UNIVERSITÄT KAISERSLAUTERN

Regularizing Properties of a Truncated Newton-CG Algorithm for Nonlinear Inverse Problems

Martin Hanke

Preprint Nr. 280

ISSN 0943-8874

Juni 1996

FACHBEREICH MATHEMATIK

REGULARIZING PROPERTIES OF A TRUNCATED NEWTON-CG ALGORITHM FOR NONLINEAR INVERSE PROBLEMS

MARTIN HANKE*

Abstract. This paper develops truncated Newton methods as an appropriate tool for nonlinear inverse problems which are ill-posed in the sense of Hadamard. In each Newton step an approximate solution for the linearized problem is computed with the conjugate gradient method as an inner iteration. The conjugate gradient iteration is terminated when the residual has been reduced to a prescribed percentage. Under certain assumptions on the nonlinear operator it is shown that the algorithm converges and is stable if the discrepancy principle is used to terminate the outer iteration. These assumptions are fulfilled, e.g., for the inverse problem of identifying the diffusion coefficient in a parabolic differential equation from distributed data.

Key words. Nonlinear ill-posed problems, inexact Newton method, conjugate gradient method, regularization, convergence analysis.

AMS subject classifications. 65J10, 65J20

1. Introduction. The stable numerical solution of nonlinear inverse problems

$$F(a) = u\,, \qquad F : \mathcal{D}(F) \subset \mathcal{X} \to \mathcal{Y}\,, \tag{1.1}$$

is one important matter of scientific computing. For example, consider the differential equation

$$u_t - \operatorname{div}(a \operatorname{grad} u) = f\,, \quad \text{for } x \in \Omega \subset \mathbb{R}^N,\ t > 0\,, \tag{1.2}$$

with suitable boundary conditions. An important inverse problem in ground water filtration is the reconstruction of the transmissivity coefficient a as a function of x in a porous medium Ω from (partial) knowledge of the piezometric head u in Ω within a given time inverval $\mathcal{T}$. Here, F in (1.1) would be the nonlinear parameter-to-solution mapping with

$$\mathcal{D}(F) = \{a \in \mathcal{L}^\infty(\Omega) \mid \operatorname{ess\,inf} a > 0\}\,,$$

and the range of F would belong to some observation space, e.g., $\mathcal{L}^2(\Omega \times \mathcal{T})$.

Inverse problems are often ill-posed in the sense that even when a is uniquely determined by the right-hand side u of (1.1), the mapping $u \mapsto a$ lacks continuity. This is a severe numerical problem when the given data $\tilde{u}$ are noisy and

$$\|\tilde{u} - u\| \le \delta \tag{1.3}$$

in the norm topology of $\mathcal{Y}$. As a consequence there is need for regularization and several possibilities for regularizing (1.1) are treated in [6].

Like for well-posed problems Newton type methods are one important option for solving (1.1) and have been applied with success in various applications: cf., e.g., [7, 20] for the parameter identification problem (1.2), [18, 21] for a related problem arising in impedance tomography, and [9, 14] for inverse scattering problems.

* Fachbereich Mathematik, Universität Kaiserslautern, D–67653 Kaiserslautern, Germany. email: hanke@math.uni-karlsruhe.de

On the other side, only few rigorous theoretical treatments of Newton type methods for ill-posed problems can be found in the literature. Exceptions are the works by Bakushinskii [1, 2], Nashed and Chen [17], and Blaschke et al. [4, 3].

The present paper develops inexact Newton type methods as a natural setting for nonlinear inverse problems. The basic idea is the computation of a regularized approximation of the linearized problem by an *inner iteration*, namely by a conjugate gradient method. Iterative methods for the linearized equation are particularly interesting for parameter identification problems where it is usually much cheaper to apply the Fréchet derivative to a single argument (which requires the solution of a differential equation, cf. Kravaris and Seinfeld [13]), than to assemble the whole (usually dense) derivative matrix and invert it afterwards. A posteriori stopping criteria for the inner and the outer iteration are suggested that make the algorithm a regularizing method in the sense of [6, Def. 3.1].

It has to be mentioned that inexact or truncated Newton methods have some tradition for large-scale well-posed problems (cf., e.g., Nash and Sofer [16] and the references given there) but those works do not address nor apply to ill-posed problems.

2. The algorithm. Throughout it will be assumed that $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces; the same notation $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ is used for the norms and inner products in $\mathcal{X}$ and $\mathcal{Y}$, respectively. For a linear operator $T : \mathcal{X} \to \mathcal{Y}$, $T^\star : \mathcal{Y} \to \mathcal{X}$ denotes the adjoint operator.

Newton's method is based on the Taylor expansion of F. Assuming that $a^\dagger$ is a solution of the nonlinear problem (1.1) and a_n is some approximation of $a^\dagger$ then

$$F(a^\dagger) - F(a_n) = F'(a_n)(a^\dagger - a_n) + R(a^\dagger; a_n)\,, \tag{2.1}$$

where $R(a^\dagger; a_n)$ is the Taylor remainder. Adding the noise term $\tilde{u} - u$ to (2.1), and solving for $a^\dagger - a_n$ this yields

$$F'(a_n)(a^\dagger - a_n) = \tilde{u} - F(a_n) \;+\; u - \tilde{u} - R(a^\dagger; a_n)\,. \tag{2.2}$$

The right-hand side of (2.2) splits in two parts: the first part, $\tilde{y}_n := \tilde{u} - F(a_n)$, is computable whereas the second part is not. In other words: the ideal update $x := a^\dagger - a_n$ solves the linear equation

$$T_n x = y_n \tag{2.3}$$

with $T_n = F'(a_n)$ and right-hand side y_n as in (2.2), however, only $\tilde{y}_n$ is known with

$$\|\tilde{y}_n - y_n\| \le \delta + \|R(a^\dagger; a_n)\|\,. \tag{2.4}$$

Here, δ is the bound on the data error (1.3) which is assumed to be available.

