On convergence rates
of inexact Newton regularizations

Andreas Rieder⋆

Institut für Wissenschaftliches Rechnen und Mathematische Modellbildung (IWRMM),
Universität Karlsruhe, 76128 Karlsruhe, Germany;
e-mail: andreas.rieder@math.uni-karlsruhe.de

Summary. REGINN is an algorithm of inexact Newton type for the regularization of nonlinear ill-posed problems [*Inverse Problems* 15 (1999), pp. 309-327]. In the present article convergence is shown under weak smoothness assumptions (source conditions). Moreover, convergence rates are established. Some computational illustrations support the theoretical results.

Mathematics Subject Classification (1991): 65J15, 65J20

1. Introduction

In [15] we proposed and analyzed an iteration of inexact Newton type (called REGINN) for the regularization of nonlinear ill-posed problems

$$(1.1) \qquad\qquad F(x) = y^\delta$$

where $F : \mathsf{D}(F) \subset X \to Y$ acts between the Hilbert spaces X and Y. Here, $\mathsf{D}(F)$ denotes the domain of definition of F and y^δ is a perturbation of the exact but unknown data $y = F(x^\dagger)$ satisfying

$$(1.2) \qquad\qquad \|y - y^\delta\|_Y \leq \delta.$$

The non-negative *noise level* δ is assumed to be known.

We have been able to verify (under reasonable assumptions) that REGINN terminates with an approximate solution $x_{N(\delta)}$ of (1.1). Moreover, there is a $\lambda \in [0, 1[$ such that the *regularization property* (1.3) holds true

---

(1.3) $\qquad \|x^\dagger - x_{N(\delta)}\|_X = \mathcal{O}(\delta^{(1-\lambda)/2})$ as $\delta \to 0$

whenever the initial guess x_0 of REGINN satisfies the *source condition*

$$x^\dagger - x_0 \in \mathsf{R}\left((F'(x^\dagger)^* F'(x^\dagger))^{1/2}\right)$$

which is an abstract smoothness assumption. Above, $F' : \mathsf{D}(F) \to \mathcal{L}(X, Y)$ is the Fréchet derivative of F which we assume to exist as a continuous mapping. By $\mathsf{R}(B)$ we denote the range of the linear operator B.

In this paper we investigate the regularization power of REGINN under weaker smoothness requirements. We will prove the existence of a positive $\kappa_{\min} < 1$ such that the source condition

(1.4) $\quad x^\dagger - x_0 \in \mathsf{R}\left((F'(x^\dagger)^* F'(x^\dagger))^{\kappa/2}\right)$ for a $\kappa \in \,]\kappa_{\min}, 1]$

implies the convergence

(1.5) $\qquad \|x^\dagger - x_{N(\delta)}\|_X = \mathcal{O}(\delta^{(\kappa-\kappa_{\min})/(1+\kappa)})$ as $\delta \to 0$.

The 'smallness' of $\kappa_{\min}$ depends on the degree of nonlinearity of F and the inner regularization scheme of REGINN used to regularize the linearized problems. The closer F is to a linear mapping the smaller $\kappa_{\min}$ becomes.

This paper is structured as follows. In the next section we formulate REGINN and recall those of its properties from [15] which we will need. In Sect. 3 we show that REGINN is well defined under (1.4) and terminates with an approximation to $x^\dagger$. Then the regularization property (1.5) will be verified (Sect. 4). Finally, we present numerical experiments for a parameter identification model problem. Here we observe an intrinsic difference between the infinite dimensional problem and its discretization.

2. Formulation of REGINN and known results

Basically, REGINN is a Newton iteration applied to (1.1). The current approximation x_n to $x^\dagger$ is updated by adding a correction step s_n: $x_{n+1} = x_n + s_n$. In the ideal case we would add the exact Newton step $s_n^e = x^\dagger - x_n$ which solves

(2.1) $\qquad F'(x_n)\, s_n^e = y - F(x_n) - E(x^\dagger, x_n) =: b_n$

where

$$E(v, w) := F(v) - F(w) - F'(w)\,(v - w)$$

denotes the Taylor remainder term. Equation (2.1) is a linearization of (1.1) about x_n (with exact data y). The right hand side b_n of (2.1) is not known in general. However the perturbation

$$b_n^\varepsilon := y^\delta - F(x_n) \quad \text{with} \quad \|b_n^\varepsilon - b_n\|_Y \leq \delta + \|E(x^\dagger, x_n)\|_Y$$

is available.

Hence, we compute the Newton correction s_n as an approximate solution of

$$(2.2) \qquad F'(x_n)\,s \;=\; y^\delta - F(x_n).$$

In general, the ill-posedness of (1.1) is passed on to (2.2), see, e.g., Engl, Hanke, and Neubauer [5, Proposition 10.1] and Hofmann and Scherzer [11] for some precise statements.

We therefore apply a regularization scheme to (2.2) and obtain the Newton update

$$(2.3) \qquad s_n \;=\; s_{n,r} \;=\; g_r(A_n^* A_n)A_n^* \, (y^\delta - F(x_n))$$

where $A_n := F'(x_n)$ and $g_r : [0, \theta] \to \mathbb{R}$, $\theta = \|A_n\|^2$, is a piecewise continuous function. The parameter $r \geq 0$ is called *regularization parameter*.

We restrict ourselves to *linear* regularization schemes $\{g_r\}_{r \in \mathbb{N}_0}$, $g_0 := 0$, satisfying the assumptions (2.4) below with $p_r(t) := 1 - t\,g_r(t)$. There exist positive constants C_g, C_p, and α such that

$$(2.4) \qquad \begin{aligned} &\sup_{t \in [0,\theta]} |g_r(t)| \;\leq\; C_g\,r^\alpha, \quad \sup_{t \in [0,\theta]} |p_r(t)| \;=\; 1, \quad \text{and} \\ &\sup_{t \in [0,\theta]} |t\,p_r(t)| \;\leq\; C_p\,r^{-\alpha}. \end{aligned}$$

Please note that the above assumptions on $\{g_r\}$ imply

$$(2.5) \qquad \widetilde{C}_g \;:=\; \sup_{r \in \mathbb{N}} \ \sup_{t \in [0,\theta]} t\,|g_r(t)| \leq 2.$$

Example 2.1. Let us look at four examples of regularization schemes satisfying (2.4).

1. The choice $g_r(t) = 1/(t+1/r)$ leads to the *Tikhonov-Phillips* regularization where $g_r(A_n^* A_n) = (A_n^* A_n + r^{-1} I)^{-1}$ and $C_g = C_p = \alpha = 1$. Here, REGINN (Fig. 2.1) is a variation of the *Levenberg-Marquardt* scheme, see, e.g., Hanke [8].

2. The *truncated singular value decomposition* is characterized by $g_r(t) = 1/t$, for $t \geq 1/r$ and $g_r(t) = 0$, otherwise. Hence, $C_g = C_p = \alpha = 1$.

3. If $g_r(t) = \sum_{j=0}^{r-1}(1 - t)^j$ and $\|A_n\| \leq 1$ then we have the *Landweber* regularization which is an iterative regularization technique where $C_g = \alpha = 1$ and $C_p = \exp(-1)$.

4. Other iterative regularization schemes are given by the *ν-methods* ($\nu > 0$) due to Brakhage [2], see also Hanke [7]. For scaled A_n, that is, $\|A_n\| \leq 1$, the function g_r has the representation $g_r(t) = (1 - \widetilde{P}_r^{(\nu)}(t))/t$ where $\widetilde{P}_r^{(\nu)}(t) = P_r^{(2\nu-1/2,-1/2)}(1 - 2t)/P_r^{(2\nu-1/2,-1/2)}(1)$ with $P_r^{(a,b)}$ denoting the Jacobi polynomials. For $\nu \geq 1$ we have $\alpha = 2$. Explicit

values for C_g and C_p are not known.

