INEXACT NEWTON REGULARIZATION USING CONJUGATE GRADIENTS AS INNER ITERATION[*]

ANDREAS RIEDER[†]

Abstract. In our papers [*Inverse Problems*, 15 (1999), pp. 309–327] and [*Numer. Math.*, 88 (2001), pp. 347–365] we proposed algorithm `REGINN`, an inexact Newton iteration for the stable solution of nonlinear ill-posed problems. `REGINN` consists of two components: the outer iteration, which is a Newton iteration stopped by the discrepancy principle, and an inner iteration, which computes the Newton correction by solving the linearized system. The convergence analysis presented in both papers covers virtually any *linear* regularization method as inner iteration, especially Landweber iteration, ν-methods, and Tikhonov–Phillips regularization. In the present paper we prove convergence rates for `REGINN` when the conjugate gradient method, which is *nonlinear*, serves as inner iteration. Thereby we add to a convergence analysis of Hanke, who had previously investigated `REGINN` furnished with the conjugate gradient method [*Numer. Funct. Anal. Optim.*, 18 (1997), pp. 971–993]. By numerical experiments we illustrate that the conjugate gradient method outperforms the ν-method as inner iteration.

Key words. nonlinear ill-posed problems, inexact Newton iteration, conjugate gradients, ν-methods, regularization

AMS subject classifications. 65J20, 65J22

DOI. 10.1137/040604029

1. Introduction. Our goal is to find a stable approximate solution of the nonlinear ill-posed problem

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

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

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

The nonnegative *noise level* δ is assumed to be known.

In [10, 11] we proposed algorithm `REGINN` for solving (1.1). As a Newton-type algorithm, `REGINN` updates the actual iterate x_n by adding a correction step s_n^δ obtained from solving a linearization of (1.1):

$$x_{n+1} = x_n + s_n^\delta, \quad n \in \mathbb{N}_0,$$

with an initial guess x_0. For obvious reasons we like to have s_n^δ as close as possible to the exact Newton step

$$s_n^{\mathrm{e}} = x^+ - x_n.$$

Assuming F to be Fréchet differentiable with derivative $F' : \mathsf{D}(F) \to \mathcal{L}(X, Y)$, the exact Newton step satisfies the linear equation

$$(1.3) \qquad F'(x_n)s_n^{\mathrm{e}} = y - F(x_n) - E(x^+, x_n) =: b_n,$$

where $E(v, w) := F(v) - F(w) - F'(w)(v - w)$ is the linearization error.

Unfortunately, the above right-hand side b_n is not available; however, we know a perturbed version

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

Therefore, we determine the correction step s_n^{δ} as a solution of

$$(1.4) \qquad F'(x_n)s = b_n^{\varepsilon}.$$

Here, we have to take into account that the ill-posedness of (1.1) is passed on to (1.4). For instance, if F is completely continuous, then $F'(x_n)$ is a compact operator (see, e.g., Zeidler [13, Proposition 7.33]); hence, (1.4) is ill-posed.

Depending on how s_n^{δ} is stably obtained from (1.4), different methods arise, for instance, the nonlinear Landweber method (Hanke, Neubauer, and Scherzer [6]), the Gauß–Newton method (see, e.g., Bakushinskii [1] and Kaltenbacher [7]), and the Levenberg–Marquardt scheme (Hanke [4]).

In the next few lines we recall briefly how `REGINN` works. First, a regularization scheme is applied to the linear system (1.4), obtaining

$$s_{n,r} := g_r(A_n^* A_n) A_n^* b_n^{\varepsilon},$$

where $A_n = F'(x_n)$ and $g_r : [0, \|A_n\|^2] \to \mathbb{R}$ is the piecewise continuous filter function of the chosen regularization method. The parameter $r \in \mathbb{N}$ is called the regularization parameter. For instance, the filter functions belonging to the Tikhonov–Phillips regularization, the Landweber iteration, and the ν-methods are explicitly known; see, e.g., [2, 12], where more examples can be found. The filter functions g_r of both latter examples are polynomials of degree $r - 1$. The conjugate gradients method (cg-method) can also be described by filter polynomials g_r of degree $r - 1$, which, however, do depend on the right-hand side b_n^{ε}: $g_r(\cdot) = g_r(\cdot, b_n^{\varepsilon})$. Therefore, the cg-method is a *nonlinear* scheme in contrast to the other mentioned examples.

Now we have to select a regularization parameter r_n. In `REGINN` r_n is picked as the smallest number at which the relative (linear) residual is smaller than a given tolerance $\mu_n \in \ ]0, 1]$, that is,

$$(1.5) \qquad \|A_n s_{n,r_n} - b_n^{\varepsilon}\|_Y < \mu_n \|b_n^{\varepsilon}\|_Y \leq \|A_n s_{n,i} - b_n^{\varepsilon}\|_Y, \quad i = 1, \ldots, r_n - 1.$$

The tolerances should not be too small to guarantee existence of r_n; see Lemma 2.1 below. A meaningful strategy to adapt the μ_n's dynamically was proposed in [10]. Setting $s_n^{\delta} := s_{n,r_n}$ we end up with the Newton iteration

$$x_{n+1} = x_n + g_{r_n}(A_n^* A_n) A_n^* b_n^{\varepsilon}, \quad n \in \mathbb{N}_0,$$

which has to be stopped in time to avoid noise amplification. A well-established stopping rule is the discrepancy principle: Choose $R > 0$ and accept iterate x_N as an approximate solution of (1.1) that fulfills

$$(1.6) \qquad \|y^{\delta} - F(x_N)\|_Y \leq R\,\delta < \|y^{\delta} - F(x_k)\|_Y, \quad k = 0, \ldots, N - 1.$$

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

FIG. 1.1.   *Algorithmic realization of* `REGINN` *(*<u>REG</u>*ularization based on* <u>IN</u>*exact* <u>N</u>*ewton iterations).*

For an algorithmic realization of `REGINN`, see Figure 1.1. The inner `repeat`-loop provides the Newton update s_{n,r_n} and the outer `while`-loop implements the Newton iteration stopped by the discrepancy principle.

In [11] we were able to verify (under reasonable assumptions) that `REGINN` with a *linear* regularization scheme $\{g_r\}_{r \in \mathbb{N}}$ is well defined and indeed terminates. Moreover, we proved the existence of a positive $\kappa_{\min} < 1$ such that the source condition[1]

$$(1.7) \qquad x^+ − x_0 \in \mathsf{R}\big(\big|F'(x^+)\big|^\kappa\big) \quad \text{for a } \kappa \in \,]\kappa_{\min}, 1]$$

implies the suboptimal convergence rate[2]

$$(1.8) \qquad \|x^+ − x_{N(\delta)}\|_X = \mathrm{O}\big(\delta^{\,(\kappa−\kappa_{\min})/(1+\kappa)}\big) \quad \text{as } \delta \to 0.$$

In the present paper we will improve upon the convergence results for `REGINN`: We will verify that (1.7) implies (1.8) even when the cg-method serves as inner iteration of `REGINN`. Thus we supplement a convergence analysis of Hanke [5], who had previously investigated `REGINN` with the cg-method as inner iteration: Under a slightly weaker version of our general assumption on the nonlinearity (see (2.1) below), Hanke proved convergence of $\{x_{N(\delta)}\}_{\delta>0}$ to a set of solutions of $F(x) = y$ as $\delta \to 0$.

This paper is structured as follows. In the next two sections we compile facts about `REGINN` and the cg-method which we will need later on in our analysis. In section 4 we show that `REGINN` is well defined under (1.7) and terminates with an approximation to x^+. Then the regularization property (1.8) will be verified (section 5). Finally, we present numerical experiments for a parameter identification model problem and end with concluding remarks in section 7. Some lengthy and technical proofs from sections 3 and 4 are shifted to Appendices A and B, respectively.

---

[1] By $\mathsf{R}(B)$ we denote the range of the operator B, and $|B|$ is the square root of B^*B.

[2] For linear inverse problems $Ax = y^\delta$ the regularization error cannot decrease faster than $\mathrm{O}(\delta^{\kappa/(1+\kappa)})$ as $\delta \to 0$ under the source condition $x^+ − x_0 \in \mathsf{R}(|A|^\kappa)$ in general; see, e.g., [2, section 3.2] or [12, Kapitel 3.2.3]. Regularization schemes attaining the maximal order are therefore called *order-optimal*.

2. General assumptions and termination of the `repeat`-loop. Throughout the paper we assume $F : \mathsf{D}(F) \subset X \to Y$ to be continuously Fréchet differentiable with derivative $F' : \mathsf{D}(F) \to \mathcal{L}(X, Y)$. Moreover, let $x^+ \in \mathsf{D}(F)$, $y = F(x^+)$, $y^\delta \in Y$ with $\|y - y^\delta\|_Y \le \delta$, $A = F'(x^+)$, and $A_n = F(x_n)$.

Our analysis relies heavily on the local factorization (2.1) of F': Let $Q : X \times X \to \mathcal{L}(X, Y)$ be a mapping such that

$$(2.1) \qquad F'(v) = Q(v, w) \, F'(w) \quad \text{with} \quad \|I - Q(v, w)\| \le C_Q \, \|v - w\|_X$$

for all $v, w \in B_\rho(x^+) \subset \mathsf{D}(F)$, the open ball of radius ρ about x^+. Here, C_Q is a positive constant. For a discussion of the nontrivial factorization (2.1) and for examples of meaningful operators satisfying (2.1), we refer to [6, 10, 11], [12, Kapitel 7.3], and the literature cited therein.