Although, in general, solving (2.3) for x is still ill-posed, there is a well-developed theory on how to regularize *linear* ill-posed problems with inexact data, cf., e.g., [8, 15, 6]. Several methods for approximating the solution of (2.3) have been suggested; CGNE, the conjugate gradient method applied to the normal equation

$$T_n^\star T_n x = T_n^\star \tilde{y}_n$$

belongs to the most powerful ones. The regularizing effect of CGNE comes from early termination of the iteration. Nemirovskii and Plato (cf. [11, Sect. 3.3]) have established

$n = 0$
while $\|\tilde{u} - F(a_n)\| > \tau\delta$ **do**
  begin % outer iteration
  $y = \tilde{u} - F(a_n)$, $T = F'(a_n)$
  $x_0 = 0$
  $r_0 = y$
  $w_0 = y$
  $\sigma_0 = 1$
  $k = 0$
  repeat % inner iteration
    $d_k = T^\star w_k$
    $\alpha_k = \|T^\star r_k\|^2 / \|T d_k\|^2$
    $x_{k+1} = x_k + \alpha_k d_k$
    $r_{k+1} = r_k - \alpha_k T d_k$
    $\beta_k = \|T^\star r_{k+1}\|^2 / \|T^\star r_k\|^2$
    $w_{k+1} = r_{k+1} + \beta_k w_k$
    $\sigma_{k+1} = 1 + \beta_k \sigma_k$
    $k = k + 1$
  until $\|r_k\| < \rho \|y\|$
  $a_{n+1} = a_n + x_k$
  $n = n + 1$
  end
return a_n.

ALGORITHM 2.1. *Truncated Newton*-CGNE

that the so-called *discrepancy principle* is a suitable stopping criterion for this purpose. This means that $x_0 = 0$ is in some sense the best possible approximation of the solution x of (2.3) if the error (2.4) dominates the right-hand side $\tilde{y}_n$, while otherwise CGNE should be stopped with iterate x_k as soon as the data fit $\|\tilde{y}_n - T_n x_k\|$ has the order of the error (2.4) in the right-hand side. This leads to the following two conclusions concerning a combination of inner and outer iteration for the nonlinear problem:

- If $\|\tilde{u} - F(a_n)\|$ has reached the order of δ then there is no sense in iterating any further.
- Even when this is not the case the outer iteration can only make any further progress via (2.3) if $\|R(a^\dagger; a_n)\| \ll \|\tilde{y}_n\|$. Otherwise the linearized equation provides little additional information.

In order to guarantee such an inequality – at least for a_n sufficiently close to $a^\dagger$ – the following assumption on the Taylor remainder term will turn out useful if not necessary: for a certain ball $\mathcal{B} \subset \mathcal{D}(F)$ around the exact solution $a^\dagger$ of (1.1), and some $C > 0$ let

$$\|F(\tilde{a}) - F(a) - F'(a)(\tilde{a} - a)\| \le C \|\tilde{a} - a\| \, \|F(\tilde{a}) - F(a)\| \tag{2.5}$$

for all $a, \tilde{a} \in \mathcal{B}$. It must be mentioned that an inequality like (2.5) is a nontrivial restriction in ill-posed problems; cf. the discussion in [12], where such an assumption has been employed for a convergence analysis of the nonlinear Landweber iteration. On the other hand, (2.5) is fulfilled for example for the inverse problem (1.2) with steady state or transient measurements of u in $\mathcal{L}^2(\Omega)$ provided that the exact solution $a^\dagger$ is sufficiently smooth, cf. [6, Ex. 11.1] and [10].

Consider the truncated Newton-CGNE scheme of Algorithm 2.1, where for the ease of notation y and T stand for $\tilde{y}_n$ and T_n, respectively. Algorithm 2.1 requires an input guess a_0 of $a^\dagger$ and two tolerance parameters ρ and τ for the stopping rules of the inner

and the outer iteration. Although any $\rho < 1$ and $\tau \geq 1$ would make sense for this purpose the theoretical results in Sect. 5 require $\rho^2\tau > 2$.

The inner iteration (CGNE) differs slightly from [11, Algorithm 2.3] in that it maintains an additional variable w_k connected to d_k via $d_k = T^\star w_k$; d_k is the same as in [11]. w_k and the additional scalar σ_k are required for the analysis in Sect. 3. In view of the theoretical results in [11] the discrepancy principle is the most natural stopping rule for CGNE; it requires an explicit upper bound for $\|\tilde{y}_n - y_n\|$. (2.4) and (2.5) yield a bound which is not implementable, namely

$$\|\tilde{y}_n - y_n\| \leq \delta + C\,\|a^\dagger - a_n\|\,\|u - F(a_n)\|\,.$$

However, during the iteration it can be presumed that $\delta \ll \|u - F(a_n)\|$, and hence any sufficiently large fraction of $\|\tilde{u} - F(a_n)\|$ may serve as an upper bound for the right-hand side when a_n is sufficiently close to $a^\dagger$. The CGNE iteration is therefore terminated as soon as

$$\|\tilde{y}_n - T_n x_k\| < \rho\,\|\tilde{y}_n\|\,, \tag{2.6}$$

where $0 < \rho < 1$ should be a fixed, but not too small parameter. According to the stopping rule (2.6), Algorithm 2.1 belongs to the general class of *inexact Newton methods* investigated in detail by Dembo, Eisenstat and Steihaug [5] for well-posed optimization problems.

The essential ingredient for the convergence analysis of Algorithm 2.1 is a monotonicity result for CGNE concerning the iteration error. The discrepancy principle *per se* is not an appropriate stopping rule for this purpose (cf. Example 3.3 in the following section), but fortunately monotonicity can nevertheless be established in the particular case (2.6) used above. From this follows that a_n converges to a solution of (1.1) as $n \to \infty$, provided the data u are given exactly and a_0 and $a^\dagger$ are sufficiently close (cf. Theorem 4.2).

For a theoretical analysis of the perturbed data case the inner iteration needs to be modified by an additional backtracking step in order to enforce equality in (2.6), cf. Sect. 5. While this does not affect the convergence analysis of Sect. 4, the advantage is that a_n depends continuously on $\tilde{u}$ through this modification. Thus, it follows from a general argument that the truncated Newton-CGNE method is a regularization method when $\rho^2\tau > 2$.

