescaling the norm in $\mathcal{Y}$. Then in particular the right hand side of (7) is well-defined.

The convergence of the IRGNM for Hölder source conditions with the optimal rate $\mathrm{O}\left(\delta^{2\mu/(2\mu+1)}\right)$ was proved by Bakushinskii [1] for $\mu = 1$ and by Kaltenbacher, Neubauer

& Scherzer [3] for $\mu \in [0,1]$ (with $\mathrm{o}(\delta)$ for $\mu = 0$). For infinitely smoothing operators F logarithmic source conditions typically correspond to natural smoothness conditions in terms of Sobolev space, see [7, 8]. Under these weaker regularity assumptions the rate of convergence of the IRGNM is $\mathrm{O}\left((-\ln\delta)^{-p}\right)$, $\delta \to 0$.

Note that for linear operators F the n-th step of the IRGNM (3) reduces to ordinary Tikhonov regularization with regularization parameter α_{n-1} and initial guess x_0. The previous iterates are not used at all. From this observation it is clear that the IRGNM can only yield optimal rates of convergence for Hölder source conditions (6) with $\mu \le 1$ since the qualification of Tikhonov regularization is $\mu_0 = 1$ (see [5]). To obtain optimal rates of convergence for $\mu > 1$, Kaltenbacher [9] suggested to replace Tikhonov regularization by a regularization method with higher qualification and consider iteration schemes of the form

$$x_{n+1}^\delta := x_0 + g_{\alpha_n}\left(F'[x_n^\delta]^* F'[x_n^\delta]\right) F'[x_n^\delta]^* \Big(y^\delta - F(x_n^\delta) + F'[x_n^\delta](x_n^\delta - x_0) \Big). \tag{9}$$

In particular, for K-times iterated Tikhonov regularization we have $g_\alpha(\lambda) = \frac{(\lambda+\alpha)^K - \alpha^K}{\lambda(\lambda+\alpha)^K}$. Recall that the qualification of iterated Tikhonov regularization is $\mu_0 = K$, and that only one matrix has to be inverted in each Newton step (see [5]). It is also possible to use Landweber iteration, which has qualification $\mu_0 = \infty$ and corresponds to the choice $g_{(1+k)^{-1}}(\lambda) = \sum_{j=0}^{k-1}(1-\lambda)^j$. The stopping index of the interior iteration plays the role of the regularization parameter with the identification $\alpha = 1/(1+k)$. Therefore, with the choice (4) the number of inner iterations grows exponentially.

A crucial ingredient of any iterative regularization method is an appropriate stopping rule. As a first step, one may consider *a-priori stopping rules* which depend on the smoothness of the solution, e.g. the index μ in (6), but not on the data y^δ (see [3, 8] for the IRGNM and [9, 11] for the methods (9)). However, the smoothness of the solution is usually not known a-priori, and therefore a-priori stopping rules do not yield optimal rates of convergence as $\delta \searrow 0$. To obtain optimal rates over a range of smoothness classes without a-priori information on the smoothness, it is necessary to use *a-posteriori stopping rules* which depend not only on δ, but also on y^δ. It has been shown that Morozov's discrepancy principle for the IRGNM leads to order-optimal convergence rates for Hölder source conditions with $\mu \le \frac{1}{2}$ ([3]) and logarithmic source condition ([7]). Moreover, it was shown in [10] that the Newton-Landweber method with the discrepancy principle converges of optimal order for Hölder source conditions with any $\nu > 0$. All these results require strong nonlinearity conditions on the operator F (see (21) below) which cannot be verified for many interesting problems.

To the best of our knowledge it is not known so far how to obtain order optimal rates for $\mu \ge 1/2$ using just Lipschitz continuity of F' without prior knowledge of μ. Moreover, even under stronger nonlinearity assumptions it is an open problem how to achieve rates of optimal order for smoothness classes with index $\mu \in (\mu_0 - 1/2, \mu_0]$ without prior knowledge of μ. These two problems will be solved using the Lepskij-type stopping rule proposed in this paper.

The principle of choosing the regularization parameter adaptively such that the propagated data noise error is roughly of the same size as other error terms has been introduced by Lepskij [13] for estimating functions in white noise. Later it has been applied to linear inverse problems with random and deterministic noise by Goldenshluger & Pereverzev [6], Mathé & Pereverzev [14] and Bauer & Pereverzev [2].

The plan of this paper is as follows. In section 2 we introduce our Lepskij-type stopping rule for iterative regularization methods (not necessarily of Newton-type) for nonlinear inverse problems and prove a general convergence theorem. In the following sections we demonstrate that the assumptions of this theorem are satisfied for the methods discussed above under certain conditions. In section 3 we require Lipschitz continuity of the Fréchet derivative F' and obtain order optimal rates of convergence as $\delta \searrow 0$ for $\mu \geq 1/2$. In section 4 we study Newton methods of the form (9) under the stronger nonlinearity condition (21) for Hölder source conditions with $\mu < 1/2$ and logarithmic source conditions and show that the Lepskij stopping rule also leads to order optimal convergence rates in this situation. Finally, in section 5 we report on tests of the proposed stopping rule for a number of inverse problems in partial differential equations.

2. Lepskij stopping rule

In this section we introduce the Lepskij-type stopping rule and prove a general convergence theorem. Later we will show that the assumptions of this theorem are satisfied for the regularized Newton methods discussed in the introduction.

Assumption 2.1 *Let x_n^δ be the sequence of iterates produced by an iterative regularization method for an initial guess x_0 from some admissible set and data (δ, y^δ) satisfying (2) for $y = F(x^\dagger)$. We assume that*

- *There exists an a-priori known index $N_{\max} = N_{\max}(\delta) \in \mathbb{N}_0$ such that x_n^δ is well defined for $0 \leq n \leq N_{\max}$.*
- *There exists an "optimal" stopping index $N \in \{0, 1, \ldots, N_{\max}\}$ and a known increasing function $\mathcal{E} : \mathbb{N}_0 \to [0, \infty)$ such that*

$$\|x_n^\delta - x^\dagger\| \leq \mathcal{E}(n)\delta \qquad \text{for } n = N, \ldots, N_{\max}. \tag{10}$$

The function $\mathcal{E}$ will be chosen such that $\frac{1}{2}\mathcal{E}(n)\delta$ is a bound on the propagated data noise error after n iteration steps. In principle the "optimal" stopping index N in Assumption 2.1 can be arbitrary, but we will always define it by some a-priori stopping rule $N(\delta, \mu, \varrho)$ for a range of smoothness classes $S_{\mu,\varrho} \subset \mathcal{X}$ given by (6) or (7) such that

$$\sup_{x^\dagger \in S_{\mu,\varrho}} \sup_{\|y^\delta - F(x^\dagger)\| \leq \delta} \mathcal{E}(N(\delta, \mu, \varrho))\delta \tag{11}$$

is an order optimal bound on the error for all μ and ϱ with $\varrho > 0$ sufficiently small. Then assumption (10) means that the total error after the "optimal" stopping index N and before the maximal iteration index $N_{\max}(\delta)$ is dominated by the propagated data noise error. For linear problems this holds true for any sufficiently large n, but for nonlinear problems a blow-up due to nonlinearity errors may occur. The crucial point is that there exists an a-priori known iteration index $N_{\max}(\delta) \geq N$ up to which the error due to the nonlinearity of the operator F is negligible compared to other error terms.

Definition 2.1 (Lepskij stopping rule) *Under assumption 2.1 define $n_* = n_*(\delta, y^\delta)$ by*

$$n_* := \min \left\{ n \in \{0, \dots, N_{\max}(\delta)\} : \begin{array}{c} \|x_n^\delta - x_m^\delta\| \leq 2\mathcal{E}(m)\delta \\ \text{for all } m = n+1, \dots, N_{\max}(\delta) \end{array} \right\}.$$

Note that the implementation of the Lepskij stopping rule does not require the knowledge of the "optimal" stopping index N.