Our stable Newton-type solver for (1.1) now has the form

$$(2.6) \quad x_{n+1} \; = \; x_n \; + \; g_{i_n}(A_n^* A_n) A_n^* \left(y^\delta - F(x_n) \right), \quad n = 0, 1, 2, \ldots ,$$

with an initial guess $x_0 \in D(F)$.

In each iteration step we determine i_n such that the relative (linear) residual is smaller than a given tolerance $\mu_n \in]0, 1]$:

$$(2.7) \qquad \|A_n s_{n,i_n} - b_n^\varepsilon\|_Y < \mu_n \, \|b_n^\varepsilon\|_Y \le \|A_n s_{n,r} - b_n^\varepsilon\|_Y,$$
$$r = 1, \ldots, i_n - 1.$$

The iteration (2.6) will be stopped by a *discrepancy principle*. We choose an $R > 0$ and accept that iterate x_N as an approximation to $x^\dagger$ that fulfills

$$(2.8) \qquad \|y^\delta - F(x_N)\|_Y \le R\,\delta \; < \; \|y^\delta - F(x_k)\|_Y,$$
$$k = 0, \ldots, N - 1.$$

See Fig. 2.1 for an implementation of (2.6) based on (2.7) and (2.8).

```
REGINN(x, R, {μ_n})
n = 0,  x_0 = x
while ‖F(x_n) − y^δ‖_Y > Rδ do
{    i_n = 0
     repeat
         i_n = i_n + 1
         s_{n,i_n} = g_{i_n}( F'(x_n)* F'(x_n) ) F'(x_n)* ( y^δ − F(x_n) )
     until ‖F'(x_n) s_{n,i_n} + F(x_n) − y^δ‖_Y < μ_n ‖F(x_n) − y^δ‖_Y
     x_{n+1} = x_n + s_{n,i_n}
     n = n + 1
}
x = x_n
```

Fig. 2.1. REGINN: REGularization based on INexact Newton iteration

Mainly we are interested in using iterative regularizations in the `repeat`-loop of REGINN. Therefore we assume that F' is scaled such that

$$(2.9) \qquad\qquad \|F'(v)\| \le 1 \quad \text{for all } v \in D(F).$$

In our analysis of REGINN we will heavily rely on the local property (2.10) for the nonlinear function F. Let $Q : X \times X \to \mathcal{L}(Y)$ be a mapping such that

$$(2.10)\, F'(v) \; = \; Q(v, w)\, F'(w) \quad \text{and} \quad \|I - Q(v, w)\| \le C_Q \, \|v - w\|_X$$

for all $v, w \in B_\rho(x^\dagger)$, the ball about $x^\dagger$ with radius ρ. This is a strong assumption which essentially forces F to be close to a linear mapping. For instance, the Fréchet derivatives of nonlinear operators with property (2.10) have a null space which is invariant in $B_\rho(x^\dagger)$, that is, $\mathsf{N}(F'(v)) = \mathsf{N}(F'(w))$ for all $v, w \in B_\rho(x^\dagger)$. However, surveying the recent literature on iterative regularization techniques, see, e.g., [1,4,8–10,12,13,16,17], one gets the impression that assumptions closely related to (2.10) are somehow necessary to have a unified convergence theory or to carry over optimality results from the linear to the nonlinear situation. A more detailed discussion of (2.10) can be found in the above cited literature.

Let $C_Q\, \rho < 1$. Then, (2.10) gives

$$(2.11) \quad \|F(v) - F(w)\|_Y \geq (1 - C_Q\, \rho)\, \|F'(w)\, (v - w)\|_Y$$

as well as

$$(2.12) \quad \|E(v, w)\|_Y \leq \omega\, \|F(v) - F(w)\|_Y \quad \text{for all} \quad v, w \in B_\rho(x^\dagger)$$

where $\omega := C_Q\, \rho/(1 - C_Q\, \rho)$, see, e.g., [15, Sect. 3]. Note that $\omega < 1$ for $C_Q\, \rho < 1/2$.

Based on (2.12) we are able to estimate the data error $\|b_n^\varepsilon - b_n\|_Y$ in terms of δ, ω, and the nonlinear defect

$$d_n := \|y^\delta - F(x_n)\|_Y = \|b_n^\varepsilon\|_Y.$$

We have, for $x_n \in B_\rho(x^\dagger)$,

$$(2.13) \quad \|b_n^\varepsilon - b_n\|_Y \leq (1 + \omega)\, \delta + \omega\, d_n := \varepsilon = \varepsilon(x_n, \delta).$$

We quote a result from [15] which gives conditions on μ_n to stop the `repeat`-loop.

Lemma 2.2. *Let $\{g_r\}_{r \in \mathbb{N}}$ satisfy (2.4) and let (2.10) hold true with $C_Q\, \rho < 1/2$. Further assume that $x_k \in B_\rho(x^\dagger)$. If $R \geq (1 + \omega)/(1 - \omega)$ then the* `repeat`-*loop of algorithm* REGINN *terminates for any*

$$(2.14) \qquad \mu_k \in \left]\omega + \frac{(1 + \omega)\, \delta}{d_k},\ 1\right].$$

3. Towards a convergence analysis: termination of REGINN

In a first step towards a convergence analysis we shall show termination of REGINN. To this end we will show that the Newton steps s_{k,i_k} decrease geometrically in k, see (3.24) below. The key estimate is (3.1). The first

relation in (2.4) and (2.5) as well as standard arguments, see, e.g., [5, 14], lead to the norm bound

$$(3.1) \qquad \|s_{k,i_k}\|_X \ \leq \ \sqrt{\widetilde{C}_g \, C_g} \ i_k^{\alpha/2} \, d_k.$$

In the following subsections we therefore bound i_k and the nonlinear defect d_k.

3.1. Bounding i_k

The analysis of this section will be rather technical. The two main results are formulated in Lemmata 3.1 and 3.2 below.

For notational convenience we introduce the ratio

$$(3.2) \qquad \tau_k \ := \ \frac{\mu_k \, d_k}{\varepsilon(x_k, \delta)} \ = \ \frac{\mu_k}{(1 + \omega) \, \delta/d_k \ + \ \omega} \, .$$

Under the hypotheses of Lemma 2.2, i_k, see (2.7), is well defined and $\tau_k > 1$.

If $i_k \geq 2$ then

$$(3.3) \qquad (\tau_k - 1) \, \varepsilon(x_k, \delta) \ \leq \ \|p_{i_k - 1}(A_k \, A_k^*) \, A_k s_k^{\mathrm{e}}\|_Y$$

where $s_k^{\mathrm{e}} = x^\dagger - x_k$ and $A_k = F'(x_k)$, see [15, Sect. 4].

To bound the right hand side of (3.3) we will make frequent use of the *interpolation inequality* (3.4). If $T \in \mathcal{L}(X, Y)$ then

$$(3.4) \quad \|(T^*T)^r x\|_X \ \leq \ \|(T^*T)^q x\|_X^{r/q} \, \|x\|_X^{1-r/q} \quad \text{ for } 0 < r \leq q,$$

see, e.g., [5, 14]. By $|T|$ we will denote $(T^*T)^{1/2}$.