3. Preliminaries about CGNE. The convergence analysis for Algorithm 2.1 requires a few properties of CGNE, which go beyond the general theory developed in [11], and which may be of independent interest.

Given a linear operator $T : \mathcal{X} \to \mathcal{Y}$ and a right-hand side $y \in \mathcal{Y}$ the kth iterate x_k of CGNE (with initial guess $x_0 = 0$) belongs to the kth Krylov subspace

$$\mathcal{K}_k(T^\star y; T^\star T) = \operatorname{span}\{T^\star y, (T^\star T)T^\star y, \ldots, (T^\star T)^{k-1}T^\star y\}\,,$$

and among all elements $x \in \mathcal{K}_k(T^\star y; T^\star T)$, x_k minimizes the residual $\|y - Tx_k\|$. If P denotes the orthogonal projector onto $\mathcal{R}(T)$ then this minimizing element is unique as long as Py does not belong to an invariant subspace of $TT^\star$ of dimension $k-1$; for the ease of presentation this will tacitly be presumed throughout the sequel and, as will be shown later on, this is no restriction for the new results to be presented below.

As in [11] it is convenient to use the connection to the so-called *residual polynomials*. Let Π_k be the set of all polynomials of degree k or less, and set

$$\Pi_k^0 := \{p \in \Pi_k \mid p(0) = 1\}\,.$$

Then there is a 1-1 relation between elements $x \in \mathcal{K}_k(T^\star y; T^\star T)$ and $p \in \Pi_k^0$ via the representation

$$y - Tx = p(TT^\star)y \tag{3.1}$$

of the corresponding residual; see [11, Sect. 2.1] for details. In particular, $p_k \in \Pi_k^0$ shall denote the residual polynomial associated with x_k, the kth CGNE iterate.

The bilinear form

$$[\varphi, \psi] := \langle\, \varphi(TT^\star)y, \psi(TT^\star)y \,\rangle \tag{3.2}$$

defines an inner product for $\varphi, \psi \in \Pi_k$, and with this inner product and (3.1) the minimizing property of x_k can be reformulated as follows: p_k solves the minimization property

$$[p, p] \longrightarrow \min \qquad \text{among } p \in \Pi_k^0\,. \tag{3.3}$$

If $q \in \Pi_{k-1}$ is an arbitrary polynomial of degree $k-1$ then the polynomial p given by $p(\lambda) = p_k(\lambda) + t\lambda q(\lambda)$ belongs to Π_k^0 for every $t \in \mathbb{R}$, and hence, by virtue of (3.3),

$$[p_k, \lambda q] = \frac{1}{2}\frac{d}{dt}[p, p]\Big|_{t=0} = 0 \qquad \text{for all } q \in \Pi_{k-1}\,. \tag{3.4}$$

In particular, defining $q = q_{k-1}$ by $p_k = 1 - \lambda q_{k-1}$, it follows from (3.4) that

$$[p_k, 1] = [p_k, 1] - [p_k, \lambda q_{k-1}] = [p_k, p_k]\,, \tag{3.5}$$

an identity which will be useful later on.

Polynomials can also be used to rewrite the update (cf. Algorithm 2.1)

$$x_{k+1} - x_k = \alpha_k d_k = \alpha_k T^\star w_k\,. \tag{3.6}$$

In fact, it is easy to see that

$$w_k = s_k(TT^\star)y \qquad \text{with} \quad s_k(\lambda) := \frac{p_k(\lambda) - p_{k+1}(\lambda)}{\alpha_k \lambda} \in \Pi_k\,. \tag{3.7}$$

However, s_k will in general not belong to Π_k^0. Instead, since the vetors w_k are updated by $w_{k+1} = r_{k+1} + \beta_k w_k$ with $r_{k+1} = y - Tx_{k+1}$, it follows from (3.7) that

$$s_{k+1}(\lambda) = p_{k+1}(\lambda) + \beta_k s_k(\lambda)\,,$$

and hence, $s_k(0)$ and σ_k of Algorithm 2.1 enjoy the same recurrence relation, i.e.,

$$s_k(0) = \sigma_k\,. \tag{3.8}$$

It is an immediate consequence of the minimization property of the CGNE iterates that $\|y - Tx_k\|$ is monotonically decreasing for $k = 0, 1, \ldots$. It is also known (cf.,

e.g., [11, Sect. 3.1]) that for $y = Tx$ the actual *error* $\|x - x_k\|$ is decreasing, too. The following result considers the iteration error for perturbed right-hand sides.

THEOREM 3.1. *Let $\gamma \geq 2$, $k_* \in \mathbb{N}$, and $x \in \mathcal{X}$ satisfy $\|y - Tx\| \leq \varepsilon$. If*

$$\|y - Tx_k\|^2 + \|y - Tx_{k+1}\|^2 \;>\; \gamma \frac{\|w_k\|}{\sigma_k} \varepsilon\,, \qquad k = 0, 1, \ldots, k_* - 1\,, \tag{3.9}$$

then $\|x - x_k\|$ is strictly monotonically decreasing for $k = 0, 1, \ldots, k_*$, and

$$\|x\|^2 - \|x - x_{k_*}\|^2 \;>\; (\gamma - 2)\,\varepsilon \sum_{k=0}^{k_*-1} \alpha_k \|w_k\|\,.$$

Proof. By virtue of (3.6),

$$\begin{aligned}
\|x - x_{k+1}\|^2 &= \|x - x_k - \alpha_k T^\star w_k\|^2 \\
&= \|x - x_k\|^2 - \alpha_k \langle\, 2x - 2x_k - \alpha_k T^\star w_k, T^\star w_k \,\rangle \\
&= \|x - x_k\|^2 - \alpha_k \langle\, Tx - Tx_k, w_k \,\rangle - \alpha_k \langle\, Tx - Tx_{k+1}, w_k \,\rangle \\
&= \|x - x_k\|^2 - \alpha_k \langle\, y - Tx_k, w_k \,\rangle - \alpha_k \langle\, y - Tx_{k+1}, w_k \,\rangle + \\
&\qquad\qquad 2\alpha_k \langle\, y - Tx, w_k \,\rangle\,.
\end{aligned}$$