Let $C_Q \, \rho < 1$. Then (2.1) gives

$$(2.2) \qquad \|F(v) - F(w)\|_Y \ge (1 - C_Q \, \rho) \, \|F'(w) \, (v - w)\|_Y$$

as well as

$$(2.3) \qquad \|E(v, w)\|_Y \le \omega \, \|F(v) - F(w)\|_Y \quad \text{for all} \ \ v, w \in B_\rho(x^+),$$

where $\omega := C_Q \, \rho / (1 - C_Q \, \rho)$; see [10, section 3] or [12, Lemma 7.3.9]. Observe that $\omega < 1$ for $C_Q \, \rho < 1/2$.

In our subsequent analysis we will frequently use the following estimate: For $x, y \in B_\rho(x^+)$ and $C_Q \rho < 1/2$ we have

$$(2.4) \qquad \| |F'(x)|^{-\kappa} \, |F'(y)|^\kappa \| \le (1 - 2C_Q\rho)^{-\kappa} =: C_{K,\kappa} \quad \text{for all} \ \ \kappa \in [0, 1],$$

which is due to Kaltenbacher [7, Lemma 2.2]; see also [12, Lemma 7.5.16].

Using (2.3) we will bound the data error $\|b_n^\varepsilon - b_n\|_Y$ in terms of δ, ω, and the nonlinear defect

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

For $x_n \in B_\rho(x^+)$ we find

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

We recall a result from [10] which gives conditions on μ_n to stop the `repeat`-loop.

LEMMA 2.1. *Let $\{g_r\}_{r \in \mathbb{N}}$ be the filter function of a linear or nonlinear regularization scheme for which the discrepancy principle returns a well-defined stopping index; that is, for $\tau > 1$ there exists a smallest index r_S with $\|A_n s_{n, r_S} - b_n^\varepsilon\|_Y \le \tau \varepsilon$. Further let (2.1) hold true with $C_Q \, \rho < 1/2$ and assume $x_n \in B_\rho(x^+)$, where $n < N$. If $R \ge (1 + \omega)/(1 - \omega)$, then the `repeat`-loop of algorithm `REGINN` terminates for any*

$$\mu_n \in \left] \omega + \frac{(1 + \omega) \, \delta}{d_n}, \, 1 \right].$$

The lower bound on R in Lemma 2.1 guarantees that the interval for μ_n is nonempty. Since $b_n \in \mathsf{R}(A_n)$ (see (1.3)), all regularization methods mentioned in section 1 (Tikhonov–Phillips, Landweber, ν-method, cg-method) satisfy the requirement of Lemma 2.1; see, e.g., Engl, Hanke, and Neubauer [2, Chapter 4.3] or [12, Kapitel 3.4].

3. The method of conjugate gradients: Preliminaries. Here we recall some basic facts of the cg-method which we will need later in the paper. More details as well as all proofs can be found in, e.g., Engl, Hanke, and Neubauer [2, Chapter 7] or [12, Kapitel 5.3].

Let $T \in \mathcal{L}(X, Y)$ and $\boldsymbol{\eta} \in Y$. The cg-method is an iteration for solving the normal equation $T^*T\boldsymbol{\zeta} = T^*\boldsymbol{\eta}$. Starting with $\boldsymbol{\xi}_0 \in X$ the cg-method produces a sequence $\{\boldsymbol{\xi}_m\}_{m \in \mathbb{N}_0}$ with the minimization property

$$\|\boldsymbol{\eta} - T\boldsymbol{\xi}_m\|_Y = \min\left\{\|\boldsymbol{\eta} - T\boldsymbol{\xi}\|_Y \,\middle|\, \boldsymbol{\xi} \in X,\ \boldsymbol{\xi} - \boldsymbol{\xi}_0 \in U_m\right\}, \quad m \geq 1,$$

where U_m is the mth Krylov space,

$$U_m := \operatorname{span}\left\{T^*r^0, (T^*T)T^*r^0, (T^*T)^2T^*r^0, \ldots, (T^*T)^{m-1}T^*r^0\right\}$$

with $r^0 := \boldsymbol{\eta} - T\boldsymbol{\xi}_0$. Therefore, $\boldsymbol{\xi}_m$, $m \geq 1$, can be expressed by

$$\boldsymbol{\xi}_m = \boldsymbol{\xi}_0 + q_{m-1}(T^*T)T^*(\boldsymbol{\eta} - T\boldsymbol{\xi}_0)$$

with a polynomial q_{m-1} of degree $m - 1$. Closely related to q_{m-1} is the residual polynomial $p_m(\lambda) = 1 - \lambda\, q_{m-1}(\lambda)$ of degree m satisfying

$$\boldsymbol{\eta} - T\boldsymbol{\xi}_m = p_m(TT^*)(\boldsymbol{\eta} - T\boldsymbol{\xi}_0).$$

Both polynomials depend on $\boldsymbol{\eta}$: $q_{m-1}(\cdot) = q_{m-1}(\cdot, \boldsymbol{\eta})$ and $p_m(\cdot) = p_m(\cdot, \boldsymbol{\eta})$. As soon as $T^*(\boldsymbol{\eta} - T\boldsymbol{\xi}_k) = 0$ holds true, the cg-sequence is finite, that is, $\boldsymbol{\xi}_m = \boldsymbol{\xi}_k$ for all $m \geq k$. Accordingly,

$$\boldsymbol{m}_{\mathrm{T}} := \sup\{m \in \mathbb{N} \,|\, T^*(\boldsymbol{\eta} - T\boldsymbol{\xi}_{m-1}) \neq 0\}$$

is called the ultimate termination index of the cg-method ($\boldsymbol{m}_{\mathrm{T}} = \infty$ is allowed and the supremum of the empty set is understood as zero).

The residual polynomials are orthogonal with respect to the inner product $\langle \varphi, \psi \rangle_\Pi := \langle \varphi(T^*T)T^*\boldsymbol{\eta}, \psi(T^*T)T^*\boldsymbol{\eta} \rangle_X$, which is defined on the space of all polynomials:

$$\langle p_i, p_j \rangle_\Pi = 0 \quad \text{for all } 1 \leq i, j \leq \boldsymbol{m}_{\mathrm{T}} \text{ with } i \neq j.$$

The orthogonality of $\{p_m\}_{1 \leq m \leq \boldsymbol{m}_{\mathrm{T}}}$ has several consequences. The residual polynomials satisfy a three-term recursion which can be used to compute $\boldsymbol{\xi}_m$ iteratively from $\boldsymbol{\xi}_{m-1}$ in a rather cheap way; see Figure 3.1. Moreover, p_m has m simple roots $\lambda_{m,j} \in\,]0, \|T\|^2[$, $j = 1, \ldots, m$, which we order by

$$0 < \lambda_{m,1} < \lambda_{m,2} < \cdots < \lambda_{m,m} < \|A\|^2.$$

Because of its normalization $p_m(0) = 1$, p_m decomposes into the following linear factors:

$$(3.1) \qquad p_m(\lambda) = \prod_{j=1}^{m}(1 - \lambda/\lambda_{m,j}) = \prod_{j=1}^{m}\frac{\lambda_{m,j} - \lambda}{\lambda_{m,j}}.$$

Although we know neither q_{m-1} nor p_m explicitly, some useful information about both polynomials is available.

LEMMA 3.1. *For $0 < \Lambda \leq \lambda_{m,1}$ and $1 \leq m \leq \boldsymbol{m}_{\mathrm{T}}$, we have that*

$$\sup_{0 \leq \lambda \leq \Lambda} |q_{m-1}(\lambda, \boldsymbol{\eta})| = q_{m-1}(0, \boldsymbol{\eta}) = -p_m'(0, \boldsymbol{\eta}) = \sum_{j=1}^{m}\lambda_{m,j}^{-1}.$$

cg-algorithm for $T \in \mathcal{L}(X,Y)$, $\boldsymbol{\eta} \in Y$ and starting guess $\boldsymbol{\xi}_0 \in X$.

$r^0 := \boldsymbol{\eta} - T\boldsymbol{\xi}_0$; $p^1 = a^0 := T^* r^0$;

$m := 1$;

```
while (a^{m-1} ≠ 0)
   { q^m := Tp^m;
```
$\qquad \alpha_m := \|a^{m-1}\|_X^2 / \|q^m\|_Y^2$;

$\qquad \boldsymbol{\xi}_m := \boldsymbol{\xi}_{m-1} + \alpha_m \ p^m$;

$\qquad r^m := r^{m-1} - \alpha_m \ q^m$;

$\qquad a^m := T^* r^m$;

$\qquad \beta_m := \|a^m\|_X^2 / \|a^{m-1}\|_X^2$;

$\qquad p^{m+1} := a^m + \beta_m \ p^m$;

$\qquad m := m + 1; \}$

FIG. 3.1. *Conjugate gradients algorithm.*

The next result is proved in Appendix A and will be used twice in our convergence analysis of `REGINN` with the cg-method as inner iteration.

LEMMA 3.2. *Let $\{\boldsymbol{\xi}_m\}_{0 \le m \le m_T}$, $\boldsymbol{\xi}_0 = 0$, be the cg-sequence with respect to $T \in \mathcal{L}(X,Y)$ and $\boldsymbol{\eta} \in Y$. Further, let $\boldsymbol{\xi}$ be in $\mathsf{D}(|T|^{-\mu})$ for a $\mu \in [0,1]$. Then, for any $\nu \in [0, \mu]$, we have that*

$$(3.2) \quad \||T|^{-\nu}(\boldsymbol{\xi}_m - \boldsymbol{\xi})\|_X \le q_{m-1}(0, \boldsymbol{\eta})^{(\nu+1)/2} \big(\|T(\boldsymbol{\xi}_m - \boldsymbol{\xi})\|_Y + \|\boldsymbol{\eta} - T\boldsymbol{\xi}\|_Y \big)$$
$$+ \ q_{m-1}(0, \boldsymbol{\eta})^{(\nu-\mu)/2} \ \||T|^{-\mu}\boldsymbol{\xi}\|_X.$$

4. Termination of `REGINN` with conjugate gradients. The convergence of `REGINN` will be established by bounding the Newton corrections s_{n,r_n} sharply enough. Indeed, we will show that the Newton corrections decrease geometrically in n. Thus, the Newton iterates stay in a ball about x_0.