Assume the existence of $w \in X$ and $\kappa \in [0, 1]$ such that

$$(3.5) \qquad s_0^{\mathrm{e}} \ = \ x^\dagger - x_0 \ = \ |A|^\kappa w$$

where $A = F'(x^\dagger)$. Since $s_k^{\mathrm{e}} \ = \ s_0^{\mathrm{e}} - \sum_{j=0}^{k-1} s_{j,i_j} \ = \ s_0^{\mathrm{e}} - \sum_{j=0}^{k-1} A_j^* 4 \, g_{i_j}(A_j A_j^*) b_j^\varepsilon$, see (2.3), we obtain

$$(3.6) \qquad s_k^{\mathrm{e}} \ = \ |A|^\kappa w - w_k \quad \text{with} \quad w_k := \sum_{j=0}^{k-1} A_j^* \, g_{i_j}(A_j A_j^*) \, b_j^\varepsilon$$

which yields

$$(3.7) \qquad \begin{aligned} p_{i_k-1}&(A_k \, A_k^*) \, A_k s_k^{\mathrm{e}} \\ &= p_{i_k-1}(A_k \, A_k^*) \, A_k |A|^\kappa w - p_{i_k-1}(A_k \, A_k^*) \, A_k w_k. \end{aligned}$$

For the following chain of inequalities we use (3.5), (2.10), (3.4), (2.4), and the abbreviation $Q_{\infty,k} := Q(x^\dagger, x_k)$:

$$
\begin{aligned}
&\|p_{i_k-1}(A_k\,A_k^*)\,A_k|A|^\kappa w\|_Y \\
&= \|p_{i_k-1}(A_k\,A_k^*)\,A_k|Q_{\infty,k}A_k|^\kappa w\|_Y \\
&\leq \|\,|Q_{\infty,k}A_k|^\kappa A_k^*\,p_{i_k-1}(A_k\,A_k^*)\|\,\|w\|_X \\
(3.8)\quad &\leq \|\,|Q_{\infty,k}A_k|\,A_k^*\,p_{i_k-1}(A_k\,A_k^*)\|^\kappa\,\|A_k^*\,p_{i_k-1}(A_k\,A_k^*)\|^{1-\kappa}\,\|w\|_X \\
&\leq \|Q_{\infty,k}\|^\kappa\,\|A_kA_k^*\,p_{i_k-1}(A_k\,A_k^*)\|^\kappa \\
&\qquad \times\|A_kA_k^*\,p_{i_k-1}(A_k\,A_k^*)\|^{(1-\kappa)/2}\,\|w\|_X \\
&\leq \frac{\widetilde{C}_Q^\kappa\,C_p^{(1+\kappa)/2}\,\|w\|_X}{(i_k-1)^{\alpha\,(1+\kappa)/2}}
\end{aligned}
$$

where $\widetilde{C}_Q$ is an upper bound of Q: $\|Q(v,z)\| \leq \widetilde{C}_Q$ for all $v,z \in B_\rho(x^\dagger)$.

We further have, by (2.10) with $Q_{k,j} = Q(x_k, x_j)$,

$$
\begin{aligned}
&\|p_{i_k-1}(A_k\,A_k^*)\,A_k w_k\|_Y \\
&\leq \sum_{j=0}^{k-1} \|p_{i_k-1}(A_k\,A_k^*)\,A_k\,A_j^*\,g_{i_j}(A_jA_j^*)b_j^\varepsilon\|_Y \\
&= \sum_{j=0}^{k-1} \|p_{i_k-1}(A_k\,A_k^*)\,Q_{k,j}A_j\,A_j^*\,g_{i_j}(A_jA_j^*)b_j^\varepsilon\|_Y \\
&\leq \sum_{j=0}^{k-1} \|p_{i_k-1}(A_k\,A_k^*)\,Q_{k,j}\,|A_j^*|^{1+\kappa}\|\,\|\,|A_j^*|^{1-\kappa}\,g_{i_j}(A_jA_j^*)b_j^\varepsilon\|_Y.
\end{aligned}
$$

Moreover, using (3.4), (2.10), and (2.4),

$$
\begin{aligned}
&\|p_{i_k-1}(A_k\,A_k^*)\,Q_{k,j}\,|A_j^*|^{1+\kappa}\| \\
&= \|\,|A_j^*|^{1+\kappa}\,Q_{k,j}^*\,p_{i_k-1}(A_k\,A_k^*)\| \\
&\leq \widetilde{C}_Q^{(1-\kappa)/2}\,\|A_jA_j^*\,Q_{k,j}^*\,p_{i_k-1}(A_k\,A_k^*)\|^{(1+\kappa)/2} \\
&= \widetilde{C}_Q^{(1-\kappa)/2}\,\|Q_{j,k}\,A_kA_k^*\,p_{i_k-1}(A_k\,A_k^*)\|^{(1+\kappa)/2} \\
&\leq \widetilde{C}_Q\,C_p^{(1+\kappa)/2}\,(i_k-1)^{-\alpha\,(1+\kappa)/2},
\end{aligned}
$$

so that

$$
\begin{aligned}
(3.9) \quad & \| p_{i_k-1}(A_k\,A_k^*)\,A_k w_k \|_Y \\
& \leq \frac{\widetilde{C}_Q\,C_p^{(1+\kappa)/2}}{(i_k-1)^{\alpha\,(1+\kappa)/2}} \sum_{j=0}^{k-1} \|\,|A_j^*|^{1-\kappa}\,g_{i_j}(A_j A_j^*) b_j^\varepsilon \|_Y .
\end{aligned}
$$

Finally, (3.7), (3.8), and (3.9) yield

$$
(3.10) \qquad \| p_{i_k-1}(A_k\,A_k^*)\,A_k s_k^{\mathrm{e}} \|_Y \;\leq\; C_w\,\frac{W(k)}{(i_k-1)^{\alpha\,(1+\kappa)/2}}
$$

where $C_w = C_p^{(1+\kappa)/2}\,\max\{\widetilde{C}_Q^\kappa,\,\widetilde{C}_Q\}$ and

$$
(3.11) \qquad W(k) \;:=\; \|w\|_X \;+\; \sum_{j=0}^{k-1} \|\,|A_j^*|^{1-\kappa}\,g_{i_j}(A_j A_j^*) b_j^\varepsilon \|_Y .
$$

We are now in a position to bound i_k.

Lemma 3.1. *Let $\{g_r\}_{r\in\mathbb{N}_0}$ fulfill (2.4). Suppose (2.10) and let the first n iterates $\{x_1,\dots,x_n\}$ of algorithm* REGINN *be well defined and stay in $B_\rho(x^\dagger)$. Moreover, let the initial guess $x_0 \in B_\rho(x^\dagger)$ be chosen such that (3.5) holds true for a $\kappa \in [0,1]$. Then, there is a constant C_I such that*

$$
(3.12) \qquad i_k \;\leq\; C_I \left(\frac{W(k)}{\tau_k - 1} \right)^{2/(\alpha\,(1+\kappa))} \varepsilon(x_k,\delta)^{-2/(\alpha\,(1+\kappa))}
$$

for $k = 0,\dots,n$ where C_I depends neither on k nor on n.

Proof. First, we consider the case $i_k \geq 2$. From (3.3) and (3.10) we obtain $(i_k - 1)^{\alpha\,(1+\kappa)/2} \leq C_w\,W(k)/(\tau_k - 1)/\varepsilon(x_k,\delta)$. Since $i_k \leq 2\,(i_k - 1)$ the inequality (3.12) is established. In the case of $i_k = 1$ the trivial estimate $\tau_k\,\varepsilon(x_k,\delta) = \mu_k\,\|b_k^\varepsilon\|_Y \leq \|b_k^\varepsilon\|_Y \leq \|A_k s_k^{\mathrm{e}}\|_Y + \varepsilon(x_k,\delta)$ together with (2.9) and similar arguments as above readily imply (3.12). $\square$

Now we wish to know how $W(k)$ behaves as k grows. Under the hypotheses of Lemma 3.1 we will establish the recursive bound

$$
(3.13) \qquad W(k) \;\leq\; \|w\|_X \;+\; C(n) \sum_{j=0}^{k-1} W(j), \quad k = 0,\dots,n,
$$

for a positive constant $C(n)$. Inductively, (3.13) implies

$$
W(k) \;\leq\; \Lambda_n^k\,\|w\|_X, \quad k = 0,\dots,n, \quad \text{with} \quad \Lambda_n = 1 + C(n).
$$

Our verification of (3.13) will rely crucially on the estimate (3.14) below. For x_j, $x_k \in B_\rho(x^\dagger)$ and $C_Q\rho < 1/2$, see (2.10), we have

$$(3.14) \qquad \| \, |A_j|^{-\kappa} \, |A_k|^{\kappa} \| \le \underbrace{(1 - 2C_Q\rho)^{-\kappa}}_{=: \, C_K} \quad \text{for all } \kappa \in [0,1]$$

which was proved by Kaltenbacher in [13, Lemma 2.2].