Theorem 2.1 *Under Assumption 2.1 the error at the stopping index n_* in Definition 2.1 satisfies*

$$\|x_{n_*}^\delta - x^\dagger\| \leq 3\mathcal{E}(N)\delta.$$

Proof
Since $\mathcal{E}$ is increasing, we have

$$\|x_m^\delta - x_N^\delta\| \leq \|x^\dagger - x_m^\delta\| + \|x^\dagger - x_N^\delta\| \leq \mathcal{E}(m)\delta + \mathcal{E}(N)\delta \leq 2\mathcal{E}(m)\delta$$

for $m = N+1, \dots, N_{\max}(\delta)$. This implies $n_* \leq N$. Therefore,

$$\|x^\dagger - x_{n_*}^\delta\| \leq \|x^\dagger - x_N^\delta\| + \|x_N^\delta - x_{n_*}^\delta\| \leq \mathcal{E}(N)\delta + 2\mathcal{E}(N)\delta = 3\mathcal{E}(N)\delta,$$

and the proof is complete.

3. Hölder source conditions with $\mu \geq \frac{1}{2}$

In this section we verify Assumption 2.1 for the Hölder source conditions (6) with $\mu \geq \frac{1}{2}$. The only condition on the operator F will be that the Fréchet derivative F' satisfy the Lipschitz condition

$$\|F'[x_1] - F'[x_2]\| \leq L\|x_1 - x_2\| \tag{12}$$

for all x_1, x_2 in $D(F)$. Moreover, we assume that the family of functions g_α defining the linear regularization method in (9) satisfies

$$\sup_{\lambda \in [0, \|F'[x^\dagger]\|^2]} |\sqrt{\lambda} g_\alpha(\lambda)| \leq \frac{C_{\mathrm{g}}}{\sqrt{\alpha}} \tag{13}$$

for all α with some constant C_{g}. This holds true with $C_{\mathrm{g}} = 1/2$ for Tikhonov regularization and $C_{\mathrm{g}} = 1$ for Landweber iteration. We cite the following recursive estimates for the errors

$$E_n = x^\dagger - x_n^\delta \tag{14}$$

from the monograph [11] which can be deduced form the definition (9) (see Lemma 4.10 for the IRGNM and the proof of Theorem 4.16 and Corollary 4.17 for general methods (9)).

Lemma 3.1 *Let x_n^δ be the iterates defined by the Newton method (9) with (iterated) Tikhonov regularization or Landweber iteration. Let μ denote the qualification of the linear method, and let C_{g} be defined by (13). Assume that (1), (2), (6) with $1/2 \le \mu \le \mu_0$, and (12) hold true and that $x_n^\delta \in B_R(x^\dagger)$. Then*

$$\|E_{n+1}\| \le C_1 \alpha_n^\mu \varrho + C_2 \varrho \|E_n\| + \frac{C_3}{\sqrt{\alpha_n}} \|E_n\|^2 + C_{\mathrm{g}} \frac{\delta}{\sqrt{\alpha_n}} \tag{15}$$

with constants C_1, C_2, and C_3 ($C_3 = \frac{1}{2} L C_{\mathrm{g}}$) independent of n, ϱ, δ, and $x^\dagger$.

To analyze the behavior of $\|E_n\|$ it is convenient to rewrite this in terms of powers of q defined in (4):

$$\|E_n\| = q^{e_n} \qquad \text{or} \qquad e_n := \frac{\ln \|E_n\|}{\ln q}$$

(for $\|E_n\| = 0$ we set $e_n := \infty$). We rewrite the other constants and quantities using this constant q as well, namely $\alpha_n = q^{n+n_0}$, $C_1 = \frac{1}{4} q^{\gamma_1}$, $C_2 = \frac{1}{4} q^{\gamma_2}$, $C_3 = \frac{1}{4} q^{\gamma_3}$, $C_g = \frac{1}{4} q^{\gamma_g}$, $\varrho = q^{r - \min\{\gamma_1, \gamma_2\}}$ and $\delta = q^{d - \gamma_g}$. With this notation we have

$$\begin{aligned} q^{e_{n+1}} &\le \frac{1}{4} \left(q^{\mu n + \mu n_0 + r} + q^{r + e_n} + q^{\gamma_3 - \frac{1}{2} n - \frac{1}{2} n_0 + 2 e_n} + q^{d - \frac{1}{2} n - \frac{1}{2} n_0} \right) \\ &\le \max \left\{ q^{\mu n + \mu n_0 + r}, q^{r + e_n}, q^{\gamma_3 - \frac{1}{2} n - \frac{1}{2} n_0 + 2 e_n}, q^{d - \frac{1}{2} n - \frac{1}{2} n_0} \right\} \\ &= q^{\min\left\{ \mu n + \mu n_0 + r, r + e_n, \gamma_3 - \frac{1}{2} n - \frac{1}{2} n_0 + 2 e_n, d - \frac{1}{2} n - \frac{1}{2} n_0 \right\}} \end{aligned}$$

or equivalently

$$e_{n+1} \ge \min \left\{ \mu(n + n_0) + r, r + e_n, \gamma_3 - \frac{1}{2}(n + n_0) + 2 e_n, d - \frac{1}{2}(n + n_0) \right\}.$$

In comparison to the original recursion inequality this one is easier to analyze.

Theorem 3.2 *Let x_n^δ be the iterates defined by the Newton method (9) with (iterated) Tikhonov regularization or Landweber iteration, and let μ_0 denote the qualification of the linear method, i.e. $\mu_0 = K$ for K-times iterated Tikhonov regularization and $\mu_0 = \infty$ for Landweber iteration. Assume that (1), (2), (6) with $1/2 \le \mu \le \mu_0$, and (12) hold true. Moreover, assume that δ, ϱ in (6) and L in (12), and $1/\alpha_0 = q^{-n_0}$ in (4) are sufficiently small (explicit bounds are given in the proof). Then Assumption 2.1 is satisfied for*

$$N - 1 = \max\left(0, \left\lfloor \frac{d - r}{\mu + \frac{1}{2}} - n_0 \right\rfloor\right), \qquad N_{\max} = \max\left(0, \lfloor d + \gamma_3 - n_0 \rfloor\right) \tag{16}$$

and $\mathcal{E}(n) := 4 C_{\mathrm{g}} \alpha_n^{-1/2}$.

In particular, the error at the Lepskij-stopping index satisfies

$$\|x_{n_*}^\delta - x^\dagger\| \le C \varrho^{\frac{1}{2\mu+1}} \delta^{\frac{2\mu}{2\mu+1}} \tag{17}$$

for some constant C independent of δ, ϱ, and $x^\dagger$.

Recall that for linear problems the optimal error bound under the source condition (6) is $\varrho^{\frac{1}{2\mu+1}} \delta^{\frac{2\mu}{2\mu+1}}$ (see e.g. [5]).

Proof

Preliminaries: We assume that

$$n_0 \le -\frac{\min\{\gamma_1,\gamma_2\}}{\mu} + 1, \tag{18a}$$

$$r \ge \max\left\{\mu, -\left(\mu - \frac{1}{2}\right) n_0 - \gamma_3 + 2\mu, -\gamma_2\left(\mu + \frac{1}{2}\right)\right\}, \tag{18b}$$

$$d \ge 2\mu(k-1) + 2r + \gamma_3, \tag{18c}$$

$$B\left(4(\alpha_0/q)^\mu \max(C_1,C_2), x^\dagger\right) \subset D(F) \tag{18d}$$

specifying the smallness conditions in the theorem. Note that (18*b*) implies in particular that $N-1 \le N_{\max}$.

The proof consists of three steps:

(i) x^δ_{n+1} well-defined for $-1 \le n \le N-1$, and

$$e_{n+1} \ge \mu(n+n_0) + r \qquad \text{for } n = -1,\dots,N-1. \tag{19}$$

(ii) x^δ_{n+1} well-defined for $N \le n \le N_{\max}-1$, and

$$e_{n+1} \ge -\frac{1}{2}(n+n_0+1) + d \qquad \text{for } N-1 \le n \le N_{\max}-1. \tag{20}$$

(iii) $\|E_{n^*}\| \le C\varrho^{\frac{1}{2\mu+1}}\delta^{\frac{2\mu}{2\mu+1}}$.