Recall the assumptions and notation from section 2 and let the cg-method be the inner iteration of `REGINN` exclusively throughout this section.

LEMMA 4.1. *Suppose s_{n,r_n} is well defined. Then*

$$(4.1) \qquad\qquad \|s_{n,r_n}\|_X < 3 \ q_{r_n - 1}(0, b_n^\varepsilon)^{1/2} \ d_n.$$

Proof. We apply Lemma 3.2 with $T = A_n$, $\mu = \nu = 0$, $\boldsymbol{\xi} = 0$, $\boldsymbol{\eta} = b_n^\varepsilon$, $\boldsymbol{\xi}_m = s_{n,r_n} = q_{r_n - 1}(A_n^* A, b_n^\varepsilon) A_n^* b_n^\varepsilon$, that is, $m = r_n$. Thus,

$$\|s_{n,r_n}\|_X \le q_{r_n - 1}(0, b_n^\varepsilon)^{1/2} \left(\|A_n s_{n,r_n}\|_Y + \|b_n^\varepsilon\|_Y \right).$$

We are done by $\|A_n s_{n,r_n}\|_Y \le \|A_n s_{n,r_n} - b_n^\varepsilon\|_Y + \|b_n^\varepsilon\|_Y \le (\mu_n + 1) \|b_n^\varepsilon\|_Y$.   □

In the following we bound each of the factors on the right-hand side of (4.1). From [10, Lemma 4.1] (see also [12, Lemma 7.5.9]) we already know that the nonlinear residuals d_n decrease linearly.

LEMMA 4.2. *Suppose that the nth iterate x_n of `REGINN` is well defined and lies in $B_\rho(x^+)$. Further, let (2.3) hold true with*

$$\omega < \eta/(2 + \eta) \quad \text{for one} \quad \eta < 1.^3$$

---

[3]This restriction is satisfied, for instance, if (2.1) holds true and ρ is small enough.

If, moreover,

$$R \geq \frac{1+\omega}{\eta - (2+\eta)\,\omega} \quad and \quad \mu_n \in \left] \omega + \frac{(1+\omega)\,\delta}{d_n},\ \eta - (1+\eta)\,\omega \right]$$

as well as $x_{n+1} \in B_\rho(x^+)$, then

$$\frac{d_{n+1}}{d_n} = \frac{\|y^\delta - F(x_{n+1})\|_Y}{\|y^\delta - F(x_n)\|_Y} < \frac{\mu_n + \omega}{1 - \omega} \leq \eta.$$

4.1. Bounding $q_{r_n-1}(0, b_n^\varepsilon)$. We assume the existence of $w \in X$ and $\kappa \in [0,1]$ such that

$$(4.2) \qquad\qquad s_0^{\mathrm{e}} = x^+ - x_0 = |A|^\kappa w,$$

where $A = F'(x^+)$. To formulate the bound for $q_{r_k-1}(0, b_k^\varepsilon)$ we introduce the ratio

$$(4.3) \qquad\qquad \tau_k := \mu_k\, d_k / \varepsilon(x_k, \delta),$$

which is greater than 1 under the hypotheses of Lemma 2.1.

LEMMA 4.3. *Let (2.1) hold true with $C_Q \rho < 1/2$ (thus, $\omega < 1$ in (2.3)) and assume that the first $n < N$ iterates $\{x_1, \ldots, x_n\}$ of* REGINN *exist and stay in $B_\rho(x^+)$. Further, let $x_0 \in B_\rho(x^+)$ satisfy the source condition (4.2).*

Then $s_k^{\mathrm{e}} = x^+ - x_k \in \mathsf{D}(|A_k|^{-\kappa})$, $0 \leq k \leq n$. Moreover, if $R \geq (1+\omega)/(1-\omega)$ and if $\mu_k \in\,]\omega + (1+\omega)\delta/d_k, 1]$, $0 \leq k \leq n$, then for any $\Theta \in\,]0,1[$ such that $\Theta \min\{\tau_0, \ldots, \tau_n\} > 1$ we have

$$q_{r_k-1}(0, b_k^\varepsilon)^{(\kappa+1)/2} \leq \frac{a_\Theta}{\Theta\,\tau_k - 1}\ \varepsilon(x_k, \delta)^{-1}\ \||A_k|^{-\kappa} s_k^{\mathrm{e}}\|_X, \quad 0 \leq k \leq n,$$

where a_Θ is a positive constant depending only on Θ and κ.

Proof. See Appendix B for the proof. □

Let us summarize what we found so far. Starting from (4.1) we are able to bound the Newton steps under the assumptions of Lemma 4.3 by

$$(4.4) \qquad \|s_{k,r_k}\|_X < 3 \left(\frac{a_\Theta}{\Theta\,\tau_k - 1} \right)^{1/(\kappa+1)} \varepsilon(x_k, \delta)^{-1/(\kappa+1)}\ \||A_k|^{-\kappa} s_k^{\mathrm{e}}\|_X^{1/(\kappa+1)}\ d_k.$$

4.2. Bounding $\||A_k|^{-\kappa} s_k^{\mathrm{e}}\|_X$. Before we are able to establish termination of REGINN by (4.4), we have to know how $\||A_k|^{-\kappa} s_k^{\mathrm{e}}\|_X$ behaves in k. Since $s_k^{\mathrm{e}} = x^+ - x_k = x^+ - x_{k-1} - s_{k-1,r_{k-1}} = s_{k-1}^{\mathrm{e}} - s_{k-1,r_{k-1}}$ we conclude that

$$\||A_k|^{-\kappa} s_k^{\mathrm{e}}\|_X = \||A_k|^{-\kappa}(s_{k-1}^{\mathrm{e}} - s_{k-1,r_{k-1}})\|_X$$
$$\overset{(2.4)}{\leq} C_{K,\kappa}\, \||A_{k-1}|^{-\kappa}(s_{k-1}^{\mathrm{e}} - s_{k-1,r_{k-1}})\|_X.$$

We estimate the norm on the right by applying Lemma 3.2 with $T = A_{k-1}$, $\mu = \nu = \kappa$, $\boldsymbol{\xi} = s^e_{k-1}$ $\boldsymbol{\eta} = b^\varepsilon_{k-1}$, and $\boldsymbol{\xi}_m = s_{k-1,r_{k-1}}$. Hence,

$$
\begin{aligned}
\||A_k|^{-\kappa} s^e_k\|_X &\leq C_{K,\kappa}\, q_{r_{k-1}-1}(0, b^\varepsilon_{k-1})^{(\kappa+1)/2} \left(\|A_{k-1}s_{k-1,r_{k-1}} - b_{k-1}\|_Y\right. \\
&\qquad\qquad\qquad \left. +\|b^\varepsilon_{k-1} - b_{k-1}\|_Y\right) + \||A_{k-1}|^{-\kappa} s^e_{k-1}\|_X \\
&\leq C_{K,\kappa}\, q_{r_{k-1}-1}(0, b^\varepsilon_{k-1})^{(\kappa+1)/2} \big(\underbrace{\mu_{k-1}\, d_{k-1}}_{\overset{(4.3)}{=}\, \tau_{k-1}\, \varepsilon(x_{k-1},\delta)} +2\,\varepsilon(x_{k-1},\delta)\big) \\
&\qquad\qquad\qquad\qquad\qquad +\||A_{k-1}|^{-\kappa} s^e_{k-1}\|_X \\
&\leq 3\, C_{K,\kappa}\, q_{r_{k-1}-1}(0, b^\varepsilon_{k-1})^{(\kappa+1)/2}\, \tau_{k-1}\, \varepsilon(x_{k-1},\delta) \\
&\qquad\qquad\qquad\qquad\qquad +\||A_{k-1}|^{-\kappa} s^e_{k-1}\|_X \\
&\leq \left(3\, C_{K,\kappa}\, a_\Theta\, \frac{\tau_{k-1}}{\Theta\, \tau_{k-1} - 1} + 1\right) \||A_{k-1}|^{-\kappa} s^e_{k-1}\|_X,
\end{aligned}
$$

where we used Lemma 4.3 in the last step. Inductively, we end up with the following lemma.

LEMMA 4.4. *Let (2.1) hold true with $C_Q \rho < 1/2$ (thus, $\omega < 1$ in (2.3)) and assume that the first $n < N$ iterates $\{x_1, \ldots, x_n\}$ of* REGINN *exist and stay in $B_\rho(x^+)$. Further, choose $R \geq (1+\omega)/(1-\omega)$ and let $x_0 \in B_\rho(x^+)$ satisfy the source condition (4.2).*

If $\mu_k \in\,]\omega + (1+\omega)\delta/d_k, 1]$, $0 \leq k \leq n$, and if $\Theta \in\,]0,1[$ is such that $\Theta\, \tau_k > 1$, $0 \leq k \leq n$, then

$$
(4.5a) \qquad\qquad \||A_k|^{-\kappa} s^e_k\|_X \leq C_{K,\kappa}\, \Lambda^k_n\, \|w\|_X \quad \text{for } k = 0, \ldots, n
$$

with

$$
(4.5b) \qquad \Lambda_n = 1 + 3\, C_{K,\kappa}\, a_\Theta\, \frac{t_n}{\Theta\, t_n - 1} \quad \text{and} \quad t_n = \min\{\tau_0, \ldots, \tau_n\}.
$$

4.3. Termination. We are now able to verify termination of REGINN with conjugate gradients as inner iteration: under reasonable technical assumptions all iterates remain in $B_\rho(x_0)$ and REGINN delivers an approximation $x_{N(\delta)}$ to x^+. The following theorem is the counterpart of Theorem 3.3 from [11] (see also [12, Satz 7.5.14]) and will be proved along the same lines.