In view of (3.11) we have to cope with $\| \, |A_j^*|^{1-\kappa} \, g_{i_j}(A_jA_j^*)b_j^\varepsilon\|$. By the triangle inequality and by $A_j s_j^e = b_j$, see (2.1), we find

$$(3.15) \begin{aligned} \| \, |A_j^*|^{1-\kappa} \, g_{i_j}(A_jA_j^*)b_j^\varepsilon\|_Y &\le \| \, |A_j^*|^{1-\kappa} \, g_{i_j}(A_jA_j^*)(b_j^\varepsilon - b_j)\|_Y \\ &+ \| \, |A_j^*|^{1-\kappa} \, g_{i_j}(A_jA_j^*) \, A_j s_j^e\|_Y. \end{aligned}$$

Each of the norms on the above right hand side will be estimated now. We begin with

$$\begin{aligned} \| \, |A_j^*|^{1-\kappa} \, &g_{i_j}(A_jA_j^*) \, (b_j^\varepsilon - b_j)\|_Y \\ &\le \|A_jA_j^* \, g_{i_j}(A_jA_j^*)\|^{(1-\kappa)/2} \, \|g_{i_j}(A_jA_j^*)\|^{(1+\kappa)/2} \, \varepsilon(x_j, \delta) \\ &\le \widetilde{C}_g^{(1-\kappa)/2} \, C_g^{(1+\kappa)/2} \, i_j^{\alpha\,(1+\kappa)/2} \, \varepsilon(x_j, \delta) \end{aligned}$$

where we used (3.4), (2.4), (2.5), and (2.13). Taking (3.12) into account gives

$$(3.16) \begin{aligned} \| \, |A_j^*|^{1-\kappa} \, &g_{i_j}(A_jA_j^*) \, (b_j^\varepsilon - b_j)\|_Y \\ &\le \widetilde{C}_g^{(1-\kappa)/2} \, C_g^{(1+\kappa)/2} \, C_I^{\alpha\,(1+\kappa)/2} \, \frac{W(j)}{\tau_j - 1}. \end{aligned}$$

Next we consider $\| \, |A_j^*|^{1-\kappa} \, g_{i_j}(A_jA_j^*) \, A_j s_j^e\|_Y$ using (3.6), (3.14), and (2.5):

$$(3.17) \begin{aligned} \| \, |A_j^*|^{1-\kappa} \, &g_{i_j}(A_jA_j^*) \, A_j s_j^e\|_Y \\ &\le \| \, |A_j^*|^{1-\kappa} \, g_{i_j}(A_jA_j^*) \, A_j|A|^{\kappa}w\|_Y \\ &\quad + \| \, |A_j^*|^{1-\kappa} \, g_{i_j}(A_jA_j^*) \, A_j w_j\|_Y \\ &\le \| \, |A_j^*|^{1-\kappa} g_{i_j}(A_jA_j^*) \, A_j|A_j|^{\kappa}|A_j|^{-\kappa} \, |A|^{\kappa}\| \, \|w\|_X \\ &\quad + \| \, |A_j^*|^{1-\kappa} \, g_{i_j}(A_jA_j^*) \, A_j w_j\|_Y \\ &\le \widetilde{C}_g \, C_K \, \|w\|_X + \| \, |A_j^*|^{1-\kappa} \, g_{i_j}(A_jA_j^*) \, A_j w_j\|_Y. \end{aligned}$$

So we are left with the investigation of

$$\begin{aligned} \| \, |A_j^*|^{1-\kappa} \, &g_{i_j}(A_jA_j^*) \, A_j w_j\|_Y \\ &= \| \, |A_j^*|^{1-\kappa} \, g_{i_j}(A_jA_j^*) \, A_j \sum_{r=0}^{j-1} A_r^* \, g_{i_r}(A_rA_r^*) \, b_r^\varepsilon\|_Y \end{aligned}$$

$$\leq \sum_{r=0}^{j-1} \| \, |A_j^*|^{1-\kappa} \, g_{i_j}(A_j A_j^*) \, A_j \, A_r^* \, g_{i_r}(A_r A_r^*) \, b_r^\varepsilon \|_Y$$

(3.18)
$$= \sum_{r=0}^{j-1} \| \, |A_j^*|^{1-\kappa} \, g_{i_j}(A_j A_j^*) \, A_j \, |A_r|^\kappa$$
$$\times |A_r|^{-\kappa} \, A_r^* \, g_{i_r}(A_r A_r^*) \, b_r^\varepsilon \|_Y$$

$$\leq \sum_{r=0}^{j-1} \| g_{i_j}(A_j A_j^*) \, |A_j^*|^{1-\kappa} \, A_j \, |A_r|^\kappa \|$$
$$\times \| \, |A_r^*|^{1-\kappa} \, g_{i_r}(A_r A_r^*) \, b_r^\varepsilon \|_Y$$

$$\overset{(\star)}{\leq} \sum_{r=0}^{j-1} \| g_{i_j}(A_j A_j^*) \, |A_j^*|^2 \| \, \| \, |A_j|^{-\kappa} \, |A_r|^\kappa \|$$
$$\times \| \, |A_r^*|^{1-\kappa} g_{i_r}(A_r A_r^*) \, b_r^\varepsilon \|_Y$$

$$\leq \widetilde{C}_g \, C_K \sum_{r=0}^{j-1} \| \, |A_r^*|^{1-\kappa} \, g_{i_r}(A_r A_r^*) \, b_r^\varepsilon \|_Y.$$

The inequality marked with $(\star)$ may be proved using the spectral representation of A_j.

Now we collect the pieces. By (3.17), (3.19), and (3.11),

$$\| \, |A_j^*|^{1-\kappa} \, g_{i_j}(A_j A_j^*) \, A_j s_j^{\mathrm{e}} \|_Y \leq \widetilde{C}_g \, C_K \, W(j)$$

which, together with (3.15) and (3.16), yields

$$\| \, |A_j^*|^{1-\kappa} \, g_{i_j}(A_j A_j^*) b_j^\varepsilon \|_Y \leq C_{W,j} \, W(j)$$

(3.19)
$$\text{with} \quad C_{W,j} := \widetilde{C}_g \, C_K + \frac{\widetilde{C}_g^{(1-\kappa)/2} \, C_g^{(1+\kappa)/2} \, C_I^{\alpha\,(1+\kappa)/2}}{\tau_j - 1}.$$

We formulate our findings with all the technical hypotheses in the following lemma. Its proof is an immediate consequence of (3.19) and (3.11).