Inserting the definitions (3.1) and (3.7) of the corresponding polynomials this becomes

$$\|x - x_k\|^2 - \|x - x_{k+1}\|^2 \;=\; \alpha_k [p_k, s_k] + \alpha_k [p_{k+1}, s_k] - 2\alpha_k \langle\, y - Tx, w_k \,\rangle\,,$$

with $[\cdot, \cdot]$ as in (3.2). By (3.8), $s_k(\lambda) = \sigma_k + \lambda q(\lambda)$ for some polynomial $q \in \Pi_{k-1}$, and hence, it follows from (3.4) and (3.5) that

$$\begin{aligned}
\|x - x_k\|^2 - \|x - x_{k+1}\|^2 &= \alpha_k \sigma_k \Big([p_k, 1] + [p_{k+1}, 1] \Big) - 2\alpha_k \langle\, y - Tx, w_k \,\rangle \\
&= \alpha_k \sigma_k \Big([p_k, p_k] + [p_{k+1}, p_{k+1}] \Big) - 2\alpha_k \langle\, y - Tx, w_k \,\rangle\,.
\end{aligned}$$

Consequently, since $[p_k, p_k] = \|y - Tx_k\|^2$, the given assumptions yield

$$\|x - x_k\|^2 - \|x - x_{k+1}\|^2 \;>\; \alpha_k \sigma_k \gamma \frac{\|w_k\|}{\sigma_k} \varepsilon \;-\; 2\alpha_k \varepsilon \|w_k\| \tag{3.10}$$

for all $k = 0, \ldots, k_* - 1$. Since $\gamma \geq 2$, the right-hand side is nonnegative which shows that the sequence $\{\|x - x_k\|\}$ is strictly decreasing for k in the given range. Furthermore, since $x_0 = 0$ the second assertion follows by taking the sum of (3.10) from $k = 0$ to $k_* - 1$. □

Corresponding to the assumptions of Theorem 3.1 denote by $k(\varepsilon)$ the stopping index of the discrepancy principle as the smallest index $k = k(\varepsilon)$ for which

$$\|y - Tx_k\| \leq \frac{\gamma}{2} \varepsilon\,.$$

It is easy to see that k_* in Theorem 3.1 can never be larger than $k(\varepsilon)$: in fact, since $\check{s}_k := s_k / \sigma_k \in \Pi_k^0$ it follows from the minimization property of CGNE that

$$\|y - Tx_{k+1}\| \leq \|y - Tx_k\| = [p_k, p_k]^{1/2} \leq [\check{s}_k, \check{s}_k]^{1/2} = \frac{1}{\sigma_k} [s_k, s_k]^{1/2} = \frac{1}{\sigma_k} \|w_k\|\,,$$

and hence, by (3.9),

$$2\|y - Tx_{k_*-1}\|^2 \geq \|y - Tx_{k_*-1}\|^2 + \|y - Tx_{k_*}\|^2 > \gamma \frac{\|w_{k_*-1}\|}{\sigma_{k_*-1}} \varepsilon \geq \gamma \|y - Tx_{k_*-1}\| \varepsilon \,.$$

This shows that $\|y - Tx_{k_*-1}\| > \frac{\gamma}{2}\varepsilon$, which in turn implies that $k_* \leq k(\varepsilon)$.

Furthermore, since $k(\varepsilon)$ is never larger than what has been called in [11] the *ultimate termination index* $k = \kappa$ of CGNE, the projected right-hand side Py cannot belong to an invariant subspace of dimension $k_* - 1$ of $TT^\star$ if k_* is as in Theorem 3.1.

LEMMA 3.2. *Under the assumptions of Theorem* 3.1 *the inequality*

$$\|y - Tx_k\|^2 \geq \gamma\varepsilon\|y\| \tag{3.11}$$

implies (3.9) *provided that* $k > 0$*; furthermore, there are only finitely many* k *for which* (3.11) *can hold.*

Proof. With the same notation as before,

$$\frac{1}{\sigma_k} \|w_k\| = [\check{s}_k, \check{s}_k]^{1/2} \,,$$

and $\check{s}_k$ is the polynomial which has been denoted $p_k^{[2]}$ in [11]. Consequently, $[\check{s}_k, \check{s}_k]$ is strictly monotonically decreasing with k by [11, Theorem 3.2]; in particular,

$$\frac{1}{\sigma_k} \|w_k\| < \frac{1}{\sigma_0} \|w_0\| = \|y\| \,,$$

so that (3.11) implies (3.9). Since the stopping index of the discrepancy principle is a well-defined finite number (cf. [11, Sect. 3.3]) the remark following Theorem 3.1 implies that (3.11) can only hold for finitely many indices k. □

The following example shows that Theorem 3.1 is sharp in the sense that γ cannot be replaced by any number smaller than two, and that the assertion would not hold for

$$\|y - Tx_k\|^2 > c\varepsilon^2 \tag{3.12}$$

instead of (3.9), whatever the value of c might be.

EXAMPLE 3.3. Let $T : \mathcal{X} \to \mathcal{Y}$ be a compact linear operator with singular system $\{u_n, v_n, \mu_n \mid n \geq 0\}$, i.e.,

$$Tu_n = \mu_n v_n \,, \quad T^\star v_n = \mu_n u_n \,, \quad \mu_n > 0 \,, \qquad n = 0, 1, \ldots,$$

and $\{u_n\}$ and $\{v_n\}$ are orthonormal bases of $\mathcal{X}$ and $\mathcal{Y}$, respectively. Assume without loss of generality that $\mu_0 = 1$; recall that $\mu_n \to 0$ as $n \to \infty$.