Step 1: To prove the statement for $n=-1$ note that $\|E_0\| = \|(F'[x^\dagger]^*F'[x^\dagger])^\mu w\| \le \|w\| \le \varrho$ due to the scaling condition (8). Therefore, $e_0 \ge r - \min\{\gamma_1,\gamma_2\} \ge r + \mu(n_0-1)$ by the assumption (18*a*) on n_0.
Suppose that (19) holds for $n-1$. Then x^δ_n belongs to $D(F)$ by assumption (18*d*) (i.e. x^δ_n is well defined) since $v_n \ge \mu(n-1+n_0)+r \ge \mu(n_0-1)+r$ and hence $\|E_n\| \le (\alpha_0/q)^\mu 4\max(C_1,C_2)$. Moreover, it follows from the induction hypothesis, the definition of N, $n \le N-1$ and the assumption (18*b*) that

$$\begin{aligned}
r + e_n &\ge r + \mu(n+n_0-1) + r = \mu(n+n_0) + r + (r-\mu)\\
&\ge \mu(n+n_0)+r\\
\gamma_3 - \frac{1}{2}(n+n_0) + 2e_n &\ge \gamma_3 - \frac{1}{2}(n+n_0) + 2\mu(n+n_0) + 2r - 2\mu\\
&= \mu(n+n_0) + r + \left(\left(\mu-\frac{1}{2}\right)(n+n_0) + r + \gamma_3 - 2\mu\right)\\
&\ge \mu(n+n_0)+r\\
d - \frac{1}{2}(n+n_0) &= \mu(n+n_0) + r + \left(-\left(\frac{1}{2}+\mu\right)(n+n_0) + d - r\right)\\
&\ge \mu(n+n_0) + r + \left(-\left(\frac{1}{2}+\mu\right)\frac{d-r}{\mu+\frac{1}{2}} + d - r\right)\\
&= \mu(n+n_0)+r
\end{aligned}$$

Taking the minimum of these terms yields the inequality (19) for n.

Step 2: The start of induction is fulfilled because of (19):

$$\begin{aligned} e_N &\geq \mu(n_0+N-1)+r \\ &= \left(\mu+\frac{1}{2}\right)\left(\left\lfloor\frac{d-r}{\mu+\frac{1}{2}}\right\rfloor - \frac{d-r}{\mu+\frac{1}{2}}+1\right) - \frac{1}{2}(n+n_0)+d \\ &\geq -\frac{1}{2}(n+n_0)+d \end{aligned}$$

Suppose that (20) holds for $n-1$. We first show that x^δ_{n+1} is well-defined. As $\|E_n\| \leq q^{-(n+n_0+1)/2+d}$ by the induction hypothesis and

$$-\frac{1}{2}(n+n_0+1)+d \geq -\frac{1}{2}(N_{\max}+n_0+1)+d \geq d/2-\gamma_3/2-\frac{1}{2} \geq \mu(n_0-1)+r$$

by the smallness assumption (18c), we obtain $\|E_n\| \leq (\alpha_0/q)^\mu 4\max(C_1,C_2)$, so $x^\delta_n \in D(F)$. Next, using the induction hypothesis, the definition of $N_{\max}$ and N, and $N-1<n\leq N_{\max}$ we obtain

$$\begin{aligned} \mu(n+n_0)+r &= -\frac{1}{2}(n+n_0+1)+d+\left(\left(\mu+\frac{1}{2}\right)(n+n_0)+r-d+\frac{1}{2}\right) \\ &\geq -\frac{1}{2}(n+n_0+1)+d+\left(\left(\mu+\frac{1}{2}\right)\frac{d-r}{\mu+\frac{1}{2}}+r-d\right) \\ &\geq -\frac{1}{2}(n+n_0+1)+d \\ r+e_n &\geq r-\frac{1}{2}(n+n_0+1)+d \geq -\frac{1}{2}(n+n_0+1)+d \\ \gamma_3-\frac{1}{2}(n+n_0)+2e_n &\geq \gamma_3-\frac{1}{2}(n+n_0)-(n+n_0+1)+2d \\ &\geq -\frac{1}{2}(n+n_0+1)+d+(\gamma_3-(n+n_0)+d) \\ &\geq -\frac{1}{2}(n+n_0)+d. \end{aligned}$$

Taking the minimum of these terms yields (20) for n.

Step 3: The inequality (20) is equivalent to (10) in Assumption 2.1 since

$$\|E_n\| = q^{e_n} \leq q^{-\frac{1}{2}(n+n_0)+d} = 4C_{\mathrm{g}}\frac{\delta}{\sqrt{\alpha_n}} = \mathcal{E}(n)\delta, \quad n=N,\ldots,N_{\max}.$$

It follows from the first two steps that Assumption 2.1 is satisfied. Therefore, Theorem 2.1 yields

$$\|E_{n^*}\| \leq 3\mathcal{E}(N)\delta \leq Cq^{-\frac{1}{2}(N+n_0)+d} \leq Cq^{-\frac{1}{2}\frac{d-r}{\mu+\frac{1}{2}}+d} \leq C\varrho^{\frac{1}{2\mu+1}}\delta^{\frac{2\mu}{2\mu+1}}$$

for a generic constant C independent of δ, ϱ, and $x^\dagger$.

4. Weak source conditions

In this section we study Newton methods under Hölder source conditions with $\mu < \frac{1}{2}$ and under logarithmic source conditions. In this case we need a stronger condition restricting the degree of nonlinearity of the operator F. We assume that there exist mappings $R : D(F) \times D(F) \to L(Y,Y)$, $Q : D(F) \times D(F) \to L(X,Y)$ and constants $\gamma_R, \gamma_Q > 0$ such that

$$\begin{aligned} &F'(\bar{x}) = R(\bar{x},x)F'(x) + Q(\bar{x},x) \\ &\|I - R(\bar{x},x)\| \le \gamma_R\|\bar{x} - x\|, \\ &\|Q(\bar{x},x)\| \le C_Q\|F'(x^\dagger)(\bar{x} - x)\| \end{aligned} \tag{21}$$

for all $x, \bar{x} \in B_R(x^\dagger)$. For a discussion of this condition and examples where it is satisfied we refer to [11].

Again, a crucial tool will be recursive inequalities for the norm of the error $E_n := x_n^\delta - x^\dagger$, and this time also for $\|F'[x^\dagger]e_n\|$. These estimates are derived in [11] (see eqs. (4.98) and (4.102) in the proof of Theorem 4.16).

Lemma 4.1 *Let x_n^δ be the iterates defined by the Newton method (9) with (iterated) Tikhonov regularization or Landweber iteration, and let C_g be defined by (13). Assume that (1), (2), and (21) hold true, and let $T := F'[x^\dagger]$. Moreover, assume that $x_n^\delta \in B_R(x^\dagger)$ and a source condition (5) with $\Lambda(t) = t^\mu$, $\mu \in (0, 1/2)$ or $\Lambda(t) = (-\ln t)^{-mu}$, $\mu \in (0,\infty)$ is satisfied. Then*

$$\|E_{n+1}\| \le C_4\Lambda(\alpha_n)\varrho + c_5\varrho\|E_n\| + c_6(\varrho + \|E_n\|)\frac{\|TE_n\|}{\sqrt{\alpha_n}} + C_{\mathrm{g}}\frac{\delta}{\sqrt{\alpha_n}} \tag{22a}$$

and

$$\begin{aligned} \|TE_{n+1}\| \le\ & C_7\Lambda(\alpha_n)\sqrt{\alpha_n}\varrho + c_8\varrho\sqrt{\alpha_n}\|E_n\| + c_9\varrho\|TE_n\| \\ &+ \left(c_{10} + c_{11}\frac{\|TE_n\|}{\sqrt{\alpha_n}}\right)\|E_n\| \cdot \|TE_n\| + \left(C_{12} + c_{13}\frac{\|TE_n\|}{\sqrt{\alpha_n}}\right)\delta \end{aligned} \tag{22b}$$

where the constants $c_5, c_6, c_8, c_9, c_{10}, c_{11}$ and c_{13} are small if γ_R and γ_Q are small.