Lemma 3.2. *Let $\{g_r\}_{r\in\mathbb{N}_0}$ fulfill (2.4). Suppose (2.10) and let the first n iterates $\{x_1, \dots, x_n\}$ of algorithm* REGINN *be well defined and stay in $B_\rho(x^\dagger)$. Moreover, let the initial guess $x_0 \in B_\rho(x^\dagger)$ be chosen such that (3.5) holds true for a $\kappa \in [0, 1]$.*

If $C_Q\,\rho < 1/2$ then

$$W(k) \;\leq\; \Lambda_n^k\,\|w\|_X, \quad k = 0,\ldots,n,$$

(3.20)
$$\text{with } \Lambda_n \;=\; 1 + \widetilde{C}_g\,C_K + \frac{\widetilde{C}_g^{(1-\kappa)/2}\,C_g^{(1+\kappa)/2}\,C_I^{\alpha\,(1+\kappa)/2}}{t_n - 1}$$

where $t_n = \min\{\tau_0,\ldots,\tau_n\} > 1$.

3.2. Termination of REGINN

Under reasonable technical assumptions all Newton iterates stay in $B_\rho(x^\dagger)$ and REGINN terminates with an approximation $x_{N(\delta)}$ to $x^\dagger$.

Theorem 3.3. *Let $\{g_r\}_{r\in\mathbb{N}}$ satisfy (2.4) and let (2.10) hold true with $C_Q\,\rho < 1/2$. Let $\tau > 1$ and set*

$$\Lambda \;=\; 1 + \widetilde{C}_g\,C_K + \frac{\widetilde{C}_g^{(1-\kappa)/2}\,C_g^{(1+\kappa)/2}\,C_I^{\alpha\,(1+\kappa)/2}}{\tau - 1}\,.$$

Suppose that (2.12) is satisfied with

$$\omega \;<\; \frac{\eta}{\eta + (1+\tau)} \quad \text{where} \quad \eta\cdot\Lambda < 1$$

(this will be true, for instance, if ρ is sufficiently small).

Assume that the starting guess $x_0 \in B_{\rho/2}(x^\dagger)$ is chosen such that the source condition (3.5) applies for $\kappa \in\,]\log_{1/\eta}\Lambda, 1]$ and that the product $\|w\|_X\,\|y^\delta - F(x_0)\|_Y$ is sufficiently small. If $\delta > 0$ and

$$R \;\geq\; \frac{\tau\,(1+\omega)}{\eta - \omega\,(\eta + (1+\tau))} \quad \text{and}$$

$$\mu_k \;\in\; \left[\,\tau\left(\omega + \frac{(1+\omega)\,\delta}{d_k}\right),\; \eta - (1+\eta)\,\omega\,\right]$$

for $k \geq 0$ then there is an $N(\delta) \in \mathbb{N}$ such that all Newton iterates $\{x_1,\ldots,x_{N(\delta)}\}$ are well defined and stay in $B_\rho(x^\dagger)$. Moreover, the final iterate $x_{N(\delta)}$ satisfies the discrepancy principle (2.8) and, for $d_0 > R\,\delta$,

(3.21)
$$N(\delta) \;\leq\; \lfloor\log_\eta(R\,\delta/d_0)\rfloor + 1.$$

Here, $\lfloor t\rfloor \in \mathbb{Z}$ for $t \in \mathbb{R}$ denotes the greatest integer: $\lfloor t\rfloor \leq t < \lfloor t\rfloor + 1$.

Proof. We will prove Theorem 3.3 by induction. Therefore, assume that, for $n \in \mathbb{N}_0$, the iterates $\{x_0, \ldots, x_n\}$ are well defined under the hypotheses of Theorem 3.3 and stay in $B_\rho(x^\dagger)$. We then have that

$$(3.22) \qquad d_k = \|y^\delta - F(x_k)\|_Y \leq \eta^k \|y^\delta - F(x_0)\|_Y = \eta^k d_0$$
$$\text{for} \quad k = 0, \ldots, n.$$

This follows from $F(x_{j+1}) - y^\delta = A_j s_{j,i_j} + F(x_j) - y^\delta + E(x_{j+1}, x_j)$, $j = 0, \ldots, n-1$, which yields

$$\|F(x_{j+1}) - y^\delta\|_Y$$
$$\leq \mu_j \|F(x_j) - y^\delta\|_Y + \omega \|F(x_{j+1}) - F(x_j)\|_Y$$
$$\leq \mu_j \|F(x_j) - y^\delta\|_Y + \omega (\|F(x_{j+1}) - y^\delta\|_Y + \|F(x_j) - y^\delta\|_Y).$$

Hence,

$$(3.23) \qquad \frac{d_{j+1}}{d_j} \leq \frac{\mu_j + \omega}{1 - \omega} \leq \eta \quad \text{for } j = 0, \ldots, n-1$$

which implies (3.22) inductively.

If $d_n \leq R\delta$ the iteration will be stopped by (2.8) with $N(\delta) = n$. Otherwise, $d_n > R\delta$ and we show that the interval determining μ_n is not empty. The bound on ω implies that the denominator of the lower bound of R is positive. The lower bound on R guarantees that $\tau(\omega + (1+\omega)\delta/d_n) < \tau(\omega + (1+\omega)/R) < \eta - (1+\eta)\omega$.

According to Lemma 2.2, i_n and thus the Newton step s_{n,i_n} are well defined. By (3.1), (3.12), and (3.20),

$$\|s_{n,i_n}\|_X \leq \sqrt{\widetilde{C}_g C_g} \; i_n^{\alpha/2} d_n$$

$$\leq \sqrt{\widetilde{C}_g C_g C_I^\alpha} \left(\frac{W(n)}{\tau_n - 1} \right)^{1/(1+\kappa)} \varepsilon(x_n, \delta)^{-1/(1+\kappa)} d_n$$

$$\leq \sqrt{\widetilde{C}_g C_g C_I^\alpha} \left(\frac{\|w\|_X}{\tau_n - 1} \right)^{1/(1+\kappa)} \Lambda_n^{n/(1+\kappa)} \varepsilon(x_n, \delta)^{-1/(1+\kappa)} d_n.$$

The lower bound on the μ_k's yields $\tau_k \geq \tau > 1, k = 0, \ldots, n$, cf. (3.2), that is, $\Lambda_n \leq \Lambda$. Moreover, $d_n/\varepsilon(x_n, \delta) \leq 1/(\tau\omega)$. Taking (3.22) into account we obtain

$$(3.24) \qquad \|s_{n,i_n}\|_X \leq C_S \|w\|_X^{1/(1+\kappa)} d_0^{\kappa/(1+\kappa)} \sigma(\kappa)^n$$

where $C_S = \sqrt{\widetilde{C}_g C_g C_I^\alpha}/((\tau - 1)\tau\omega)^{1/(1+\kappa)}$ and

$$(3.25) \qquad \sigma(\kappa) := (\Lambda \cdot \eta^\kappa)^{1/(1+\kappa)} < 1$$

(note that $\sigma(\kappa)$ is smaller than 1 since $\kappa > \log_{1/\eta} \Lambda$). Now, let $\|w\|_X \, d_0$ be small enough such that

(3.26)
$$a(\delta) := C_S \, \|w\|_X^{1/(1+\kappa)} \, \|F(x_0) - y^\delta\|_X^{\kappa/(1+\kappa)}$$
$$\Big/ (1 - \sigma(\kappa)) \leq \rho/2.$$

Then, the new iterate $x_{n+1} = x_n + s_{n,i_n} = x_0 + \sum_{k=0}^n s_{k,i_k}$ is in $B_\rho(x^\dagger)$:

$$\|x^\dagger - x_{n+1}\|_X \leq \|x^\dagger - x_0\|_X + \sum_{k=0}^n \|s_{k,i_k}\|_X \leq \rho/2 + a(\delta) \leq \rho.$$

Further, $d_{n+1} \leq \eta^{n+1} \, d_0$. This completes the inductive step, thereby finishing the proof of Theorem 3.3.   □

Our next result shows that the reduction rate d_{k+1}/d_k for the nonlinear residuals approximates the tolerance μ_k as the iteration progresses.