THEOREM 4.5. *Let (2.1) hold true with $C_Q \rho < 1/2$ (thus, $\omega < 1$ in (2.3)). Let $\tau > 1$ and let $\Theta \in\,]0,1[$ be such that $\Theta\, \tau > 1$. Set*

$$
\Lambda = 1 + 3\, C_{K,\kappa}\, a_\Theta\, \frac{\tau}{\Theta\, \tau - 1},
$$

where $C_{K,\kappa} = (1 - C_Q \rho)^{-\kappa}$ (see (2.4)), and a_Θ comes from the estimate in Lemma 4.3. Suppose that (2.3) is satisfied with

$$
\omega < \frac{\eta}{\eta + \tau + 1}, \quad \text{where} \quad \eta\, \Lambda < 1.^4
$$

Assume that the starting guess $x_0 \in B_{\rho/2}(x^+)$ is chosen such that the source condition (4.2) applies for κ restricted to $]\log_{1/\eta} \Lambda, 1]$ and that the product $\|w\|_X\, \|y -$

---

[4]This restriction is satisfied, for instance, if (2.1) holds true and ρ is small enough.

$F(x_0)\|_Y^\kappa$ is sufficiently small. If

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

for $k \geq 0$, then there is an $N(\delta) \in \mathbb{N}$ and a $\overline{\delta} > 0$ such that all iterates $\{x_1, \ldots, x_{N(\delta)}\}$ are well defined and stay in $B_\rho(x^+)$ for all noise levels $\delta \in\ ]0, \overline{\delta}]$. Moreover, the final iterate $x_{N(\delta)}$ satisfies the discrepancy principle (1.6) and, for $d_0 > R\,\delta$,

(4.6) $N(\delta) \leq \lfloor \log_\eta(R\,\delta/d_0)\rfloor + 1.$[5]

   Proof. We will prove Theorem 4.5 by induction. Therefore, assume that the first n iterates $\{x_0, \ldots, x_n\}$ are well defined under the hypotheses of Theorem 4.5 and stay in $B_\rho(x^+)$.
   If $d_n \leq R\,\delta$, the iteration will be stopped by (1.6) 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 2.1, r_n and thus the Newton step s_{n,r_n} are well defined. By (4.4) and (4.5),

$$\|s_{n,r_n}\|_X \leq 3\left(\frac{C_{K,\kappa}\,a_\Theta\,\|w\|_X}{\Theta\,\tau_n - 1}\right)^{1/(\kappa+1)} \Lambda_n^{n/(\kappa+1)}\ \varepsilon(x_n, \delta)^{-1/(\kappa+1)}\ d_n.$$

The lower bound on the μ_k's yields $\tau_k \geq \tau > 1$, $k = 0, \ldots, n$ (cf. (4.3)), that is, $\Lambda_n \leq \Lambda$. Moreover, $d_n/\varepsilon(x_n, \delta) \leq 1/\omega$. Taking Lemma 4.2 into account, we obtain

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

where $C_S = 3\big(C_{K,\kappa}\,a_\Theta/(\Theta\,\tau - 1)/\omega\big)^{1/(\kappa+1)}$ and

(4.7) $\sigma(\kappa) := \big(\Lambda\,\eta^\kappa\big)^{1/(\kappa+1)} < 1.$[6]

We define the quantity

(4.8) $\alpha(\delta) := C_S\ \|w\|_X^{1/(\kappa+1)}\ \|F(x_0) - y^\delta\|_X^{\kappa/(\kappa+1)}\Big/\big(1 - \sigma(\kappa)\big).$

In our formulation of Theorem 4.5 we assumed the product $\|w\|_X\|F(x_0) - y\|_X^\kappa$ to be sufficiently small. Now we can be more precise: assume that $\|w\|_X\|F(x_0) - y\|_X^\kappa$ is so small that $\alpha(0) < \rho/2$. Then there exists a $\overline{\delta} > 0$ yielding $\alpha(\delta) < \rho/2$ for all $\delta \in\ ]0, \overline{\delta}]$ and the new iterate $x_{n+1} = x_n + s_{n,r_n} = x_0 + \sum_{k=0}^n s_{k,r_k}$ is in $B_\rho(x^+)$:

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

Further, $d_{n+1} \leq \eta^{n+1}d_0$ uniformly in $\delta \in\ ]0, \overline{\delta}]$ (Lemma 4.2). This completes the inductive step, thereby finishing the proof of Theorem 4.5.   □

----

   [5]Here, $\lfloor t\rfloor \in \mathbb{Z}$ for $t \in \mathbb{R}$ denotes the greatest integer: $\lfloor t\rfloor \leq t < \lfloor t\rfloor + 1$.
   [6]Note that $\sigma(\kappa)$ is smaller than 1 since $\kappa > \log_{1/\eta}\Lambda$.

5. Convergence with rates. Finally, we are able to verify the regularization property of `REGINN` with conjugate gradients as inner iteration, that is, we will show convergence of $x_{N(\delta)}$ to x^+ as the noise level δ decreases.

As an additional tool we will use the *interpolation inequality* (5.1): If $T \in \mathcal{L}(X,Y)$, then

$$(5.1) \qquad \left\|(T^*T)^\alpha x\right\|_X \leq \|(T^*T)^\beta x\|_X^{\alpha/\beta} \, \|x\|_X^{1-\alpha/\beta} \quad \text{for} \ \ 0 < \alpha \leq \beta;$$

see, e.g., [2, 12].

Under the hypotheses of Theorem 4.5 we have to control the reconstruction error $s_k^{\mathrm{e}} = x^+ - x_k$ of the kth iterate, $0 \leq k \leq N(\delta)$:

$$
\begin{aligned}
\|s_k^{\mathrm{e}}\|_X^2 &= \langle |A_k|^\kappa s_k^{\mathrm{e}}, |A_k|^{-\kappa} s_k^{\mathrm{e}}\rangle_X \leq \||A_k|^\kappa s_k^{\mathrm{e}}\|_X \, \||A_k|^{-\kappa} s_k^{\mathrm{e}}\|_X \\
&\overset{(5.1)}{\leq} \||A_k|s_k^{\mathrm{e}}\|_X^\kappa \, \|s_k^{\mathrm{e}}\|_X^{1-\kappa} \, C_{K,\kappa} \, \Lambda^k \, \|w\|_X,
\end{aligned}
$$

where we also applied (4.5) with $\Lambda_k \leq \Lambda$ to obtain the last inequality. Thus,

$$
(5.2)
\begin{aligned}
\|s_k^{\mathrm{e}}\|_X &\leq C_{K,\kappa}^{1/(\kappa+1)} \, \Lambda^{k/(\kappa+1)} \, \|w\|_X^{1/(\kappa+1)} \, \|A_k s_k^{\mathrm{e}}\|_Y^{\kappa/(\kappa+1)} \\
&\overset{(2.2)}{\leq} C_{K,\kappa}^{1/(\kappa+1)} \, \Lambda^{k/(\kappa+1)} \, \|w\|_X^{1/(\kappa+1)} \left(\frac{\|y - F(x_k)\|_Y}{1 - C_Q \, \rho} \right)^{\kappa/(\kappa+1)}.
\end{aligned}
$$

Relying on the above estimate, we are now able to copy the proof of Theorem 4.1 from [11] (see also [12, Satz 7.5.17]) to yield the announced convergence result.

THEOREM 5.1. *Adopt the assumptions of Theorem 4.5; especially, let the source condition* (4.2) *be satisfied with κ restricted to $\,]\log_{1/\eta} \Lambda, 1]$. Additionally, assume that $\alpha(0) < \rho/2$ (see* (4.8)*), as well as $F(x_0) \neq y = F(x^+)$. Then*

$$(5.3) \qquad \|x^+ - x_{N(\delta)}\|_X \leq C_\kappa \, \|w\|_X^{1/(\kappa+1)} \, \delta^{(\kappa-\log_{1/\eta} \Lambda)/(\kappa+1)} \quad \text{as} \ \ \delta \to 0,$$

where C_κ is a suitable constant.

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

$$\|x^+ - x_k\|_X = \mathrm{O}\big(\sigma(\kappa)^k\big) \quad \text{as} \ \ k \to \infty$$

with $\sigma(\kappa)$ from (4.7)*.*

Proof. Plugging $k = N(\delta)$ into (5.2) and taking (1.6) into account give

$$\|s_{N(\delta)}^{\mathrm{e}}\|_X \leq C_{K,\kappa}^{1/(\kappa+1)} \, \|w\|_X^{1/(\kappa+1)} \left(\frac{R+1}{1 - C_Q \, \rho} \right)^{\kappa/(\kappa+1)} \Lambda^{N(\delta)/(\kappa+1)} \, \delta^{\kappa/(\kappa+1)}.$$

Thus, (5.3) follows from (4.6). Convergence in the noise-free setting is obtained from (5.2) in combination with Lemma 4.2. $\square$

6. Computational example. By computational experiments we will demonstrate the increase in numerical efficiency of `REGINN` when replacing the ν-method by the conjugate gradient iteration as inner iteration. We distinguish the two variants by ν-`REGINN` and cg-`REGINN`. Throughout this section let $\nu = 1$.[7]

---

[7] Any $\nu \geq 1$ is admissible [11, Example 2.1]. However, larger ν slow down ν-`REGINN` in the numerical computations presented here.