To analyze the behavior of $\|E_n\|$ and $\|TE_n\|$ we again rewrite the inequalities (4.1) in terms of powers of q by introducing $e_n, s_n \in \mathbb{R} \cup \{\infty\}$ such that

$$\|E_n\| = q^{e_n} \qquad \text{and} \qquad \|TE_n\| = q^{s_n}.$$

We also rewrite the other constants and quantities using this constant q, namely $\alpha_n = q^{n+n_0}$, $\Lambda(\alpha_n) = q^{l(n+n_0)}$, $\varrho\max\{C_4, c_5, c_6, C_7, c_8, c_9, 1\} = \frac{1}{7}q^r$ and $\delta C_g = \frac{1}{7}q^d$ and $C_{12} = \frac{1}{2}C_g q^{\widetilde{\gamma}}$. Furthermore we assume for our convenience that $c_6 \le \frac{1}{4}$, $c_{10} \le \frac{1}{7}$, $c_{11} \le \frac{1}{7}$, $c_{13} \le \frac{1}{2}C_g$ and $\gamma = \min\{\widetilde{\gamma}, -1\}$. Then

$$\begin{aligned} e_{n+1} \ge \min\Big\{ & l(n+n_0) + r, r + e_n, r + s_n - \frac{1}{2}(n+n_0), \\ & e_n + s_n - \frac{1}{2}(n+n_0), d - \frac{1}{2}(n+n_0)\Big\} \end{aligned} \tag{23a}$$

$$\begin{aligned} s_{n+1} \ge \min\{ & l(n+n_0) + \frac{1}{2}(n+n_0) + r, r + \frac{1}{2}(n+n_0) + e_n, r + s_n, \\ & e_n + s_n, e_n + 2s_n - \frac{1}{2}(n+n_0), d + \gamma, s_n - \frac{1}{2}(n+n_0) + d\} \end{aligned} \tag{23b}$$

The analysis of this coupled recursive set of inequalities is a bit more involved:

Theorem 4.2 *Let x_n^δ be the iterates defined by the Newton method (9) with (iterated) Tikhonov regularization or Landweber iteration. Assume that (1), (2), and (21) hold true and that a source condition (5) with $\Lambda(t) = t^\mu$, $\mu \in (0, 1/2)$ or $\Lambda(t) = (-\ln t)^{-\mu}$, $\mu \in (0,\infty)$ is satisfied. Moreover, assume that γ_R, γ_Q in (21), ϱ in (5), and and $1/\alpha_0$ in (4) are sufficiently small (explicit bounds are given in the proof). Then Assumption 2.1 is satisfied if $N \in \mathbb{N}_0$ is chosen such that*

$$\sqrt{\alpha_{N-2}}\Lambda(\alpha_{N-2}) > 7C_{\mathrm{g}}\frac{\delta}{\varrho} \geq \sqrt{\alpha_{N-1}}\Lambda(\alpha_{N-1}), \tag{24}$$

and

$$N_{\max} = \lfloor 2(d+\gamma-1)\rfloor - n_0, \qquad \mathcal{E}(n) := 7C_{\mathrm{g}}\alpha_n^{-1/2}.$$

In particular, the error at the Lepskij-stopping index n_* satisfies

$$\|x_{n_*}^\delta - x^\dagger\| \leq C\varrho^{\frac{1}{2\mu+1}}\delta^{\frac{2\mu}{2\mu+1}} \tag{25}$$

for the Hölder source conditions and

$$\|x_{n_*}^\delta - x^\dagger\| \leq C\varrho\left(-\ln\frac{\delta}{\varrho}\right)^{-\mu} \tag{26}$$

for logarithmic source conditions.

Proof

Preliminaries: We assume that

$$l(n+n_0-1) \geq l(n+n_0) - \frac{1}{2} \qquad \text{for all } n \in \mathbb{N}_0, \tag{27a}$$
$$r \geq \max\{3/2, -\gamma\}, \tag{27b}$$
$$l(n_0) + r - 3/2 \geq 0, \tag{27c}$$
$$B\left(7\varrho\max\{C_4, c_5, c_6, C_7, c_8, c_9\}\Lambda(\alpha_{n_0}), x^\dagger\right) \subset D(F) \tag{27d}$$
$$d/2 - \gamma/2 - \frac{1}{2} \geq l(n_0) + r \tag{27e}$$
$$l(n_0-1) \leq 0 \text{ and } (n_0-1) \leq 0 \tag{27f}$$

specifying the smallness conditions in the theorem.

After rewriting the definition (24) of N as

$$l(N+n_0-2) + \frac{1}{2}(N+n_0-2) < d-r \leq l(N+n_0-1) + \frac{1}{2}(N+n_0-1), \tag{28}$$

it follows from assumption (27*b*) that $N \leq N_{\max}$. Moreover, we define $B \in \mathbb{R}$ implicitly by $l(B+n_0) + \frac{1}{2}(B+n_0) = d+\gamma-r$ and note that $B < N-1$ because $\gamma \leq -1$ and the definition of r.

The proof consists of four steps:

(i) x_{n+1}^δ well-defined for $-1 \leq n \leq B$ and

$$e_{n+1} \geq l(n+n_0) + r, \tag{29a}$$
$$s_{n+1} \geq \frac{1}{2}(n+n_0) + l(n+n_0) + r. \tag{29b}$$

(ii) x^δ_{n+1} well-defined for $B \le n \le N-1$ and

$$e_{n+1} \ge l(n+n_0)+r, \tag{30a}$$
$$s_{n+1} \ge d+\gamma. \tag{30b}$$

(iii) x^δ_{n+1} well-defined for $N-1 \le n \le N_{max}$ and

$$e_{n+1} \ge -\frac{1}{2}(n+n_0+1)+d, \tag{31a}$$
$$s_{n+1} \ge d+\gamma. \tag{31b}$$

(iv) $\|E_{n_*}\| \le C\varrho^{\frac{1}{2\mu+1}}\delta^{\frac{2\mu}{2\mu+1}}$ or $\|E_{n_*}\| \le C\varrho\left(-\ln\frac{\delta}{\varrho}\right)^{-\mu}$, respectively.

Step 1: The first three statements are shown by induction. To prove the first statement for $n=-1$, note that $\|E_0\| = \|(F'[x^\dagger]^*F'[x^\dagger])^\mu w\| \le \|w\| \le \varrho$ due to the scaling condition (8). Therefore, $e_0 \ge l(n_0-1)+r$ and $s_0 \ge \frac{1}{2}(n_0-1)+l(n_0-1)+r$ by the assumption on n_0.

Let $n \in \mathbb{N}_0$ and assume the assertion holds true with n replaced by $n-1$. To prove that x^δ_{n+1} is well-defined, we have to show that x^δ_n belongs to $D(F)$. By assumption (27d), this follows from the induction hypothesis since $\|x_n - x^\dagger\| \le q^{e_n} \le q^{l(n_0)+r} \le 7\varrho\max\{C_4,c_5,c_6,C_7,c_8,c_9\}\Lambda(\alpha_{n_0})$. To prove the inequalities, we consider the terms on the right hand side of (23a) and (23a) separately using the induction hypothesis, the definition of N, $n \le N-1$, and (27a)–(27c):

$$\begin{aligned}
r+e_n &\ge r+l(n+n_0)+r-\frac{1}{2} \ge l(n+n_0)+r,\\
r+s_n-\frac{1}{2}(n+n_0) &\ge r+l(n+n_0)+\frac{1}{2}(n+n_0)+r-1-\frac{1}{2}(n+n_0)\\
&\ge l(n+n_0)+r,\\
e_n+s_n-\frac{1}{2}(n+n_0) &\ge l(n+n_0)+r+\frac{1}{2}(n+n_0)+l(n+n_0)+r-\frac{1}{2}(n+n_0)\\
&\ge l(n+n_0)+r,\\
d-\frac{1}{2}(n+n_0) &\ge l(n+n_0)+r+\left(d-\frac{1}{2}(n+n_0)-l(n+n_0)-r\right)\\
&\ge l(n+n_0)+r
\end{aligned}$$