For a given $n \in \mathbb{N}$ let $y := \mu_n^3 v_0 + v_n$, so that the CGNE iterates are

$$x_0 = 0 \,, \qquad x_1 = \alpha(\mu_n^3 u_0 + \mu_n u_n) \,, \qquad x_2 = \mu_n^3 u_0 + \frac{1}{\mu_n} u_n \,,$$

with

$$\alpha = \mu_n^{-2} \frac{1 + \mu_n^4}{1 + \mu_n^2} \,.$$

Consequently, for $x := \mu_n^3 u_0 + (\mu_n^{-1} - \mu_n)u_n$ the "iteration error" $\|x - x_k\|$ behaves for large n like

$$\|x - x_0\| \sim \mu_n^{-1}, \qquad \|x - x_1\| = \mu_n - 2\mu_n^3 + O(\mu_n^5), \qquad \|x - x_2\| = \mu_n .$$

Therefore, if n is sufficiently large, $\|x - x_k\|$ is decreasing in the first iteration, and increasing in the second one.

Consider now the quantities $\|y - Tx_1\|$ and $\|w_1\|/\sigma_1$. Straightforward computations show that

$$y - Tx_1 = y - \alpha TT^\star y, \qquad w_1/\sigma_1 = y - \frac{\alpha}{1+\beta} TT^\star y,$$

with $\frac{\alpha}{1+\beta} = \frac{1+\mu_n^2}{2\mu_n^2}$, and hence,

$$\|y - Tx_1\|^2 = \mu_n^2 \frac{(1-\mu_n^2)^2}{1+\mu_n^2}, \qquad \frac{\|w_1\|}{\sigma_1} = \frac{1}{2}(1-\mu_n^2)(1+\mu_n^2)^{1/2}.$$

Since the assumptions of Theorem 3.1 hold with $\varepsilon = \|y - Tx\| = \mu_n^2$, the right-hand side of (3.9) behaves like $\frac{\gamma}{2}\mu_n^2$ for $k = 1$ and $n \to \infty$, whereas the corresponding right-hand side of (3.12) behaves like $c\mu_n^4$. As a matter of fact, (3.9) with $\gamma \geq 2$ will not hold, but (3.12) will hold for $k = 1$ and n sufficiently large.

Another result that will be required in Sect. 5 is the following straightforward extension of the stability analysis in [11, Sect. 2.6].

LEMMA 3.4. *For $\delta > 0$ let $T_\delta : \mathcal{X} \to \mathcal{Y}$ be bounded linear operators and $y_\delta \in \mathcal{Y}$. Furthermore, denote by x_k^δ the kth* CGNE *iterate for $T_\delta x = y_\delta$. If $T_\delta \to T$ and $y_\delta \to y$ as $\delta \to 0$, and if Py does not belong to a $k-1$ dimensional invariant subspace of $TT^\star$, then x_k^δ is well-defined for all δ sufficiently small, and x_k^δ converges to the kth iterate of* CGNE *for $Tx = y$ as $\delta \to 0$.*

The proof of this lemma is exactly the same as the proof of Theorem 2.11 in [11] because the corresponding moments $\langle y_\delta, T_\delta^m y_\delta \rangle$, $0 \leq m \leq 2k-1$, still converge to $\langle y, T^m y \rangle$ as $\delta \to 0$, and this is all that is required for the proof to go through.

4. Convergence analysis for exact data. After these preliminaries reconsider the nonlinear operator F whose Taylor remainder satisfies (2.5), i.e.,

$$\|F(\tilde{a}) - F(a) - F'(a)(\tilde{a} - a)\| \leq C\|\tilde{a} - a\|\,\|F(\tilde{a}) - F(a)\| \tag{4.1}$$

for some $C > 0$ and all $a, \tilde{a}$ in a certain ball $\mathcal{B} \subset \mathcal{D}(F)$. It will be assumed throughout this section that $F(a) = u$ has a solution $a^\dagger \in \mathcal{B}$, and that the right-hand side $u \in \mathcal{Y}$ is given exactly. The following lemma applies the monotonicity result from the previous section to the nonlinear context.

LEMMA 4.1. *Consider the $(n+1)$st outer iteration of Algorithm 2.1. Let $\gamma > 2$, $0 < \rho < 1$, and assume that (4.1) holds for some $C > 0$. If $a_n \in \mathcal{B}$ with $\|a^\dagger - a_n\| \leq \rho^2/(\gamma C)$ then the inner iteration terminates after $k_n < \infty$ steps, and*

$$a_{n+1} = a_n + x_{k_n} = a_n + F'(a_n)^\star v_n$$

with a certain $v_n \in \mathcal{Y}$. Moreover, the following inequalities hold:

$$\|u - F(a_n)\|\,\|v_n\| < \frac{\gamma}{\gamma - 2}\frac{1}{\rho^2}\left(\|a^\dagger - a_n\|^2 - \|a^\dagger - a_{n+1}\|^2\right), \tag{4.2}$$

$$\|u - F(a_n)\|^2 < \frac{\gamma}{\gamma - 2}\frac{\|F'(a_n)\|^2}{\rho^2}\left(\|a^\dagger - a_n\|^2 - \|a^\dagger - a_{n+1}\|^2\right). \tag{4.3}$$

Proof. According to Sect. 2, $x = a^\dagger - a_n$ is a solution of (2.3), and $\tilde{y}_n = u - F(a_n)$ satisfies (2.4) with $\delta = 0$. By (4.1) and the closeness assumption of the lemma,

$$\|\tilde{y}_n - y_n\| \le C\,\|a^\dagger - a_n\|\,\|u - F(a_n)\| \le \frac{\rho^2}{\gamma}\,\|u - F(a_n)\|\,,$$

and hence, x satisfies the requirements of Theorem 3.1 with $\varepsilon = \frac{\rho^2}{\gamma}\|u - F(a_n)\|$. Substituting $\gamma\varepsilon$ in (3.11) accordingly, it follows from Lemma 3.2 that the stopping rule (2.6) determines a finite stopping index k_n for the inner iteration, and that (3.9) is fulfilled with $k_* = k_n$. (Note that for $k = 0$ (3.9) does always hold.) In other words, Theorem 3.1 applies to the inner iteration with $k_* = k_n$.