For our numerical experiments we select a model problem which satisfies our main assumption (2.1). We like to identify the bivariate parameter $c \geq 0$ in the two-dimensional elliptic PDE

(6.1)
$$-\Delta u + c\,u = f \quad \text{in } \Omega,$$
$$u = g \quad \text{on } \partial\Omega$$

from the knowledge of u in the box $\Omega =]0,1[^2$. In (6.1), $-\Delta$ is the Laplacian. Further, let f and g be continuous functions. If u has no zeros in Ω, then c can be recovered explicitly by $c = (f + \Delta u)/u$. Thus, c is uniquely determined by u but does not depend continuously on it. In the case of noise-corrupted data the inversion formula is useless. Further details about our model problem can be found in Hanke, Neubauer, and Scherzer [6, Example 4.2]. Since we already used our model problem for numerical experiments in [10, 11, 12] we will be brief in what follows.

We discretize (6.1) by finite differences with respect to the grid points $(x_i, y_j) = (i\,h, j\,h)$, $1 \leq i, j \leq m$, where $m \in \mathbb{N}$ and $h = 1/(m+1)$ is the discretization step size. Ordering the grid points lexicographically yields the $m^2 \times m^2$ linear system

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

where $\mathbf{A}$ comes from the difference star of the Laplacian $-\Delta$ and where the components of $\mathbf{c} = (\mathbf{c}_1, \dots, \mathbf{c}_{m^2})^t$ are given by $\mathbf{c}_{\ell(i,j)} = c(x_i, y_j)$ with $\ell : \{1, \dots, m\}^2 \to \{1, \dots, m^2\}$ denoting lexicographical ordering. The boundary values g are incorporated into the right-hand side $\mathbf{f}$. From the convergence theory for finite differences (see, e.g., Hackbusch [3]), we know that the solution $\mathbf{u}$ of the above linear system satisfies $\mathbf{u}_{\ell(i,j)} = u(x_i, y_j) + \mathrm{O}(h^2)$ as $h \to 0$ whenever u is sufficiently smooth.

In this discrete setting we like to reconstruct $\mathbf{c}$ from $\mathbf{u}$. Thus, we have to solve the nonlinear equation

(6.2)
$$F(\mathbf{c}) = \mathbf{u}$$

with $F : \mathbb{R}_{\geq 0}^{m^2} \to \mathbb{R}^{m^2}$, $F(\mathbf{c}) = (\mathbf{A} + \operatorname{diag}(\mathbf{c}))^{-1}\,\mathbf{f}$. The function F is differentiable with Jacobi-matrix $F'(\mathbf{c})w = -\big(\mathbf{A} + \operatorname{diag}(\mathbf{c})\big)^{-1}\big(F(\mathbf{c}) \odot w\big)$, where $\odot$ denotes componentwise multiplication of vectors. Moreover, F' can be factorized according to (2.1) in a neighborhood of any $\mathbf{c} > 0$ (componentwise), where also $F(\mathbf{c}) > 0$.[8] In our numerical experiments the parameter to be identified is

$$c^+(x, y) = 1.5\,\sin(2\pi\,x)\,\sin(3\pi\,y) + 3\,\big((x - 0.5)^2 + (y - 0.5)^2\big) + 2.$$

We have chosen f and g such that $u(x, y) = 16\,x\,(x - 1)\,y\,(1 - y) + 1$ solves (6.1) with respect to c^+. As perturbed right-hand side $\mathbf{u}^\delta$ for (6.2) we worked with $\mathbf{u}^\delta = \widetilde{\mathbf{u}} + \delta\,\mathbf{v}$, where $\widetilde{\mathbf{u}}_{\ell(i,j)} = u(x_i, y_j)$ and $\mathbf{v} = \mathbf{z}/\|\mathbf{z}\|_h$. The entries of the random vector $\mathbf{z}$ are uniformly distributed in $[-1, 1]$. Therefore, $\|\mathbf{u} - \mathbf{u}^\delta\|_h \leq \delta + \mathrm{O}(h^2)$ measured in the weighted Euclidean norm $\|\mathbf{z}\|_h = h\,\big(\sum_{i=0}^{m^2} \mathbf{z}_i^2\big)^{1/2}$, which approximates the L^2-norm over Ω.

In all computations below we started `REGINN` with initial guess $\mathbf{c}_0$, where $(\mathbf{c}_0)_{\ell(i,j)} = c_0(x_i, x_j)$ and

$$c_0(x, y) = 3\,\big((x - 0.5)^2 + (y - 0.5)^2\big) + 2 + 128\,x\,(x - 1)\,y\,(1 - y).$$

---

[8]The implication $\mathbf{c} > 0 \Rightarrow F(\mathbf{c}) > 0$ holds true, for instance, if $\mathbf{f} > 0$ since $\mathbf{A} + \operatorname{diag}(\mathbf{c})$ is an M-matrix whose inverse has only nonnegative entries.
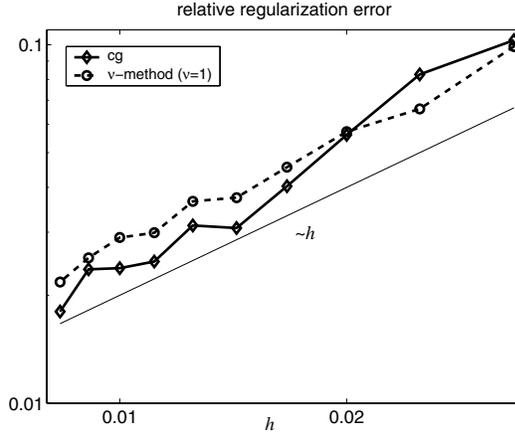
FIG. 6.1. *Relative reconstruction error* (6.4) *of* REGINN *as a function of h (solid line with ◇: cg-method as inner iteration; dashed line with ○: ν-method as inner iteration). The thin solid line indicates decay* O(h) *as h → 0.*

Further, we always used $R = 1.5$, and we adapted the tolerances $\{\mu_k\}$ in (1.5) dynamically according to scheme (6.3) below, which was proposed in [10, section 6] (see also [12, Kapitel 7.5.3.4]): Initialize $\mu_{\text{start}} = 0.1$, $\gamma = 0.9$, $\mu_{\max} = 0.999$, and $\widetilde{\mu}_0 = \widetilde{\mu}_1 := \mu_{\text{start}}$. For $k = 0, \ldots, N(\delta) - 1$ set

$$(6.3) \qquad \mu_k := \mu_{\max} \max \big\{ R\,\delta / \|F(\mathbf{c}_k) - \mathbf{u}^\delta\|_h,\ \widetilde{\mu}_k \big\},$$

where $\mathbf{c}_k$ is the kth iterate and

$$\widetilde{\mu}_k := \begin{cases} 1 - \frac{r_{k-2}}{r_{k-1}}\,(1 - \mu_{k-1}) & : \quad r_{k-1} \geq r_{k-2}, \\ \gamma\,\mu_{k-1} & : \qquad \text{else.} \end{cases}$$

Figure 6.1 shows relative reconstruction errors by ν-REGINN and cg-REGINN. More precisely, we plotted

$$(6.4) \qquad \text{err}(h) := \|\mathbf{c}_{N(\delta(h))} - \mathbf{c}^+\|_h / \|\mathbf{c}^+\|_h \quad \text{with}\ \ \delta(h) = 10\,h^2$$

as a function of $h \in \{(10\,k)^{-1} \,|\, k = 3, \ldots, 12\}$, where $\mathbf{c}^+_{\ell(i,j)} = c^+(x_i, y_j)$ and where $\mathbf{c}_{N(\delta(h))}$ is the output of either ν-REGINN or cg-REGINN. The auxiliary line in Figure 6.1 represents exact decay O(h) as $h \to 0$. Thus, our computations indicate that $\text{err}(h) = \text{O}(h)$ as $h \to 0$. Since $\|\mathbf{u} - \mathbf{u}^{\delta(h)}\|_h \leq \delta(h) + \text{O}(h^2) := \boldsymbol{\delta}(h) = \text{O}(h^2)$ the regularization error achieves maximal order of convergence according to Theorem 5.1, namely, $\text{err}(\boldsymbol{\delta}) = \text{O}(\boldsymbol{\delta}^{1/2})$ as $\boldsymbol{\delta} \to 0$, that is, $\kappa = 1$ and $\Lambda = 1$.

Both variants of REGINN deliver errors in comparable magnitude. In this respect there is not much difference between cg-REGINN and ν-REGINN. However, looking at the numerical efficiency we observe a significant difference. In Figure 6.2 we plotted the ratio

$$(6.5) \qquad q(h) := \frac{\text{cpu-time of } \nu\text{-REGINN to compute } \mathbf{c}_{N(\delta(h))}}{\text{cpu-time of cg-REGINN to compute } \mathbf{c}_{N(\delta(h))}},$$
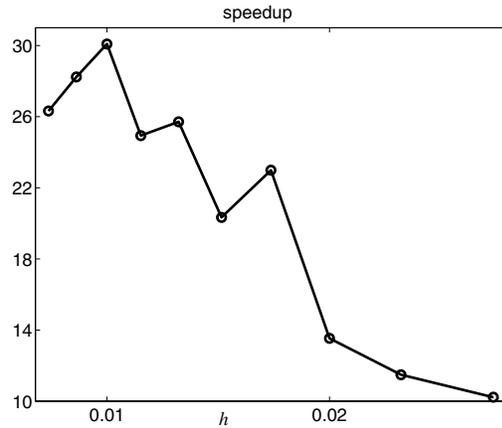
FIG. 6.2. *Speedup q* (6.5) *of cg-*REGINN *over ν-*REGINN.