and

$$\begin{aligned}
r+\frac{1}{2}(n+n_0)+e_n &\ge r+\frac{1}{2}(n+n_0)+l(n+n_0)+r-\frac{1}{2}\\
&\ge l(n+n_0)+\frac{1}{2}(n+n_0)+r,\\
r+s_n &\ge r+l(n+n_0)+\frac{1}{2}(n+n_0)+r-1 \ge l(n+n_0)+\frac{1}{2}(n+n_0)+r\\
e_n+s_n &\ge l(n+n_0)+r+l(n+n_0)+\frac{1}{2}(n+n_0)+r-\frac{3}{2}\\
&\ge l(n+n_0)+\frac{1}{2}(n+n_0)+r,\\
e_n+2s_n-\frac{1}{2}(n+n_0) &\ge l(n+n_0)+r+\frac{1}{2}(n+n_0)+2l(n+n_0)+2r-5/2
\end{aligned}$$

$$\begin{aligned}
&\geq l(n+n_0)+\frac{1}{2}(n+n_0)+r,\\
d+\gamma \quad &= l(n+n_0)+\frac{1}{2}(n+n_0)+r\\
&\qquad -\left(l(n+n_0)+\frac{1}{2}(n+n_0)+r-d-\gamma\right)\\
&\geq l(n+n_0)+\frac{1}{2}(n+n_0)+r,\\
s_n-\frac{1}{2}(n+n_0)+d \quad &\geq l(n+n_0)+\frac{1}{2}(n+n_0)+r-\frac{1}{2}(n+n_0)+d-\frac{1}{2}\\
&= l(n+n_0)+r+d-\frac{1}{2}\\
&\geq l(n+n_0)+\frac{1}{2}(n+n_0)+r.
\end{aligned}$$

Taking the minimum of these terms yields the inequalities (29a) and (29b).

Step 2: For $n = \lfloor B \rfloor$, (30a) and (30b) follow immediately from the first step of the proof and the definition of B. Let $n \in \{\lfloor B \rfloor + 1, \dots, N-1\}$ and assume the assertion holds true with n replaced by $n-1$. It follows as in step 1 of the proof that x_{n+1}^{δ} is well-defined. To prove the inequalities we again consider the terms separately, using the induction hypothesis, $B \leq n \leq N-1$, the definitions of B and N, and (27b):

$$\begin{aligned}
r+e_n \quad &\geq r+l(n+n_0)+r-\frac{1}{2}\geq l(n+n_0)+r,\\
r+s_n-\frac{1}{2}(n+n_0) \quad &\geq r+d+\gamma-\frac{1}{2}(n+n_0)\\
&\geq l(n+n_0)+r+\left(d+\gamma-\frac{1}{2}(n+n_0)-l(n,n_0)\right)\\
&\geq l(n+n_0)+r+\left(d-r-\frac{1}{2}(n+n_0)-l(n,n_0)\right)\geq l(n+n_0)+r,\\
e_n+s_n-\frac{1}{2}(n+n_0) &\geq l(n+n_0)+r+\left(d+\gamma-\frac{1}{2}(n+n_0)\right)\\
&\geq l(n+n_0)+r+\left(d-r-\frac{1}{2}(n+n_0)\right)\geq l(n+n_0)+r,\\
d-\frac{1}{2}(n+n_0) \quad &\geq l(n+n_0)+r+\left(d-\frac{1}{2}(n+n_0)-l(n+n_0)-r\right)\geq l(n+n_0)+r
\end{aligned}$$

and

$$\begin{aligned}
r+\frac{1}{2}(n+n_0)+e_n \quad &\geq r+\frac{1}{2}(n+n_0)+l(n+n_0)+r-\frac{1}{2}\geq d+\gamma,\\
r+s_n \quad &\geq r+d+\gamma\geq d+\gamma,\\
e_n+s_n \quad &\geq l(n+n_0)+r+d+\gamma-\frac{1}{2}\\
&\geq d+\gamma,\\
e_n+2s_n-\frac{1}{2}(n+n_0) &\geq d+\gamma+\left(l(n+n_0)+r+d+\gamma-\frac{1}{2}(n+n_0)\right)-\frac{1}{2}=d+\gamma,\\
s_n-\frac{1}{2}(n+n_0)+d \quad &\geq d+\gamma-\frac{1}{2}(n+n_0)+d\geq d+\gamma.
\end{aligned}$$

Taking the minimum of these terms yields the inequalities (30a) and (30b).

Step 3: For $n = N-1$ we get from (30a) and the definition of N (or (28)) that

$$\begin{aligned} e_N &\geq l(N+n_0-1)+r \\ &= -\frac{1}{2}(N+n_0)+d+l(N+n_0-1)+r+\frac{1}{2}(N+n_0-1)-d+\frac{1}{2} \\ &\geq -\frac{1}{2}(N+n_0)+d, \end{aligned}$$

so (31a) holds true. (31b) follows from the second step of the proof.

Now let $n \in \{N, \ldots, N_{\max}\}$ and assume that the assertion holds true with n replaced by $n-1$. We first show that x_{n+1}^δ is well-defined. By the induction hypothesis we have $\|E_n\| \leq q^{-(n+n_0+1)/2+d}$, and

$$-\frac{1}{2}(n+n_0+1)+d \geq -\frac{1}{2}(N_{\max}+n_0+1)+d \geq d/2-\gamma/2-\frac{1}{2} \geq l(n_0)+r$$

by the definition of $N_{\max}$. Using (27d), this implies $x_n^\delta \in D(F)$. From the induction hypothesis, the definition of $N_{\max}$ and N, $N < n \leq N_{\max}$ and (27b) we obtain

$$\begin{aligned} l(n+n_0)+r &\geq -\frac{1}{2}(n+n_0+1)+d+\left(l(n+n_0)+r+\frac{1}{2}(n+n_0+1)-d\right) \\ &\geq -\frac{1}{2}(n+n_0+1)+d, \\ r+e_n &\geq r-\frac{1}{2}(n+n_0+1)+d+\frac{1}{2} \geq -\frac{1}{2}(n+n_0+1)+d, \\ r+s_n-\frac{1}{2}(n+n_0) &\geq r+d+\gamma-\frac{1}{2}(n+n_0+1)+\frac{1}{2} \geq -\frac{1}{2}(n+n_0+1)+d, \\ e_n+s_n-\frac{1}{2}(n+n_0) &\geq -\frac{1}{2}(n+n_0+1)+d-\frac{1}{2}(n+n_0+1)+d+\gamma \\ &\geq -\frac{1}{2}(n+n_0+1)+d \end{aligned}$$

and

$$\begin{aligned} l(n+n_0)+\frac{1}{2}(n+n_0)+r &\geq d+\gamma, \\ r+\frac{1}{2}(n+n_0)+e_n &\geq r+\frac{1}{2}(n+n_0)-\frac{1}{2}(n+n_0+1)+d+\frac{1}{2} \geq d+\gamma, \\ r+s_n &\geq r+d+\gamma \geq d+\gamma, \\ e_n+s_n &\geq -\frac{1}{2}(n+n_0+1)+d+d+\gamma \geq d+\gamma, \\ e_n+2s_n-\frac{1}{2}(n+n_0) &\geq -\frac{1}{2}(n+n_0+1)+d+2d+2\gamma-\frac{1}{2}(n+n_0) \geq d+\gamma, \\ s_n-\frac{1}{2}(n+n_0)+d &\geq d+\gamma-\frac{1}{2}(n+n_0)+d \geq d+\gamma. \end{aligned}$$

This yields the inequalities (31a) and (31b).