Consider the updates of a_n and x_k in Algorithm 2.1. It follows that (writing $F'(a_n)$ for T again)

$$a_{n+1} = a_n + x_{k_n} = a_n + \sum_{k=0}^{k_n-1} \alpha_k d_k = a_n + F'(a_n)^\star v_n \quad \text{with } v_n = \sum_{k=0}^{k_n-1} \alpha_k w_k\,.$$

Since

$$x = a^\dagger - a_n \qquad \text{and} \qquad x - x_{k_n} = a^\dagger - a_{n+1}\,,$$

Theorem 3.1 asserts that $\|a^\dagger - a_{n+1}\| < \|a^\dagger - a_n\|$, and that

$$\|a^\dagger - a_n\|^2 - \|a^\dagger - a_{n+1}\|^2 > \frac{\gamma - 2}{\gamma}\rho^2\,\|u - F(a_n)\| \sum_{k=0}^{k_n-1} \alpha_k \|w_k\|\,. \tag{4.4}$$

Since α_k is always nonnegative, the right-hand side of (4.4) can be estimated from below by $\frac{\gamma-2}{\gamma}\rho^2\|u - F(a_n)\|\,\|v_n\|$ which yields (4.2). The right-hand side of (4.4) can alternatively be estimated from below by $\frac{\gamma-2}{\gamma}\rho^2\alpha_0\|u - F(a_n)\|\,\|w_0\|$. Since $w_0 = \tilde{y}_n = u - F(a_n)$ according to Algorithm 2.1, and since

$$\alpha_0 = \|F'(a_n)^\star \tilde{y}_n\|^2 / \|F'(a_n)F'(a_n)^\star \tilde{y}_n\|^2 \ge \|F'(a_n)\|^{-2}\,, \tag{4.5}$$

this yields

$$\|a^\dagger - a_n\|^2 - \|a^\dagger - a_{n+1}\|^2 > \frac{\gamma - 2}{\gamma}\,\frac{\rho^2}{\|F'(a_n)\|^2}\,\|u - F(a_n)\|^2\,, \tag{4.6}$$

as was to be shown. □

Lemma 4.1 states that the inner iteration is a well-defined terminating loop, provided that $a^\dagger - a_n$ is sufficiently small. It is easy to see that the same inequalities (4.2) and (4.3) would hold if the inner iteration is terminated *before* the stopping criterion (2.6) is met. This is important for practical purposes because usually the number of inner iterations is constrained by some maximum number $k_{\max}$.

THEOREM 4.2. *Assume that $\tilde{u} = u = F(a^\dagger)$ for some $a^\dagger \in \mathcal{D}(F)$, and that F satisfies* (4.1) *for some $C > 0$ in a ball $\mathcal{B} \subset \mathcal{D}(F)$ around $a^\dagger$. Let $0 < \rho < 1$. If $a_0 \in \mathcal{B}$ and $\|a^\dagger - a_0\| < \rho^2/(2C)$ then the iterates $\{a_n\}$ of Algorithm 2.1 converge to a solution of* (1.1) *as $n \to \infty$.*

Proof. Lemma 4.1 can be applied with $\gamma = \rho^2/(C\,\|a^\dagger - a_0\|)$. This shows that $\|a^\dagger - a_n\|$ is monotonically decreasing, and therefore stopping rule (2.6) returns a well-defined stopping index k_n for each inner iteration.

It will be shown next that the iteration errors $e_n = a^\dagger - a_n$, $n \in \mathbb{N}$, form a Cauchy sequence. Given $m, n \in \mathbb{N}$ with $m > n$ let $l \in \{n, \ldots, m\}$ be chosen in such a way that

$$\|u - F(a_l)\| \le \|u - F(a_i)\|\,, \qquad i = n, \ldots, m\,. \tag{4.7}$$

Consider now

$$\|e_l - e_n\|^2 = 2\langle\, e_l - e_n, e_l \,\rangle + \|e_n\|^2 - \|e_l\|^2\,. \tag{4.8}$$

Inserting the definition of v_n from Lemma 4.1 it follows that

$$|\langle\, e_l - e_n, e_l \,\rangle| \;=\; \Big|\sum_{i=n}^{l-1} \langle\, F'(a_i)^* v_i, e_l \,\rangle\Big| \;\le\; \sum_{i=n}^{l-1} \|v_i\|\,\|F'(a_i)e_l\|\,,$$

where the last factor can be estimated by using (4.1) and the definition (4.7) of l:

$$\begin{aligned}
\|F'(a_i)e_l\| \;&=\; \|F'(a_i)e_i \;-\; F'(a_i)(a_l - a_i)\| \\
&\le\; \|u - F(a_i) - F'(a_i)e_i\| + \|F(a_l) - F(a_i) - F'(a_i)(a_l - a_i)\| + \|u - F(a_l)\| \\
&\le\; C\,\|a^\dagger - a_i\|\,\|u - F(a_i)\| + C\,\|a_l - a_i\|\,\|F(a_l) - F(a_i)\| + \|u - F(a_l)\| \\
&\le\; \frac{3}{2}\,\|u - F(a_i)\| \;+\; 2\,\|u - F(a_l)\| \\
&\le\; \frac{7}{2}\,\|u - F(a_i)\|\,.
\end{aligned}$$

Therefore, (4.2) implies that

$$|\langle\, e_l - e_n, e_l \,\rangle| \le \frac{\gamma}{\gamma - 2}\,\frac{7}{2\rho^2}\left(\|a^\dagger - a_n\|^2 - \|a^\dagger - a_l\|^2\right),$$

which, together with (4.8) yields

$$\|e_l - e_n\|^2 \le c\left(\|a^\dagger - a_n\|^2 - \|a^\dagger - a_l\|^2\right),$$

where $c = \frac{\gamma}{\gamma - 2}\frac{7}{\rho^2} + 1$ does not depend on l, n, or m. In the same way one obtains

$$\|e_m - e_l\|^2 \le c\Big(\|a^\dagger - a_l\|^2 - \|a^\dagger - a_m\|^2\Big)\,,$$

so that

$$\begin{aligned}
\|a_m - a_n\|^2 \;&=\; \|e_m - e_n\|^2 \;\le\; 2\|e_m - e_l\|^2 + 2\|e_l - e_n\|^2 \\
&\le\; 2c\Big(\|a^\dagger - a_n\|^2 - \|a^\dagger - a_m\|^2\Big)\,.
\end{aligned}$$

The right-hand side tends to zero for $n, m \to \infty$ because of the monotonicity of the iteration error, and hence, $\{a_n\}$ is a Cauchy sequence.