TABLE 6.1
*Convergence history of ν-*REGINN *for $h = 0.01$, $\delta(h) = 10\, h^2$, with respect to the tolerance selection* (6.3), *where $\mu_{\mathrm{start}} = 0.1$, $\mu_{\mathrm{max}} = 0.999$, and $\gamma = 0.9$. Overall cpu-time: 240.72 seconds.*

k	μ_{k-1}	r_{k-1}	d_k/d_{k-1}	$d_k/(R\,\delta)$	e_k
1	0.0999	40	0.2464	43.68	0.4280
2	0.0999	75	0.1128	4.925	0.2204
3	0.5194	99	0.5153	2.538	0.1311
4	0.6353	105	0.6324	1.605	0.0836
5	0.6555	176	0.6550	1.051	0.0443
6	0.9504	242	0.9503	0.999	0.0290

TABLE 6.2
*Convergence history of cg-*REGINN *for $h = 0.01$, $\delta(h) = 10\, h^2$, with respect to the tolerance selection* (6.3), *where $\mu_{\mathrm{start}} = 0.1$, $\mu_{\mathrm{max}} = 0.999$, and $\gamma = 0.9$. Overall cpu-time: 8.00 seconds.*

k	μ_{k-1}	r_{k-1}	d_k/d_{k-1}	$d_k/(R\,\delta)$	e_k
1	0.0999	1	0.3997	70.86	0.5598
2	0.0999	1	0.1188	8.418	0.3090
3	0.1187	6	0.1223	1.030	0.0239
4	0.9704	1	0.9629	0.991	0.0238

where we did not count cpu-time for preprocessing steps performed by both variants.[9] Figure 6.2 reveals that cg-REGINN is 10 to 30 times faster than ν-REGINN in our example.

Tables 6.1 and 6.2 record the convergence history of ν-REGINN and cg-REGINN in full detail for the discretization step size $h = 0.01$. In both tables

$$d_k := \|F(\mathbf{c}_k) - \mathbf{u}^{\delta(h)}\|_h \quad \text{and} \quad e_k := \|\mathbf{c}_k - \mathbf{c}^+\|_h / \|\mathbf{c}^+\|_h$$

denote the nonlinear defect and the relative L^2-error of the kth iterate, respectively.

Among all Krylov-subspace methods the conjugate gradient iteration is the most efficient one when the discrepancy principle is the used stopping rule; see, e.g., [2,

---

[9]The experiments were carried out under MATLAB 6.5 on a 2.6GHz Intel Pentium 4 processor.

Chapter 7.1] or [12, Kapitel 5.3.6]. As expected, cg-`REGINN` outperforms ν-`REGINN` since it takes much fewer inner iterations to yield the correction step which we can observe clearly in the tables (one iteration step of the cg-methods is only slightly more expensive than one iteration step of the ν-method).

7. Concluding remarks. In this paper we proved local convergence with rates for a regularization scheme of inexact Newton type with the cg-method as inner iteration. Theoretical aspects are emphasized; ideas and techniques have been presented to cope with the nonlinearity of the conjugate gradient iteration.

As far as the author knows, the restrictive factorization assumption (2.1) has not been verified for real applications such as, e.g., impedance tomography, ultrasound tomography, and SPECT (single photon emission computed tomography). Therefore the most pressing improvement of the presented analysis is to weaken or to get rid of (2.1).

Nevertheless the practitioner may benefit from our theoretical results in at least two ways: (1) The adaptive tolerance selection scheme (6.3) has a sound justification for cg-`REGINN` and can be expected to perform well also for more general nonlinearities. (2) A potential convergence analysis of cg-`REGINN` for a specific application, which does not fall into the general category considered, can be based upon techniques developed here.

Appendix A. Proof of Lemma 3.2. For the sake of simplicity we only prove Lemma 3.2 for a compact operator T (the general result will follow by integration over the spectral family of T^*T). Most of our arguments have been used before by Plato [9, Lemma 5.4] (see also [12, Lemma 5.3.11]) to prove another error estimate for the cg-method.

Let $\{(\sigma_j; v_j, u_j) | j \in \mathbb{N}\} \subset {]0, \infty[} \times X \times Y$ be the singular system of T, that is, $Tx = \sum_{j=1}^{\infty} \sigma_j \langle x, v_j \rangle_X u_j$ with $\lim_{j \to \infty} \sigma_j = 0$ monotonically, and $\{v_j\}$ and $\{u_j\}$ are orthonormal bases in $\mathsf{N}(T)^{\perp}$ and $\overline{\mathsf{R}(T)}$, respectively.[10]

We introduce the spectral family $\{\mathcal{E}_\Lambda\}_{\Lambda > 0} \subset \mathcal{L}(X)$ of T^*T by[11]

$$(A.1) \qquad \mathcal{E}_\Lambda x := \sum_{j \in \mathcal{J}(\Lambda)} \langle x, v_j \rangle_X \, v_j + P_{\mathsf{N}(T)} x, \quad \mathcal{J}(\Lambda) := \{j \in \mathbb{N} \,|\, \sigma_j^2 \leq \Lambda\},$$

and start with

$$\||T|^{-\nu}(\boldsymbol{\xi}_m - \boldsymbol{\xi})\|_X \leq \||T|^{-\nu}(I - \mathcal{E}_\Lambda)(\boldsymbol{\xi}_m - \boldsymbol{\xi})\|_X + \||T|^{-\nu}\mathcal{E}_\Lambda(\boldsymbol{\xi}_m - \boldsymbol{\xi})\|_X.$$

We proceed by

$$\begin{aligned}
\||T|^{-\nu}(I - \mathcal{E}_\Lambda)(\boldsymbol{\xi}_m - \boldsymbol{\xi})\|_X^2 &= \sum_{k=1}^{\infty} \sigma_k^{-2(\nu+1)} \left|\left\langle |T|(I - \mathcal{E}_\Lambda)(\boldsymbol{\xi}_m - \boldsymbol{\xi}), v_k \right\rangle_X\right|^2 \\
&= \sum_{k \notin \mathcal{J}(\Lambda)} \sigma_k^{-2(\nu+1)} \left|\left\langle |T|(\boldsymbol{\xi}_m - \boldsymbol{\xi}), v_k \right\rangle_X\right|^2 \\
&\leq \Lambda^{-(\nu+1)} \|T(\boldsymbol{\xi}_m - \boldsymbol{\xi})\|_Y^2
\end{aligned}$$

---

[10]$\mathsf{N}(B)$ and $\mathsf{R}(B)$ denote the null space and the range of a linear operator, respectively.
[11]$P_M \in \mathcal{L}(X)$ denotes the orthogonal projection onto the closed subspace M of X.

and

$$\|\,|T|^{-\nu}\mathcal{E}_\Lambda(\boldsymbol{\xi}_m - \boldsymbol{\xi})\|_X$$

$$\leq \|\,|T|^{-\nu}\mathcal{E}_\Lambda p_m(T^*T, \boldsymbol{\eta})\boldsymbol{\xi}\|_X + \|\,|T|^{-\nu}\mathcal{E}_\Lambda q_{m-1}(T^*T, \boldsymbol{\eta})T^*(\boldsymbol{\eta} - T\boldsymbol{\xi})\|_X$$

$$\leq \|\mathcal{E}_\Lambda |T|^{\mu-\nu}p_m(T^*T, \boldsymbol{\eta})\|_X \,\|\,|T|^{-\mu}\boldsymbol{\xi}\|_X$$
$$+ \|\,|T|^{-\nu}\mathcal{E}_\Lambda q_{m-1}(T^*T, \boldsymbol{\eta})T^*\|_X \,\|\boldsymbol{\eta} - T\boldsymbol{\xi}\|_X$$

$$\leq \|\,|T|^{-\mu}\boldsymbol{\xi}\|_X \sup_{0\leq\lambda\leq\Lambda} \lambda^{(\mu-\nu)/2}\,|p_m(\lambda, \boldsymbol{\eta})|$$
$$+ \|\mathcal{E}_\Lambda q_{m-1}(T^*T, \boldsymbol{\eta})|T|^{-\nu}T^*\|_X \,\|\boldsymbol{\eta} - T\boldsymbol{\xi}\|_X.$$

Further,

$$\|\mathcal{E}_\Lambda q_{m-1}(T^*T, \boldsymbol{\eta})|T|^{-\nu}T^*\|^2$$

$$= \|\mathcal{E}_\Lambda\, q_{m-1}(T^*T, \boldsymbol{\eta})|T|^{2(1-\nu)}q_{m-1}(T^*T, \boldsymbol{\eta})\mathcal{E}_\Lambda\|$$

$$\leq \|\mathcal{E}_\Lambda\, q_{m-1}(T^*T, \boldsymbol{\eta})\| \,\|\,|T|^{2(1-\nu)}q_{m-1}(T^*T, \boldsymbol{\eta})\mathcal{E}_\Lambda\|$$

$$\leq \sup_{0\leq\lambda\leq\Lambda} |q_{m-1}(\lambda, \boldsymbol{\eta})| \sup_{0\leq\lambda\leq\Lambda} \lambda^{1-\nu}\,|q_{m-1}(\lambda, \boldsymbol{\eta})|.$$