Step 4: Step 3 implies that

$$\|E_n\| \leq q^{-\frac{1}{2}(n+n_0)+d} = 7C_{\mathrm{g}}\frac{\delta}{\sqrt{\alpha_n}} = \mathcal{E}(n)\delta$$

for $n = N, \ldots, N_{\max}$, i.e. (10) in Assumption 2.1 is satisfied.

Therefore, Theorem 2.1 yields for the Hölder source condition:

$$\|E_{n^*}\| \le 3\mathcal{E}(N)\delta \le \widetilde{C} q^{-\frac{1}{2}(N+n_0)+d} \le \hat{C} q^{-\frac{1}{2}\frac{d-r}{\mu+\frac{1}{2}}+d} \le C\varrho^{\frac{1}{2\mu+1}}\delta^{\frac{2\mu}{2\mu+1}}$$

for some constants $C, \hat{C}, \widetilde{C}$ which are independent of $x^\dagger$, δ and ϱ.

For logarithmic source conditions we obtain from the second inequality in (24) that

$$\|E_N\| \le 3\mathcal{E}(N)\delta = 21 \cdot C_{\mathrm{g}} \frac{\delta}{\sqrt{\alpha_N}} \le C\varrho(-\ln \alpha_N)^{-\mu} \tag{32}$$

with C independent of δ and ϱ, and $x^\dagger$. Since $\sqrt{\alpha_{N-2}} > \Lambda(\alpha_{N-2})\sqrt{\alpha_{N-2}} > 7C_{\mathrm{g}}\frac{\delta}{\varrho}$ by the first inequality in (24), we get $\alpha_N \le C(\delta/\varrho)^2$. Plugging this into (32) yields (26).

5. Numerical results

We test the Lepskij stopping rule on a number of examples:

A parameter identification problem with distributed measurements. We consider the identification of the parameter $c \ge 0$ in the boundary value problem

$$\begin{aligned} &-u'' + cu = f \qquad \text{on } (0,1),\\ &u(0) = u_0, \qquad u(1) = u_1 \end{aligned}$$

given measurements of u in $[0,1]$ (see [5]). Here the right-hand side $f \in L^2([0,1])$ and the boundary values $u_0, u_1 \in \mathbb{R}$ are known, in our example $f \equiv 1$, $u_0 = 1$ and $u_1 = 2$. We define the operator $F : D(F) \to L^2([0,1])$ with $D(F) := \{c \in L^2([0,1]) : c \ge 0 \text{ a.e.}\}$, which maps a coefficient $c \in D(F)$ to the corresponding solution $u \in H^2([0,1])$. Here c plays the role of x and u the role of y. This problem was discretized by a finite difference method, and data noise was simulated by adding a centered Gaussian random variable to $u(x_i)$ at each measurement point x_i.

As discussed in [5], the Hölder source condition 6 is satisfied if $c^\dagger - c_0$ is $H^{4\mu}$-smooth and satisfies certain boundary conditions. In particular, for our choice $c^\dagger(x) := 1 + \sin^4(\pi x)\sin(10x)$ and $c_0(x) := 1$, a Hölder source condition with $\mu = 1$ is satisfied. Since Tikhonov regularization has qualification $\mu_0 = 1$, we expect the rate $\mathrm{O}\left(\delta^{2/3}\right)$ for the IRGNM if the proposed Lepskij-type stopping rule is used, but only the rate $\mathrm{O}\left(\delta^{1/2}\right)$ corresponding to $\mu = 1/2$ for the discrepancy principle. This effect can clearly be observed in Fig. 1. A further discussion of the numerical results is given below.

An inverse potential problem. Our next problem concerns the identification of the shape of a heat source $\Omega \subset \mathbb{R}^2$ from measurement of the heat flux $\frac{\partial u}{\partial n}$ and the temperature u on some boundary Γ surrounding the heat source (see [7]). We can arrange things such that $u|_\Gamma = 0$ by subtracting a solution to the Laplace equation. Then the forward problem is described by the boundary value problem

$$\begin{aligned} &\Delta u = \chi_\Omega,\\ &u = 0 \qquad \text{on } \Gamma \end{aligned}$$

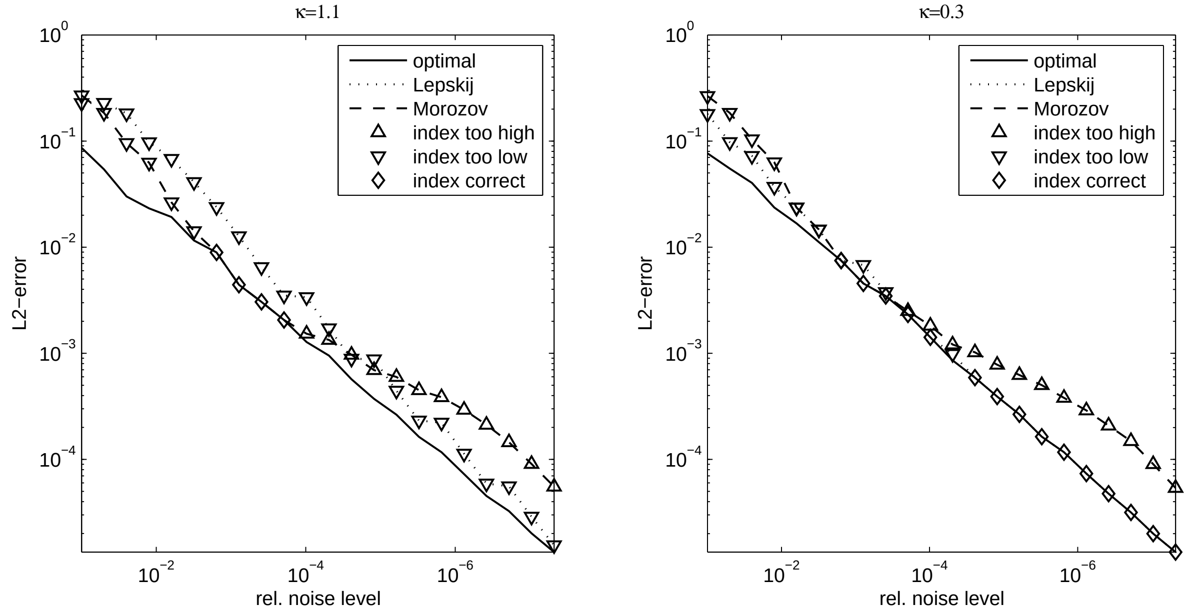

Figure 1. parameter identification problem with random noise: The plots show the L^2-errors of the reconstructed parameter c as a function of the noise level δ.

where χ_Ω is the characteristic function of the domain Ω, which we assume to be star-shaped with respect to the origin. Then $\partial\Omega = \{q(t)(\cos t, \sin t)^\top : t \in [0, 2\pi]\}$ for a positive, 2π-periodic function q. The inverse problem consists in identifying the shape of Ω given the Neumann data $\frac{\partial u}{\partial n}$ of the solution on Γ. Therefore, we define F as the operator mapping q to $\frac{\partial u}{\partial n}$. We have chosen $\Gamma := \{x : |x|_2 = 2\}$, $q^\dagger(t) := (1 + 0.9\cos(t) + 0.1\sin(2t))/(1 + 0.75\cos(t))$ and $q_0(t) := 1$. Random noise was generated as in the previous problem.

It has been shown in [7] that logarithmic source conditions are equivalent to smoothness conditions in terms of Sobolev spaces if $\partial\Omega$ and Γ are concentric circles. Therefore, we expect that the convergence behavior is described by Theorem 4.2 (although it has not been possible to verify assumption (21)).