Denote the limit of a_n by a. Since $\|F'(a_n)\|$ remains uniformly bounded it follows from (4.3) by summation that $\sum_{n=0}^{\infty} \|u - F(a_n)\|^2$ converges, and therefore $F(a_n) \to u$ as $n \to \infty$. Thus, it has been shown that a is a solution of (1.1), and the proof is complete. □

Note that the theorem makes no assertion as to *which* solution $\{a_n\}$ does converge.

The proof of Theorem 4.2 uses a technique from [12] which has been developed for the convergence analysis of the nonlinear Landweber iteration. In [12] a somewhat weaker assumption on F has been employed, namely

$$\|F(\tilde{a}) - F(a) - F'(a)(\tilde{a} - a)\| \le \eta \|F(\tilde{a}) - F(a)\|, \qquad \eta < 1/2. \tag{4.9}$$

It is easy to see that the proof of Theorem 4.2 remains valid under the same assumption (4.9), provided that it holds in a ball $\mathcal{B}$ around $a^\dagger$; in this case, however, ρ must be a number in the interval $(\sqrt{2\eta}, 1)$.

5. Regularizing properties for inexact data. So far, Algorithm 2.1 has been considered for exactly given right-hand side u only. In practice only an approximation $\tilde{u} = u^\delta$ will be known with

$$\|u^\delta - u\| \le \delta. \tag{5.1}$$

To emphasize this point the corresponding iterates will be denoted by a_n^δ further on. In case of perturbed data it is important to stop the outer iteration sufficiently early to prevent divergence. Algorithm 2.1 terminates the outer loop as soon as the residual norm is of the order of the noise level δ: more precisely, if τ is a fixed positive number then the stopping index $n(\delta)$ is the smallest iteration index $n \in \mathbf{N}_0$ for which

$$\|u^\delta - F(a_n^\delta)\| \le \tau\delta. \tag{5.2}$$

The following result shows that this stopping criterion actually does terminate the outer iteration for adequate values of τ.

PROPOSITION 5.1. *Let $0 < \rho < 1$ and $\tau > 2/\rho^2$. Furthermore, let a be a solution of (1.1) with F satisfying (4.1) for some $C > 0$ in a ball $\mathcal{B} \subset \mathcal{D}(F)$ around a. If $a_0^\delta \in \mathcal{B}$ is sufficiently close to a, i.e., $\|a - a_0^\delta\| < \frac{\rho^2\tau - 2}{2C(1+\tau)}$, then Algorithm 2.1 is well-defined and terminates after $n(\delta) < \infty$ outer iterations. Moreover, for $n = 0, 1, \ldots, n(\delta)$, $\|a - a_n^\delta\|$ is monotonically decreasing.*

Proof. Without loss of generality it will be assumed that $n(\delta) > 0$. The proof goes by induction on n. Assume that

$$\|a - a_n^\delta\| < \frac{\rho^2\tau - 2}{2C(1+\tau)} \tag{5.3}$$

for some $n < n(\delta)$. By assumption this is fulfilled for $n = 0$. It will be shown that the associated inner iteration does terminate, and that $\|a - a_{n+1}^\delta\| < \|a - a_n^\delta\|$. According to (2.3) and (2.4) the given right-hand side $\tilde{y}_n = u^\delta - F(a_n^\delta)$ is an approximation of the right-hand side y_n of (2.2) with

$$\|\tilde{y}_n - y_n\| \le \delta + \|R(a; a_n^\delta)\|.$$

This can be further estimated by using (4.1), namely

$$\begin{aligned} \|\tilde{y}_n - y_n\| &\le \delta + C\|a - a_n^\delta\| \, \|u - F(a_n^\delta)\| \\ &\le \left(1 + C\|a - a_n^\delta\|\right)\delta + C\|a - a_n^\delta\| \, \|u^\delta - F(a_n^\delta)\|. \end{aligned}$$

Since $\|u^\delta - F(a_n^\delta)\| > \tau\delta$ as $n < n(\delta)$ this yields

$$\|\tilde{y}_n - y_n\| \le \varepsilon := \frac{1 + (1+\tau)C\,\|a - a_n^\delta\|}{\tau}\,\|u^\delta - F(a_n^\delta)\|\,. \tag{5.4}$$

Defining

$$\gamma = \frac{\rho^2\tau}{1 + (1+\tau)C\,\|a - a_n^\delta\|}\,,$$

it follows that $\gamma\varepsilon = \rho^2\|u^\delta - F(a_n^\delta)\|$ so that (3.11) and (3.9) hold for all inner iterations up to the stopping index. Since $\gamma > 2$ by assumption (5.3), the stopping index k_n of (2.6) for the inner iteration is well-defined according to Lemma 3.2. Moreover, Theorem 3.1 shows that the iteration error is decreasing up to the stopping index, and hence,

$$\|a - a_{n+1}^\delta\| < \|a - a_n^\delta\|\,.$$

This implies that a_{n+1}^δ satisfies (5.3) again, which completes the induction step.