Corollary 3.4. *Adopt the assumptions of Theorem 3.3. Then, for $k = 0, \ldots,$ $N(\delta) - 1$,*

$$\frac{\|y^\delta - F(x_{k+1})\|_Y}{\|y^\delta - F(x_k)\|_Y} \leq \min\left\{ \frac{\mu_k + \omega}{1 - \omega}, \; \mu_k + C_D \, \sigma(\kappa)^k \right\}$$

where $C_D = C_Q \, C_S \, \|w\|_X^{1/(1+\kappa)} \, \|y^\delta - F(x_0)\|_Y^{\kappa/(1+\kappa)}$ and $\sigma(\kappa)$ is from (3.25).

Proof. In view of (3.23) it suffices to verify that $d_{k+1}/d_k \leq \mu_k + C_D \, \sigma(\kappa)^k$. The estimate $d_{k+1} < (C_Q \, \|s_{k,i_k}\|_X + \mu_k) \, d_k$ was shown in the proof of Corollary 4.7 in [15]. Now, the assertion follows from (3.24).   □

4. Convergence analysis

In this section we will verify the regularization property of algorithm REGINN, that is, the convergence of $x_{N(\delta)}$ to $x^\dagger$ under the hypotheses of Theorem 3.3. To this end we study

(4.1)
$$\|x^\dagger - x_k\|_X^2 = \langle s_k^e, s_k^e \rangle_X$$
$$\overset{(3.6)}{=} \langle s_k^e, |A|^\kappa w \rangle_X - \sum_{j=0}^{k-1} \langle s_k^e, A_j^* \, g_{i_j}(A_j A_j^*) b_j^\varepsilon \rangle_X.$$

Ideas used before lead us to

$$|\langle s_k^{\mathrm{e}}, |A|^\kappa w\rangle_X| = |\langle s_k^{\mathrm{e}}, |Q_{\infty,k}A_k|^\kappa w\rangle_X| \le \||Q_{\infty,k}A_k|^\kappa s_k^{\mathrm{e}}\|_X \|w\|_X$$
$$\le \||Q_{\infty,k}A_k|s_k^{\mathrm{e}}\|_X^\kappa \|s_k^{\mathrm{e}}\|_X^{1-\kappa} \|w\|_X$$
$$\le \widetilde{C}_Q^\kappa \|A_k s_k^{\mathrm{e}}\|_Y^\kappa \|s_k^{\mathrm{e}}\|_X^{1-\kappa} \|w\|_X.$$

Note that $A_j^* \, g_{i_j}(A_j A_j^*)b_j^\varepsilon \in \mathrm{D}(|A_j|^{-\kappa})$. Hence,

$$|\langle s_k^{\mathrm{e}}, A_j^* \, g_{i_j}(A_j A_j^*)b_j^\varepsilon\rangle_X|$$
$$= |\langle |A_j|^\kappa s_k^{\mathrm{e}}, |A_j|^{-\kappa} A_j^* \, g_{i_j}(A_j A_j^*)b_j^\varepsilon\rangle_X|$$
$$\le \||A_j|^\kappa s_k^{\mathrm{e}}\|_X \||A_j|^{-\kappa} A_j^* \, g_{i_j}(A_j A_j^*)b_j^\varepsilon\|_X$$
$$\le \||Q_{j,k}A_k|s_k^{\mathrm{e}}\|_X^\kappa \|s_k^{\mathrm{e}}\|_X^{1-\kappa} \||A_j^*|^{1-\kappa} \, g_{i_j}(A_j A_j^*)b_j^\varepsilon\|_Y$$
$$\le \widetilde{C}_Q^\kappa \|A_k s_k^{\mathrm{e}}\|_Y^\kappa \|s_k^{\mathrm{e}}\|_X^{1-\kappa} \||A_j^*|^{1-\kappa} \, g_{i_j}(A_j A_j^*)b_j^\varepsilon\|_Y.$$

Recalling (3.11) the latter two displayed inequalities together with (4.1) result in

$$\|s_k^{\mathrm{e}}\|_X^2 \le \widetilde{C}_Q^\kappa \, W(k) \, \|A_k s_k^{\mathrm{e}}\|_Y^\kappa \, \|s_k^{\mathrm{e}}\|_X^{1-\kappa}.$$

Thus, we end up with

$$(4.2) \qquad \|s_k^{\mathrm{e}}\|_X \le \widetilde{C}_Q^{\kappa/(1+\kappa)} \, W(k)^{1/(1+\kappa)} \, \|A_k s_k^{\mathrm{e}}\|_Y^{\kappa/(1+\kappa)}.$$

We are now well prepared for our convergence result.

Theorem 4.1. *Adopt the assumptions of Theorem 3.3, especially let the source condition (3.5) hold for $\kappa \in \,]\log_{1/\eta} \Lambda, 1]$. Further, suppose that $a(0) < \rho/2$, cf. (3.26).*

If $d_0 = \|y^\delta - F(x_0)\|_Y > R\,\delta > 0$ (for instance, $F(x_0) \ne y$ and δ sufficiently small) then

$$(4.3) \qquad \|x^\dagger - x_{N(\delta)}\|_X = \mathcal{O}(\delta^{(\kappa - \log_{1/\eta} \Lambda)/(1+\kappa)}) \quad \text{as } \delta \to 0.$$

In the noise free situation, $\delta = 0$, we have that

$$(4.4) \qquad \|x^\dagger - x_k\|_X = \mathcal{O}(\sigma(\kappa)^k) \quad \text{as } k \to \infty$$

with $\sigma(\kappa) < 1$ from (3.25).

Proof. Note that the elements of the Newton sequence produced by REGINN depend on δ, that is, $x_k = x_k^\delta$, $k = 1, \ldots, N(\delta)$. Since $a(0) < \rho/2$ there exists a $\overline{\delta} > 0$ such that $a(\delta) \leq \rho/2$ for all $0 < \delta \leq \overline{\delta}$, i.e., $\{\, x_k^\delta \,|\, 0 < \delta \leq \overline{\delta},\ k = 1, \ldots, N(\delta)\,\} \subset B_\rho(x^\dagger)$. We infer from (4.2), (2.11), (1.2), and (2.8) that

$$
\begin{aligned}
&\|x^\dagger - x_{N(\delta)}\|_X \\
&\leq \frac{\widetilde{C}_Q^{\kappa/(1+\kappa)}}{(1 - C_Q\,\rho)^{\kappa/(1+\kappa)}}\ W(N(\delta))^{1/(1+\kappa)}\ \|y - F(x_{N(\delta)})\|_X^{\kappa/(1+\kappa)} \\
&\leq \frac{\widetilde{C}_Q^{\kappa/(1+\kappa)}\,(R+1)^{\kappa/(1+\kappa)}}{(1 - C_Q\,\rho)^{\kappa/(1+\kappa)}}\ W(N(\delta))^{1/(1+\kappa)}\ \delta^{\kappa/(1+\kappa)} \\
&\leq \frac{\widetilde{C}_Q^{\kappa/(1+\kappa)}\,(R+1)^{\kappa/(1+\kappa)}}{(1 - C_Q\,\rho)^{\kappa/(1+\kappa)}}\ \|w\|_X^{1/(1+\kappa)}\ \Lambda^{N(\delta)/(1+\kappa)}\ \delta^{\kappa/(1+\kappa)}.
\end{aligned}
$$

For the last inequality we used (3.20) with $k = n = N(\delta)$ and $\Lambda_{N(\delta)} \leq \Lambda$.

Since $N(\delta) \leq \log_\eta(R\,\delta/d_0) + 1$, see (3.21), we obtain that $\Lambda^{N(\delta)} \leq \Lambda\,\Lambda^{\log_\eta(R\,\delta/d_0)} = \Lambda\,(R\,\delta/d_0)^{\log_\eta \Lambda}$. Further, $\log_\eta \Lambda = -\log_{1/\eta} \Lambda$ which verifies (4.3).