Inverse obstacle scattering problems. Our last example concerns the scattering of N time-harmonic acoustic waves $u_i^{(j)}(x) = \exp(ikx \cdot d^{(j)})$ with directions $d^{(j)}$ and wave number k by M disjoint, simply connected scattering obstacles $\Omega_1, \dots, \Omega_M \subset \mathbb{R}^2$. On each of the boundaries of the domain Ω_l we impose either Dirichlet or a homogeneous Neumann boundary conditions for the total field $u^{(j)} = u_i^{(j)} + u_s^{(j)}$ and require that the scattered fields $u_s^{(j)}$ satisfy the Helmholtz equation $(\Delta + k^2)u_s^{(j)} = 0$ in $\mathbb{R}^2 \setminus \bigcup_l \Omega_l$ and the Sommerfeld radiation condition (see Colton & Kress [4]). We assume that the number M of domains, an initial guess, and the boundary condition for each of these domains are known. Such information can be obtained by other methods, see e.g. [12]. Then our task is to reconstruct the shape of the scatters given the far field patterns (or scattering amplitudes) $u_\infty^{(j)}$ of the scattered fields $u_s^{(j)}$. This problem is again exponentially ill-

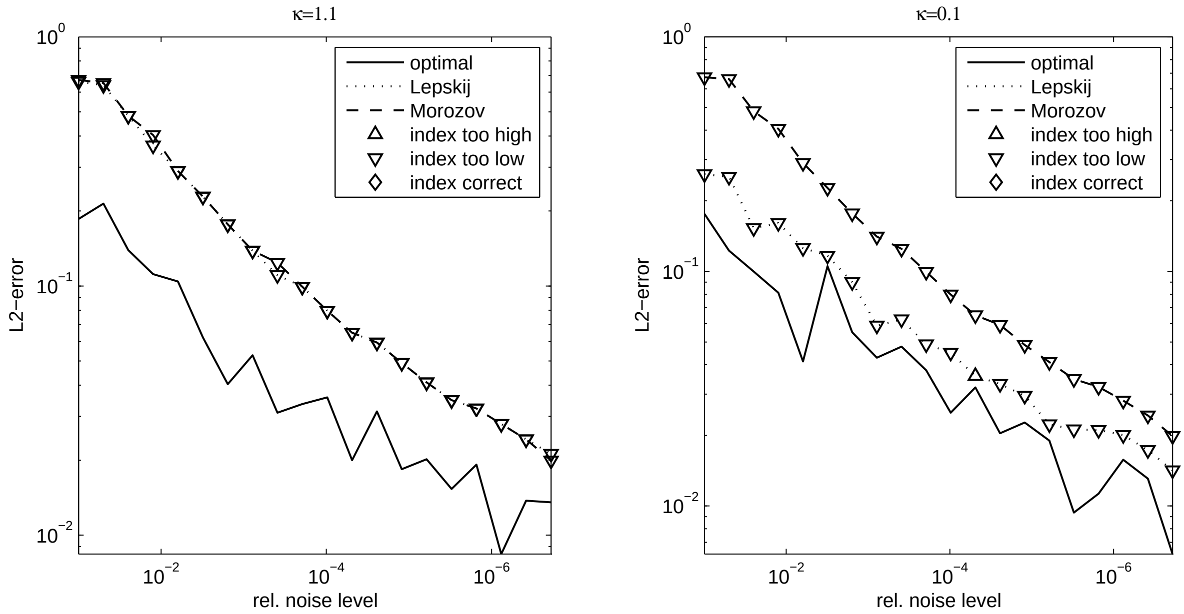

Figure 2. potential problem with random noise: The plots show the L^2-errors of the reconstructed parametrization of the curve as a function of the noise level δ.

posed, and only logarithmic source conditions can be expected to be fulfilled.

In our numerical experiments we have used the scatterers shown in Fig. 5 with $N = 8$ incident waves and wave number $k = 1$. On the upper obstacle we imposed a Dirichlet boundary condition and on the lower obstacle Neumann boundary condition. The initial guess consisted of two circles of radius 1. Random noise was generated by adding a centered Gaussian random variable to the real and imaginary part of the far-field patterns at each measurement point.

In Figure 4 we simulated deterministic noise by adding a fixed positive number to the real and imaginary part of the far-field patterns at each measurement point. The curves for Morozov's discrepancy principle and the Lepskij stopping rule do not differ much from the left plot in Fig. 3 since both stopping rules always terminate too early when the total error is still dominated by the approximation error.

Discussion of the numerical results. We have compared our results with those for the stopping index n_*^{discr} defined by Morozov's discrepancy principle:

$$n_*^{\mathrm{discr}} := \min\{n \in \mathbb{N}_0 : \|F(x_n^\delta) - y^\delta\| \le \tau\delta\}$$

Here $\tau > 1$ is a fixed parameter which is required to be sufficiently large in some convergence proofs. In Fig. 1 we chose $\tau = 2$, which yields too small stopping indices for large δ and too large stopping indices for small δ. For the exponentially ill-posed problems in Fig. 2 and 3 we had to choose $\tau = 4$ since for smaller values of τ the residual never fell below $\tau\delta$ for small δ.

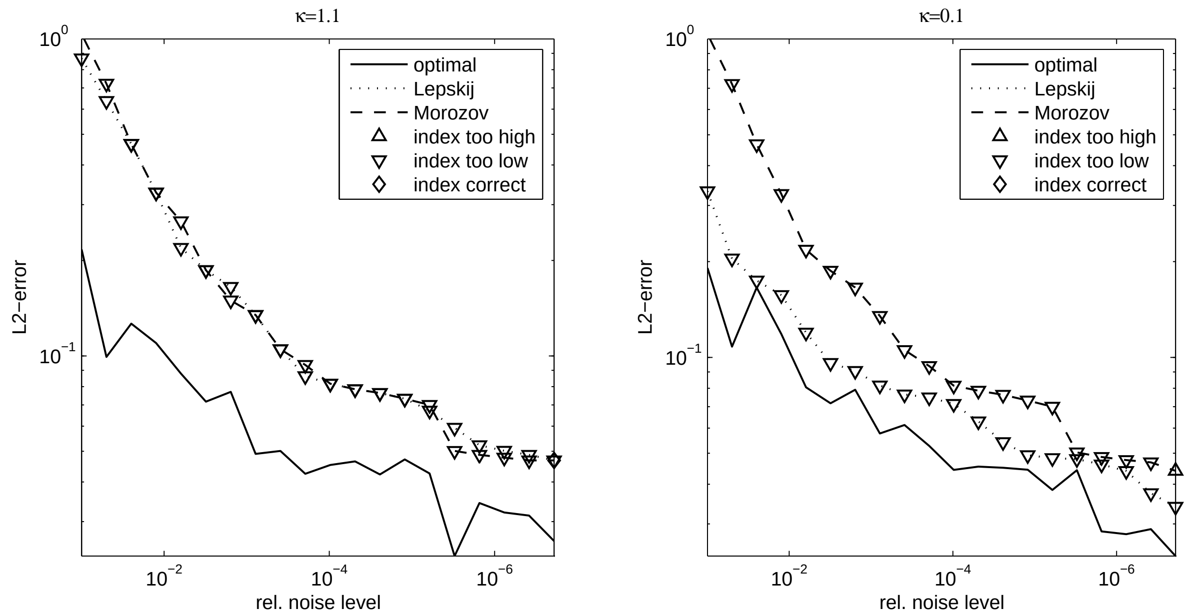

Figure 3. inverse scattering problem with random noise. The plots show the L^2-errors of the reconstructed parametrizations of the curves as a function of the noise level δ.

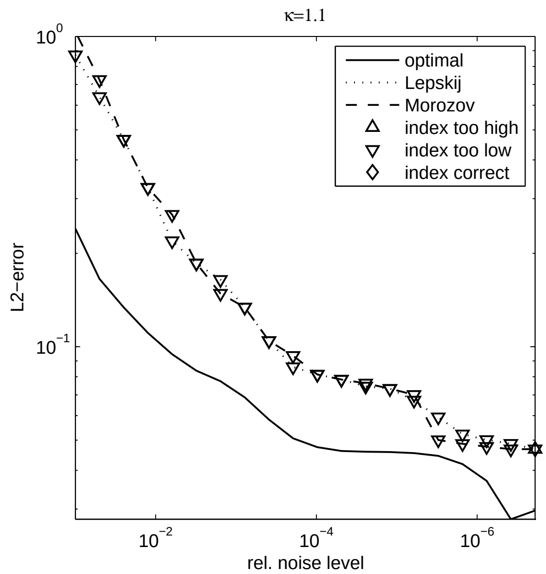

Figure 4. inverse scattering problem with deterministic noise.