By Lemma 3.1 we have

$$\sup_{0\leq\lambda\leq\Lambda} |q_{m-1}(\lambda, \boldsymbol{\eta})| \leq q_{m-1}(0, \boldsymbol{\eta}) \quad \text{whenever } 0 < \Lambda \leq \lambda_{m,1}.$$

The representation (3.1) of p_m shows that $0 \leq p_m(\lambda) \leq 1$ for $\lambda \in [0, \lambda_{m,1}]$. Since $p_m(\lambda) = 1 - \lambda\, q_{m-1}(\lambda)$ we derive that

$$\sup_{0\leq\lambda\leq\Lambda} \lambda^{1-\nu}\,|q_{m-1}(\lambda, \boldsymbol{\eta})| \leq \Big(\sup_{0\leq\lambda\leq\Lambda} \lambda\,|q_{m-1}(\lambda, \boldsymbol{\eta})|\Big)^{1-\nu} \Big(\sup_{0\leq\lambda\leq\Lambda} |q_{m-1}(\lambda, \boldsymbol{\eta})|\Big)^{\nu}$$

$$\leq q_{m-1}(0, \boldsymbol{\eta})^{\nu}$$

whenever $0 < \Lambda \leq \lambda_{m,1}$. Finally, for $0 < \Lambda \leq \lambda_{m,1}$, we obtain that

$$\|\,|T|^{-\nu}(\boldsymbol{\xi}_m - \boldsymbol{\xi})\|_X \leq \Lambda^{-(\nu+1)/2}\,\|T(\boldsymbol{\xi}_m - \boldsymbol{\xi})\|_Y + \|\,|T|^{-\mu}\boldsymbol{\xi}\|_X\,\Lambda^{(\mu-\nu)/2}$$
$$+ q_{m-1}(0, \boldsymbol{\eta})^{(\nu+1)/2}\,\|\boldsymbol{\eta} - T\boldsymbol{\xi}\|_X,$$

which yields the stated inequality (3.2) by setting $\Lambda = 1/q_{m-1}(0, \boldsymbol{\eta})$. This choice for Λ is admissible since $1/q_{m-1}(0, \boldsymbol{\eta}) < \lambda_{m,1}$; see Lemma 3.1.

Appendix B. Proof of Lemma 4.3. Before we can prove Lemma 4.3, we need some auxiliary results (Lemmas B.1 and B.2 and Corollary B.3 below). Here we rely on arguments laid out by Plato [9, Kapitel 5] and Nemirovskii [8] (see also [12, Kapitel 5.3]).

Suppose that the first n iterates $\{x_1, \ldots, x_n\}$ of REGINN exist and stay in $B_\rho(x^+)$. Point of departure is the inequality

(B.1) $$\|b_k^\varepsilon - A_k s_{k,m}\|_Y \leq \|\mathcal{F}_{\lambda_{m,1}}\varphi_m(A_k A_k^*, b_k^\varepsilon)b_k\|_Y + \varepsilon, \quad 1 \leq m \leq \boldsymbol{m}_\mathrm{T},$$

where $0 \leq k \leq n$, $\boldsymbol{m}_\mathrm{T} = \boldsymbol{m}_\mathrm{T}(k)$, and $\mathcal{F}_\Lambda \in \mathcal{L}(Y)$, $\Lambda > 0$, is the orthogonal projection

(B.2) $$\mathcal{F}_\Lambda y := \sum_{i\in\mathcal{J}(\Lambda)} \langle y, u_i\rangle_Y\, u_i + P_{\mathsf{N}(A_k^*)}y$$

with index set $\mathcal{J}$ as in (A.1). In defining $\mathcal{F}_\Lambda$ we used the singular system of A_k.[12] Further, the function $\varphi_m(\cdot, b_k^\varepsilon) \in C(\mathbb{R})$ in (B.1) is

$$\varphi_m(\lambda, b_k^\varepsilon) := p_m(\lambda, b_k^\varepsilon) \left(\frac{\lambda_{m,1}}{\lambda_{m,1} - \lambda} \right)^{1/2},$$

where $\lambda_{m,1}$ is the smallest zero of the mth residual polynomial $p_m(\cdot, b_k^\varepsilon)$ of the cg-method with respect to A_k and b_k^ε. A proof of (B.1) is presented, e.g., by Engl, Hanke, and Neubauer [2, Proof of Theorem 7.10].

As $s_k^e = x^+ - x_k = s_0^e - \sum_{j=0}^{k-1} s_{j,r_j}$ we obtain

$$s_k^e \overset{(4.2)}{=} |A|^\kappa w - \sum_{j=0}^{k-1} A_j^* q_{r_j-1}(A_j A_j^*, b_j^\varepsilon) b_j^\varepsilon.$$

Note that $s_k^e \in D(|A_i|^{-\kappa})$, $i = 0, \dots, n$. Indeed, in using $A_j = Q_{j,i} A_i$ with $Q_{j,i} = Q(x_j, x_i)$ (see (2.1)), we obtain that

$$\||A_i|^{-\kappa} s_k^e\|_X \;\leq\; \||A_i|^{-\kappa}|A|^\kappa w\|_X + \sum_{j=0}^{k-1} \||A_i|^{-\kappa} A_j^* q_{r_j-1}(A_j A_j^*, b_j^\varepsilon) b_j^\varepsilon\|_X$$

$$\overset{(2.4)}{\leq} C_{K,\kappa} \|w\|_X + \underbrace{\||A_i|^{-\kappa} A_i^*\|}_{=\|A_i\|^{1-\kappa}} \sum_{j=0}^{k-1} \|Q_{j,i}^* q_{r_j-1}(A_j A_j^*, b_j^\varepsilon) b_j^\varepsilon\|_Y.$$

Thus, by $A_k s_k^e = b_k$,

(B.3)
$$\|\mathcal{F}_{\lambda_{m,1}} \varphi_m(A_k A_k^*, b_k^\varepsilon) b_k\|_Y$$
$$= \|\mathcal{F}_{\lambda_{m,1}} \varphi_m(A_k A_k^*, b_k^\varepsilon) A_k |A_k|^\kappa |A_k|^{-\kappa} s_k^e\|_Y$$
$$\leq \sup_{0 \leq \lambda \leq \lambda_{m,1}} \lambda^{(\kappa+1)/2} \varphi_m(\lambda, b_k^\varepsilon) \||A_k|^{-\kappa} s_k^e\|_X.$$

Techniques from elementary calculus, together with Lemma 3.1, yield

$$\sup_{0 \leq \lambda \leq \lambda_{m,1}} \lambda^{(\kappa+1)/2} \varphi_m(\lambda, b_k^\varepsilon) \leq 2 \, q_{m-1}(0, b_k^\varepsilon)^{-(\kappa+1)/2}$$

for $\kappa \in [0,1]$; see, e.g., [2, (7.8)] or [12, (5.65)]. Hence,

(B.4) $\qquad \|\mathcal{F}_{\lambda_{m,1}} \varphi_m(A_k A_k^*, b_k^\varepsilon) b_k\|_Y \leq 2 \, q_{m-1}(0, b_k^\varepsilon)^{-(\kappa+1)/2} \||A_k|^{-\kappa} s_k^e\|_X.$

Finally, (B.1) and (B.4) yield the following lemma.

LEMMA B.1. *Let (2.1) hold true and assume that the first n iterates $\{x_1, \dots, x_n\}$ of* REGINN *exist and stay in $B_\rho(x^+)$. If $x_0 \in B_\rho(x^+)$ satisfies (4.2), then, for $0 \leq k \leq n$ and $1 \leq m \leq \boldsymbol{m}_\mathrm{T}(k)$,*

$$\|b_k^\varepsilon - A_k s_{k,m}\|_Y \leq \varepsilon + 2 \, q_{m-1}(0, b_k^\varepsilon)^{-(\kappa+1)/2} \||A_k|^{-\kappa} s_k^e\|_X.$$

We need a second auxiliary lemma.

---

[12]More precisely, $\{\mathcal{F}_\Lambda\}_{\Lambda>0}$ is the spectral family of $A_k A_k^*$.

LEMMA B.2. *Let (2.1) hold true and assume that the first n iterates $\{x_1, \ldots, x_n\}$ of* REGINN *exist and stay in $B_\rho(x^+)$. Further, let $x_0 \in B_\rho(x^+)$ satisfy (4.2). Choose $\vartheta > 2$ and $2 < r \le 2(\vartheta - 1)$. Let $0 \le k \ge n$ and $1 \le m \le \boldsymbol{m}_\mathrm{T}(k)$.*

If $\vartheta\, q_{m-2}(0, b_k^\varepsilon) \le q_{m-1}(0, b_k^\varepsilon)$, then

$$\frac{r-2}{r-1}\, \|b_k^\varepsilon - A_k s_{k,m-1}\|_Y \le \varepsilon + \alpha^{(\kappa+1)/2}\, q_{m-1}(0, b_k^\varepsilon)^{-(\kappa+1)/2}\, \||A_k|^{-\kappa} s_k^\mathrm{e}\|_X,$$

where $\alpha = r/(1 - \vartheta^{-1})$.