In the noise free situation under the assumptions of Theorem 3.3 the Newton sequence $\{x_k\}_{k \in \mathbb{N}_0}$ is well defined and infinite. The convergence result (4.4) follows immediately from (4.2) and (3.22). $\qquad\square$

5. Computational illustrations

Some numerical experiments shall illustrate the mode of action of REGINN. We will realize an essential difference between the infinite dimensional setting and the finite dimensional computations.

We like to reconstruct c in the 2D-elliptic problem

(5.1)
$$
\begin{aligned}
-\Delta u + c\,u &= f \quad \text{in } \Omega \\
u &= g \quad \text{on } \partial\Omega
\end{aligned}
$$

from the knowledge of u in $\Omega = \,]0,1[^2$ where Δ is the Laplacian. Further, $f \in L^2(\Omega)$ and g is the trace of a function in $H^2(\Omega)$. Let $F : \mathrm{D}(F) \to L^2(\Omega)$ be the operator mapping the parameter c to the solution u of (5.1). Here, $\mathrm{D}(F) = \{c \in L^2(\Omega)\,|\,\|c - \widetilde{c}\|_{L^2} \leq \beta \text{ for some } \widetilde{c} \geq 0\}$ for a positive β small enough, see Colonius and Kunisch [3, Lemma 2.1].

Identifying c thus reduces to solve the nonlinear problem

(5.2)
$$
F(c) = u.
$$

If u has no zeroes in Ω then (5.2) admits a unique solution $c^\dagger$ which does not depend continuously on the data. Hanke, Neubauer, and Scherzer have been able to verify (2.10) in the vicinity of any $c \in \mathrm{D}(F)$ such that $F(c) > 0$ a.e., see [10, Example 4.2]. The abstract smoothness condition (1.4) for $\kappa = 1$ may be formulated as

(5.3) $$(c^\dagger - c_0)/F(c_0) \in H^2(\Omega) \cap H_0^1(\Omega),$$

especially, $(c^\dagger - c_0)|_{\partial\Omega} = 0$, see [15].

For our computations we discretize (5.1) by finite differences w.r.t. the grid points $(x_i, y_j) = (ih, jh) \in \Omega$, $0 \le i, j \le n + 1$, where $n \in \mathbb{N}$ and $h = 1/(n + 1)$ is the discretization step size. A lexicographical ordering of the grid points yields the $n^2 \times n^2$-linear system

$$(\mathbf{A} + \mathrm{diag}(\mathbf{c}))\,\mathbf{u} = \mathbf{f}$$

where $\mathbf{A}$ approximates $-\Delta$ and $\mathrm{diag}(\mathbf{c}) = \mathrm{diag}(\mathbf{c}_1, \dots, \mathbf{c}_{n^2})$ is the diagonal matrix with entries $\mathbf{c}_{\ell(i,j)} = c(x_i, y_j)$. By $\ell : \{1, \dots, n\}^2 \to \{1, \dots, n^2\}$ we denote the lexicographical ordering. The details of finite differences can be found, e.g., in Hackbusch [6].

In the discrete situation we wish to recover $\mathbf{c}$ from $\mathbf{u}$. The corresponding nonlinear equation is

(5.4) $$\mathbf{F}(\mathbf{c}) = \mathbf{u}$$

with $\mathbf{F} : \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ defined by $\mathbf{F}(\mathbf{c}) = (\mathbf{A} + \mathrm{diag}(\mathbf{c}))^{-1}\,\mathbf{f}$. The function $\mathbf{F}$ is differentiable with Jacobian

$$\mathbf{F}'(\mathbf{c}) = -(\mathbf{A} + \mathrm{diag}(\mathbf{c}))^{-1}\mathrm{diag}(\mathbf{F}(\mathbf{c})).$$

In our numerical experiments below we identify the parameter $c^\dagger(x, y) = 1.5\sin(4\pi x) \cdot \sin(6\pi y) + 3\left((x - 0.5)^2 + (y - 0.5)^2\right) + 2$. Further, f and g are such that $u(x, y) = 16\,x\,(x - 1)\,y\,(1 - y) + 1$ is the solution of (5.1) w.r.t. $c^\dagger$.

The perturbed right hand side $\mathbf{u}^\delta$ of (5.4) is $\mathbf{u}^\delta = \mathbf{u} + \delta\,\mathbf{v}$. Here, $\mathbf{u}_{\ell(i,j)} = u(x_i, y_j)$ and $\mathbf{v} = \mathbf{z}/\|\mathbf{z}\|_h$ with $\mathbf{z}$ being a vector with random entries uniformly distributed in $[-1, 1]$. Hence, $\|\mathbf{u} - \mathbf{u}^\delta\|_h = \delta$ measured in the weighted Euclidean norm $\|\cdot\|_h = h\,\|\cdot\|_2$ on $\mathbb{R}^{n^2}$ which approximates the $L^2(\Omega)$-norm.

The scaling requirement (2.9) will be satisfied for $\mathbf{F}'$ provided $\mathbf{c}_k \ge 0$ for all k and $h < 1$, see [15, Sect. 7]. We are thus allowed to use the ν-method, $\nu = 1$, as inner regularization scheme. The tolerances $\{\mu_k\}$ are adapted dynamically during the iteration based on the strategy from [15, Sect. 6] with parameters $\mu_{\mathrm{start}} = 0.1$, $\mu_{\max} = 0.999$ and $\gamma = 0.9$. The results presented below are based on the parameter $R = 3$, see (2.8), and the discretization step size $h = 1/50$.

We ran REGINN on (5.4) with three different starting vectors $\mathbf{c}_0^m$, $(\mathbf{c}_0^m)_{\ell(i,j)} = c_0^m(x_i, y_j)$, $m = 1, 2, 3$, where

$$
\begin{aligned}
c_0^m(x, y) &= 3\left((x - 0.5)^2 + (y - 0.5)^2\right) + 2 + 3\,d_m(x)\,d_m(y), \\
&\quad m = 1, 2, \\
c_0^3(x, y) &= c_0^1(x, y) + \mathrm{random}(x, y).
\end{aligned}
$$

The functions

$$
d_1(t) = 5 \cdot \begin{cases} t & : \quad 0 \le t \le 1/2 \\ 1 - t & : \quad 1/2 < t \le 1 \\ 0 & : \quad \text{otherwise} \end{cases}
$$

$$
\text{and} \quad d_2(t) = \sqrt{10} \cdot \begin{cases} t & : \quad 0 \le t \le 1/2 \\ 2\,(1 - t) & : \quad 1/2 < t \le 1 \\ 0 & : \quad \text{otherwise} \end{cases}
$$

determine the smoothness of c_0^m: $c_0^1 \in H^s(\Omega)$ and $c_0^2 \in H^{s-1}(\Omega)$ for any $s < 3/2$. The third initial guess c_0^3 has no smoothness at all because $\mathrm{random}(\cdot, \cdot)$ is a uniformly distributed random variable with values in $[0, 1]$.

Observe that $(c_0^m - c^\dagger)|_{\partial\Omega} = 0$ for $m = 1, 2, 3$. However, no starting guess c_0^m satisfies (5.3). Therefore, we expect that ($\mathbf{c}^\dagger$ is $c^\dagger$ evaluated at the grid points)

(5.5) $$\|\mathbf{c}_{N(\delta)} - \mathbf{c}^\dagger\|_h = \mathcal{O}(\delta^r) \quad \text{as } \delta \to 0$$

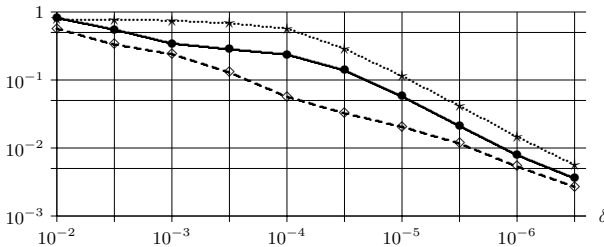where r is clearly smaller than $1/2$ and we expect r to increase with the smoothness of c_0^m.