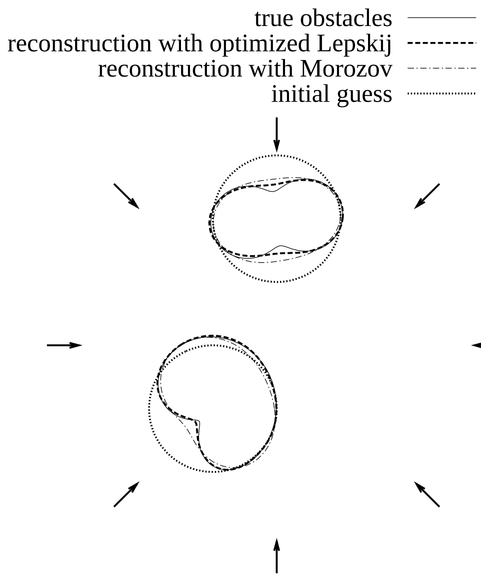

Figure 5. inverse scattering problem: reconstructions for 1% noise.

Recall that the Lepskij-stopping index in Theorems 3.2 and 4.2 was defined by

$$n_* := \min \left\{ n \in \{0, \dots, N_{\max}(\delta)\} : \begin{array}{c} \|x_n^\delta - x_m^\delta\| \le \frac{2\kappa C_g \delta}{\sqrt{\alpha_m}} \\ \text{for all } m = n+1, \dots, N_{\max}(\delta) \end{array} \right\}. \quad (33)$$

Here $N_{\max}(\delta) := \max(0, c + s\lfloor \ln_q \frac{\delta}{\alpha_0} \rfloor)$ with some problem dependent constant $c \in \mathbb{R}$ and $s = 1$ for the weak non-linearity conditions, $s = 2$ for the strong non-linearity conditions. A certain disadvantage of the proposed method is that the constant c used in the proofs can hardly be estimated analytically or computed numerically. Fortunately, the method is not sensitive to the choice of c. c just has to be sufficiently small to prevent nonlinearity blow-off for $n \leq N_{\max}$, but it does not influence the asymptotic rate.

The constants $\kappa = 4$ and $\kappa = 7$ in Theorems 3.2 and 4.2 were chosen for the convenience of the proofs. With more technical proofs one could choose smaller values of $\kappa > 1$, but even for linear problems Assumption 2.1 is not satisfied for $\kappa \leq 1$ if (13) holds true with equality: In this case

$$\|g_{\alpha_n}(F'[x_n^\delta]^* F'[x_n^\delta]) F'[x_n^\delta]^*(y^\delta - y)\| \leq \frac{C_g \delta}{\sqrt{\alpha_n}}$$

is a sharp worst case bound on the propagated data noise error.

However, for stochastic noise it turns out that the left-hand side is usually overestimated by an order of magnitude with high probability. Therefore, one usually obtains much better results for smaller values of κ. To illustrate the potential of the proposed stopping rule, we included convergence plots with $\kappa < 1$ chosen by trial and error on the right hand sides of Fig. 1, 2 and 3. We point out again, that a choice $\kappa \leq 1$ cannot be justified in a worst-case setting. An analysis of the Lepskij stopping rule in a statistical setting which also has to address a proper choice of $\mathcal{E}$ (not necessarily of the form $\mathcal{E}(\alpha) = \kappa C_g \delta / \sqrt{\alpha}$) is intended as future research.

Let us summarize our results:

- We proved that the proposed Lepskij-type stopping rule leads to optimal rates of convergence both for Hölder source conditions with $0 < \mu \leq \mu_0$ and for logarithmic source conditions without a-priori knowledge of the smoothness of the solution.
- It does not suffer from the well-known saturation effect of the discrepancy principle. This can be observed in numerical experiments with mildly ill-posed problems and smooth solutions.
- For exponentially ill-posed problems the numerical results with the Lepskij stopping rule are as good as with the discrepancy principle if we choose $\kappa > 1$ in (33) as required by our theoretical results.
- The results can be significantly improved by choosing $\kappa < 1$. A justification would require a statistical framework for which the Lepskij stopping rule was originally developed.

Acknowledgments

We would like to thank Barbara Kaltenbacher, Andreas Neubauer and Otmar Scherzer for a preliminary version of parts of their book [11]. Moreover, we acknowledge the support of the Graduiertenkolleg 1023 "Identification in Mathematical Models".

References

[1] A. Bakushinskii. The problem of the convergence of the iteratively regularized Gauss-Newton method. *Comput. Maths. Math. Phys.*, 32:1353–1359, 1992.

[2] F. Bauer and S. Pereverzev. Regularization without preliminary knowledge of smoothness and error behavior. *European Journal of Applied Mathematics*, 2005. to appear.

[3] B. Blaschke, A. Neubauer, and O. Scherzer. On convergence rates for the iteratively regularized Gauß-Newton method. *IMA Journal of Numerical Analysis*, 17:421–436, 1997.

[4] D. Colton and R. Kreß. *Inverse Acoustic and Electromagnetic Scattering Theory.* Springer Verlag, Berlin, Heidelberg, New York, second edition, 1997.

[5] H. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems.* Kluwer Academic Publisher, Dordrecht, Boston, London, 1996.

[6] A. Goldenshluger and S. Pereverzev. Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Prob. Theor. Rel. Fields*, 118:169–186, 2000.

[7] T. Hohage. Logarithmic convergence rates of the iteratively regularized Gauß-Newton method for an inverse potential and an inverse scattering problem. *Inverse Problems*, 13:1279–1299, 1997.

[8] T. Hohage. *Iterative Methods in Inverse Obstacle Scattering: Regularization Theory of Linear and Nonlinear Exponentially Ill-Posed Problems.* PhD thesis, University of Linz, 1999.

[9] B. Kaltenbacher. Some Newton-type methods for the regularization of nonlinear ill-posed problems. *Inverse Problems*, 13:729–753, 1997.

[10] B. Kaltenbacher. A posteriori parameter choice strategies for some Newton type methods for the regularization of nonlinear ill-posed problems. *Numer. Math*, 79:501–528, 1998.

[11] B. Kaltenbacher, A. Neubauer, and O. Scherzer. *Iterative Regularization Methods for Nonlinear Ill-Posed Problems.* Springer, Dordrecht, 2005. to appear.

[12] A. Kirsch. Characterization of the shape of the scattering obstacle by the spectral data of the far field operator. *Inverse Problems*, 14:1489–1512, 1998.

[13] O. V. Lepskij. On a problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.*, 35:454–466, 1990.

[14] P. Mathé and S. Pereverzev. Geometry of ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19:789–803, 2003.

Institut für Numerische und Angewandte Mathematik
Universität Göttingen
Lotzestr. 16-18
D - 37083 Göttingen

Telefon: 0551/394512
Telefax: 0551/393944

Email: trapp@math.uni-goettingen.de URL: http://www.num.math.uni-goettingen.de

Verzeichnis der erschienenen Preprints:

2005-01	A. Schöbel, S. Scholl:	Line Planning with Minimal Traveling Time
2005-02	A. Schöbel	Integer programming approaches for solving the delay management problem
2005-03	A. Schöbel	Set covering problems with consecutive ones property
2005-04	S. Mecke, A. Schöbel, D. Wagner	Station location - Complex issues
2005-05	R. Schaback	Convergence Analysis of Methods for Solving General Equations
2005-06	M. Bozzini, L. Lenarduzzi, R. Schaback	Kernel B–Splines and Interpolation
2005-07	R. Schaback	Convergence of Unsymmetric Kernel–Based Meshless Collocation Methods
2005-08	T. Hohage, F.–J. Sayas	Numerical solution of a heat diffusion problem by boundary element methods using the Laplace transform
2005-09	T. Hohage	An iterative method for inverse medium scattering problems based on factorization of the far field operator
2005-10	F. Bauer, T. Hohage	A Lepskij-type stopping rule for regularized Newton methods