Proof. Under our assumptions Plato [9, (5.16)] (see also [12, (5.67)]) established the bound

$$\frac{r-2}{r-1}\, \|b_k^\varepsilon - A_k s_{k,m-1}\|_Y \le \|\mathcal{F}_{r\lambda_{m,1}} p(A_k A_k^*) b_k^\varepsilon\|_Y$$

with the polynomial $p(\lambda) = p_m(\lambda, b_k^\varepsilon)/(1 - \lambda/\lambda_{m,1})$ of degree $m - 1$. The triangle inequality leads to

$$\|\mathcal{F}_{r\lambda_{m,1}} p(A_k A_k^*) b_k^\varepsilon\|_Y \le \varepsilon \sup_{0 \le \lambda \le r\lambda_{m,1}} |p(\lambda)| + \|\mathcal{F}_{r\lambda_{m,1}} p(A_k A_k^*) b_k\|_Y.$$

To bound $\|\mathcal{F}_{r\lambda_{m,1}} p(A_k A_k^*) b_k\|_Y$ we are able to apply exactly the same arguments as were used in estimating $\|\mathcal{F}_{\lambda_{m,1}} \varphi_m(A_k A_k^*) b_k\|_Y$; cf. (B.3). Accordingly, if

(B.5a)
$$\sup_{0 \le \lambda \le r\lambda_{m,1}} |p(\lambda)| \le 1$$

as well as

(B.5b)
$$\sup_{0 \le \lambda \le r\lambda_{m,1}} \lambda^{(\kappa+1)/2} |p(\lambda)| \le \alpha^{(\kappa+1)/2}\, q_{m-1}(0, b_k^\varepsilon)^{-(\kappa+1)/2},$$

then Lemma B.2 is true. Assume, for the moment, that

(B.6)
$$r\lambda_{m,1} < 2\, \lambda_{m,2} \quad \text{and} \quad \lambda_{m,1} \le q_{m-1}(0, b_k^\varepsilon)^{-1}/(1 - \vartheta^{-1})$$

hold true. The left inequality yields $r\lambda_{m,1}/\lambda_{m,j} < 2$, $j = 2, \ldots, m$, whence

$$|p(\lambda)| = \prod_{j=2}^{m} |1 - \lambda/\lambda_{m,j}| < 1 \quad \text{for } 0 < \lambda \le r\lambda_{m,1}.$$

Therefore, (B.6) implies (B.5) and we are left with verifying (B.6).

First we look into the estimate on the right in (B.6). Since the residual polynomials $\{p_m(\cdot, b_k^\varepsilon)\}_{1 \le m \le \boldsymbol{m}_\mathrm{T}}$ are orthogonal, their zeros interlace, that is, $\lambda_{m-1,j} < \lambda_{m,j+1}$, $j = 1, \ldots, m - 1$. By Lemma 3.1 we therefore have

(B.7)
$$q_{m-1}(0, b_k^\varepsilon) = \lambda_{m,1}^{-1} + \sum_{j=1}^{m-1} \lambda_{m,j+1}^{-1}$$
$$< \lambda_{m,1}^{-1} + \sum_{j=1}^{m-1} \lambda_{m-1,j}^{-1} = \lambda_{m,1}^{-1} + q_{m-2}(0, b_k^\varepsilon).$$

The hypothesis $q_{m-2}(0, b_k^\varepsilon) \le \vartheta^{-1} q_{m-1}(0, b_k^\varepsilon)$ implies $q_{m-1}(0, b_k^\varepsilon) \le \lambda_{m,1}^{-1} + \vartheta^{-1} q_{m-1}(0, b_k^\varepsilon)$ and thus the right inequality in (B.6). Since

$$(\vartheta - 1)\, \lambda_{m-1,1}^{-1} \quad < \quad (\vartheta - 1) \sum_{j=1}^{m-1} \lambda_{m-1,j}^{-1} \;=\; \vartheta\, q_{m-2}(0, b_k^\varepsilon) - q_{m-2}(0, b_k^\varepsilon)$$

$$\overset{\substack{\text{by} \\ \text{assumpt.}}}{\le} \quad q_{m-1}(0, b_k^\varepsilon) - q_{m-2}(0, b_k^\varepsilon) \overset{(B.7)}{<} \lambda_{m,1}^{-1},$$

we obtain $\lambda_{m,1} < (\vartheta - 1)^{-1} \lambda_{m-1,1} < (\vartheta - 1)^{-1} \lambda_{m,2}$ (interlacing property). In view of $r/(\vartheta - 1) \le 2$ we conclude that the left inequality in (B.6) holds true as well, thereby finishing the proof of Lemma B.2. $\qquad\square$

Both latter lemmas merge in the next corollary.

COROLLARY B.3. *Let* (2.1) *hold true and assume the first n iterates* $\{x_1, \ldots, x_n\}$ *of* REGINN *exist and stay in* $B_\rho(x^+)$. *Further, let* $x_0 \in B_\rho(x^+)$ *satisfy* (4.2). *Then, to any* $\Theta \in\,]0, 1[$, *there exists a number* a_Θ *such that, for* $0 \le k \le n$,

$$\Theta\, \|b_k^\varepsilon - A_k s_{k,m-1}\|_Y \le \varepsilon + a_\Theta\, q_{m-1}(0, b_k^\varepsilon)^{-(\kappa+1)/2}\, \||A_k|^{-\kappa} s_k^{\mathrm{e}}\|_X,$$

where $1 \le m \le \boldsymbol{m}_\mathrm{T}(k)$. *The number* a_Θ *only depends on* Θ *and* κ.

Proof. There is exactly one $r = r(\Theta) > 2$ such that $\Theta = \frac{r-2}{r-1}$. Let this r be fixed and define $\vartheta = r/2 + 1 > 2$, that is, $r = 2(\vartheta - 1)$. Exactly one of the following two cases holds true.

1. In the case of $\vartheta\, q_{m-2}(0, b_k^\varepsilon) \le q_{m-1}(0, b_k^\varepsilon)$ the assertion follows immediately from Lemma B.2 when setting

$$a_{\Theta,1} := \left(\frac{r}{1 - \vartheta^{-1}}\right)^{(\kappa+1)/2} = \left(\frac{r}{1 - (r/2 + 1)^{-1}}\right)^{(\kappa+1)/2}.$$

2. In the case of $q_{m-1}(0, b_k^\varepsilon) < \vartheta\, q_{m-2}(0, b_k^\varepsilon)$ we argue with Lemma B.1. Since $\Theta < 1$ and $q_{m-2}(0, b_k^\varepsilon)^{-1} < \vartheta\, q_{m-1}(0, b_k^\varepsilon)^{-1}$ we have

$$\Theta\, \|b_k^\varepsilon - A s_{k,m-1}\|_Y \le \varepsilon + 2\, q_{m-2}(0, b_k^\varepsilon)^{-(\kappa+1)/2}\, \||A_k|^{-\kappa} s_k^{\mathrm{e}}\|_X$$

$$< \varepsilon + a_{\Theta,2}\, q_{m-1}(0, b_k^\varepsilon)^{-(\kappa+1)/2}\, \||A_k|^{-\kappa} s_k^{\mathrm{e}}\|_X$$

with

$$a_{\Theta,2} := 2\, \vartheta^{(\kappa+1)/2} = 2\, (r/2 + 1)^{(\kappa+1)/2}.$$

So, the assertion of Corollary B.3 is verified with $a_\Theta = \max\{a_{\Theta,1}, a_{\Theta,2}\}$. $\qquad\square$

Finally we are in a position to verify Lemma 4.3: The μ_k's and R satisfy the requirements of Lemma 2.1. Hence, $\tau_k > 1$ (see (4.3)), and $\tau_k\, \varepsilon \le \|b_k^\varepsilon - A_k s_{k,r_k-1}\|_Y$ (see (1.5)). Since $1 \le r_k \le \boldsymbol{m}_\mathrm{T}(k)$ we obtain

$$\Theta\, \tau_k\, \varepsilon \le \Theta\, \|b_k^\varepsilon - A_k s_{k,r_k-1}\|_Y \le \varepsilon + a_\Theta\, q_{r_k-1}(0, b_k^\varepsilon)^{-(\kappa+1)/2}\, \||A_k|^{-\kappa} s_k^{\mathrm{e}}\|_X$$

by Corollary B.3. A simple rearrangement of terms yields the assertion of Lemma 4.3.

REFERENCES

[1] A. B. BAKUSHINSKII, *The problem of the convergence of the iteratively regularized Gauss-Newton method*, Comput. Math. Phys., 32 (1992), pp. 1353–1359.

[2] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Math. Appl. 375, Kluwer Academic, Dordrecht, The Netherlands, 1996.

[3] W. Hackbusch, *Elliptic Differential Equations: Theory and Numerical Treatment*, Springer Ser. Comput. Math. 18, Springer-Verlag, Heidelberg, 1992.

[4] M. Hanke, *A regularizing Levenberg-Marquardt scheme, with applications to inverse ground-water filtration problems*, Inverse Problems, 13 (1997), pp. 79–95.

[5] M. Hanke, *Regularizing properties of a truncated Newton-CG algorithm for non-linear inverse problems*, Numer. Funct. Anal. Optim., 18 (1997), pp. 971–993.

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

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

[8] A. S. Nemirovskii, *The regularizing properties of the adjoint gradient method in ill-posed problems*, USSR Comp. Math. Math. Phys., 26 (1986), pp. 7–16.

[9] R. Plato, *Über die Diskretisierung und Regularisierung schlecht gestellter Probleme*, Ph.D. thesis, Fachbereich Mathematik der Technischen Universität Berlin, Berlin, 1990. Available online from www.math.tu-berlin.de/~plato/promo.ps.

[10] A. Rieder, *On the regularization of nonlinear ill-posed problems via inexact Newton iterations*, Inverse Problems, 15 (1999), pp. 309–327.

[11] A. Rieder, *On convergence rates of inexact Newton regularizations*, Numer. Math., 88 (2001), pp. 347–365.

[12] A. Rieder, *Keine Probleme mit Inversen Problemen*, Vieweg, Wiesbaden, Germany, 2003.

[13] E. Zeidler, *Nonlinear Functional Analysis and Its Applications* I: *Fixed-Point Theorems*, Springer-Verlag, New York, 1993.