Fig. 5.1. Relative errors vs. noise level δ (dashed line with $\diamond$: c_0^1, solid line with $\bullet$: c_0^2, dotted line with $\star$: c_0^3)

Figure 5.1 displays the relative errors $\|\mathbf{c}_{N(\delta)} - \mathbf{c}^\dagger\|_h / \|\mathbf{c}^\dagger\|_h$ for $\delta \in \{10^{-(k+1)/2} \mid k = 3, \ldots, 12\}$ and for the initial vectors $\mathbf{c}_0^m$, $m = 1, 2, 3$ (the standardizations of the c_0^m's guarantee initial errors $\|\mathbf{c}_0^m - \mathbf{c}^\dagger\|_h$ of comparable magnitude). As long as the noise dominates the discretization

error, that is, $\delta \geq 10^{-4}$, the decay rates behave exactly as expected. However, as soon as δ is smaller than the discretization error, we basically solve a discrete noise free problem and the errors decrease with optimal order $r = 1/2$.

Do the computational results for $\delta < 10^{-4}$ contradict our theoretical results? The answer is: no! If the noise level is too small the finite dimensional problem cannot be considered anymore a model for the underlying infinite dimensional setting.

In the latter situation we have to interpret (5.4) completely from the finite dimensional point of view. Now, $\mathsf{R}((B^t B)^{\kappa/2}) = \mathsf{N}(B)^\perp$ for any matrix B and any $\kappa > 0$. Since $\mathsf{N}(\mathbf{F}'(\mathbf{c}^\dagger)) = \{0\}$ we thus have the source condition

$$\mathbf{c}^\dagger - \mathbf{c}_0^m \in \mathsf{R}\Big((\mathbf{F}'(\mathbf{c}^\dagger)^t \mathbf{F}'(\mathbf{c}^\dagger))^{1/2} \Big) \;=\; \mathsf{R}\Big(\mathbf{F}'(\mathbf{c}^\dagger)^t \Big), \quad m = 1, 2, 3.$$

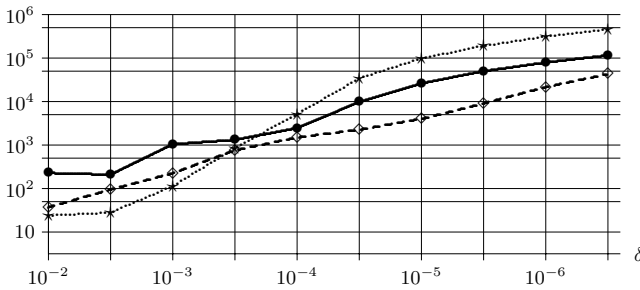Theorem 4.1 tells us to expect (5.5) with r near $1/2$ for δ small enough.

Fig. 5.2. The overall number $\sum_{k=0}^{N(\delta)-1} i_k$ of inner iteration steps vs. noise level δ (dashed line with $\diamond$: $\mathbf{c}_0^1$, solid line with $\bullet$: $\mathbf{c}_0^2$, dotted line with $\star$: $\mathbf{c}_0^3$)

Even for δ small, a smooth initial guess pays. The increasing roughness of the starting iterates can be realized not only in the increasing error value for fixed δ (Fig. 5.1) but also, even more pronounced, in the increasing numerical effort to compute $\mathbf{c}_{N(\delta)}$. In Fig. 5.2 we plotted the overall number $S_\delta := \sum_{k=0}^{N(\delta)-1} i_k$ of inner iteration steps vs. δ. The numerical value of S_δ is a reliable measure for the computational work. For δ small, the rougher the initial error $\mathbf{c}^\dagger - \mathbf{c}_0^m$ becomes, the more computer time is needed to terminate REGINN.

This behavior is supported by the theory. We have that $\mathbf{c}^\dagger - \mathbf{c}_0^m = \mathbf{F}'(\mathbf{c}^\dagger)^t \mathbf{w}^m$ where

$$\mathbf{w}^m \;=\; -\left(\mathbf{A} + \mathrm{diag}(\mathbf{c}^\dagger) \right) \mathrm{diag}(\mathbf{F}(\mathbf{c}^\dagger))^{-1} \left(\mathbf{c}^\dagger - \mathbf{c}_0^m \right).$$

The norm $\|\mathbf{w}^m\|_h$ enters our upper bound for i_k crucially, see [15, Lemmata 4.4 and 4.5] and compare Lemma 3.1. Recall that $\mathbf{A}$ approximates the

Laplacian differential operator. Hence, $\|\mathbf{w}^m\|_h$ grows with the roughness of $\mathbf{c}^\dagger - \mathbf{c}_0^m$.

References

1. B. Blaschke(-Kaltenbacher), A. Neubauer, O. Scherzer: On convergence rates for the iteratively regularized Gauss-Newton method. IMA J. Numer. Anal. **17**, 421–436 (1997)
2. H. Brakhage: On ill-posed problems and the method of conjugate gradients, in Inverse and Ill-Posed Problems, H. W. Engl and C. W. Groetsch, eds., vol. 4 of Notes and Reports in Mathematics in Science and Engeneering, Boston, 1987, Academic Press, pp. 165–175
3. F. Colonius, K. Kunisch: Stability for parameter estimation in two point boundary value problems, J. Reine u. Angewandte Mathematik **370**, 1–29 (1986)
4. P. Deuflhard, H. W. Engl, O. Scherzer: A convergence analysis of iterative methods for the solution of nonlinear ill-posed problems under affinely invariant conditions. Inverse Problems, **14**, 1081–1106 (1998)
5. H. W. Engl, M. Hanke, A. Neubauer: Regularization of Inverse Problems. vol. 375 of Mathematics and Its Applications, Kluwer Academic Publishers, Dordrecht, 1996
6. W. Hackbusch: Elliptic Differential Equations: Theory and Numerical Treatment, vol. 18 of Springer Series in Computational Mathematics, Springer, Heidelberg, 1992
7. M. Hanke: Accelerated Landweber iterations for the solution of ill-posed problems. Numer. Math. **60**, 341–373 (1991)
8. M. Hanke: A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems. Inverse Problems **13**, 79–95 (1997)
9. M. Hanke: Regularizing properties of a truncated Newton-CG algorithm for non-linear inverse problems. Numer. Funct. Anal. Optim. **18**, 971–993 (1998)
10. M. Hanke, A. Neubauer, O. Scherzer: A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. Numer. Math. **72**, 21–37 (1995)
11. B. Hofmann, O. Scherzer: Local ill-posedness and source conditions of operator equations in Hilbert spaces. Inverse Problems **14**, 1189–1206 (1998)
12. B. Kaltenbacher: Some Newton-type methods for the regularization of nonlinear ill-posed problems. Inverse Problems **13**, 729–753 (1997)
13. 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)
14. A. K. Louis: Inverse und schlecht gestellte Probleme. Studienbücher Mathematik, B. G. Teubner, Stuttgart, Germany, 1989
15. A. Rieder: On the regularization of nonlinear ill-posed problems via inexact Newton iterations. Inverse Problems **15**, 309–327 (1999)
16. O. Scherzer: An iterative multi level algorithm for solving nonlinear ill-posed problems. Numer. Math. **80**, 579–600 (1998)
17. U. Tautenhahn: On a general regularization scheme for nonlinear ill-posed problems. Inverse Problems **13**, 1427–1437 (1997)