It remains to show that the outer iteration terminates. From Theorem 3.1 follows that

$$\|a - a_n^\delta\|^2 - \|a - a_{n+1}^\delta\|^2 > \frac{\gamma - 2}{\gamma}\,\frac{\rho^2}{\|F'(a_n^\delta)\|^2}\,\|u^\delta - F(a_n^\delta)\|^2\,,$$

compare (4.6), which means that for some $c > 0$, independent of n,

$$\|u^\delta - F(a_n^\delta)\|^2 \le c\Big(\|a - a_n^\delta\|^2 - \|a - a_{n+1}^\delta\|^2\Big)\,.$$

Consequently,

$$\sum_{n=0}^{\infty} \|u^\delta - F(a_n^\delta)\|^2 \le c\|a - a_0^\delta\|^2 < \infty\,,$$

showing that $F(a_n^\delta) \to u^\delta$ as $n \to \infty$ if the outer iteration *would not* terminate. However, this would also imply that (5.2) holds for some finite $n(\delta)$ which is a contradiction. It follows that the outer iteration does indeed terminate according to (5.2). □

Note that the constraint $\tau > 2/\rho^2 > 2$ is somewhat restrictive: in practice, one would like to choose τ close to 1 to enforce a good data-fit of the final reconstruction; on the other hand, the tolerance ρ for the inexact Newton step should be sufficiently small to benefit from the quadratic Newton approximation. Advice on how to choose ρ and τ is given in Section 6.

It is possible to extend Proposition 5.1 to the case that F satisfies (4.9) instead of (4.1). This, however, yields even stronger restrictions on possible combinations of ρ and τ, namely

$$1 > \rho^2 > 2\eta, \qquad \tau > (2 + 2\eta)/(\rho^2 - 2\eta)\,,$$

where η is the constant in (4.9). Note that the lower bound for τ is similar to the one in [12].

To establish regularizing properties of Algorithm 2.1, the inner iteration has to be slightly modified by a final linear backtracking step in order to satisfy (2.6) with equality:

MODIFIED INNER ITERATION. *Let $\{x_k \mid k = 0, \ldots, k_n\}$ be the iterates of the inner iteration with residuals $\{r_k\}$ as in Algorithm 2.1 within the $(n+1)$st outer iteration, so that $\|r_{k_n}\| < \rho\|\tilde{y}_n\| \le \|r_{k_n-1}\|$. Then $\|\tilde{y}_n - T_n(x_{k_n-1} + \lambda\alpha_{k_n-1}T_n^\star w_{k_n-1})\| = \rho\|\tilde{y}_n\|$ for*

$$\lambda = \frac{\|r_{k_n-1}\|^2 - \rho^2\|\tilde{y}_n\|^2}{\|r_{k_n-1}\|^2 - \|r_{k_n}\|^2 + \left(\rho^2\|\tilde{y}_n\|^2 - \|r_{k_n}\|^2\right)^{1/2}\left(\|r_{k_n-1}\|^2 - \|r_{k_n}\|^2\right)^{1/2}},$$

and the computation of a_{n+1}^δ in Algorithm 2.1 is modified in the following way: define

$$\hat{\alpha}_{k_n-1} = \lambda\alpha_{k_n-1}, \qquad \hat{x}_{k_n} = x_{k_n-1} + \hat{\alpha}_{k_n-1}T^\star w_{k_n-1}, \tag{5.5}$$

and let

$$a_{n+1}^\delta = a_n^\delta + \hat{x}_{k_n}. \tag{5.6}$$

Note that $0 \le \lambda < 1$ so that $\hat{x}_{k_n}$ lies on the linear line segment between x_{k_n-1} and x_{k_n}, and the step from x_{k_n-1} to $\hat{x}_{k_n}$ can be interpreted as a damped CGNE step. It follows that all previous results for Algorithm 2.1 remain valid for this modification, after replacing α_{k_n-1} by $\hat{\alpha}_{k_n-1}$ everywhere. When $k_n = 1$ then $\|r_{k_n-1}\| = \|r_0\| = \|\tilde{y}_n\|$ and it follows that in this case λ is bounded from below by $1-\rho$. Consequently, when $k_n = 1$ then $\hat{\alpha}_0 = \lambda\alpha_0 \ge (1-\rho)\alpha_0$, and hence, $\|F'(a_n)\|^2$ in (4.5), (4.6), and (4.3) must be replaced by $\frac{1}{1-\rho}\|F'(a_n)\|^2$ for the modified algorithm.

When the inner iteration is terminated after a maximum number of $k_{\max}$ iterations *without* matching the stopping criterion (2.6) then, of course, no backtracking step is performed.

To formulate the following results consider a set of approximations $\{u^\delta\}$ corresponding to different noise levels $0 < \delta < \delta_0$. Throughout, let $n(\delta)$ be the stopping index of the outer iteration corresponding to the right-hand side u^δ. As before, a_n and a_n^δ denote the outer iterates of Algorithm 2.1 for the right-hand sides u and u^δ, respectively.

LEMMA 5.2. *With the above modification (5.5), (5.6), Algorithm 2.1 is stable in the following sense: if $n \le n(\delta)$ for all δ sufficiently small, then $a_n^\delta \to a_n$ as $\delta \to 0$.*

Proof. The proof goes by induction on n, where nothing is to prove for $n = 0$. Assume that $a_n^\delta \to a_n$ as $\delta \to 0$, and that $n+1 \le n(\delta)$ for all δ sufficiently small. Denote by $T_\delta = F'(a_n^\delta)$ and $y_\delta = u^\delta - F(a_n^\delta)$ the linear operator and right-hand side for the inner iteration of Algorithm 2.1 with data u^δ. In the same way let $T = F'(a_n)$ and $y = u - F(a_n)$ correspond to the inner iteration with exact right-hand side u. The modified updates in (5.5) are denoted by $\hat{x}^\delta$ and $\hat{x}$, respectively.

For the unperturbed right-hand side the inner iteration terminates after k_n steps, say, and according to the remarks following Theorem 3.1, Py does not belong to an invariant subspace of dimension $k_n - 1$ of $TT^\star$. Now, by assumption,

$$T_\delta \to T, \quad y_\delta \to y, \qquad \delta \to 0,$$

and hence, Lemma 3.4 applies to the present setting. It follows that

$$x_{k_n}^\delta \to x_{k_n}\,, \qquad x_{k_n-1}^\delta \to x_{k_n-1}\,, \qquad x_{k_n-2}^\delta \to x_{k_n-2} \quad (\text{when } k_n > 1)\,,$$

as $\delta \to 0$. On the other hand, the right-hand side of (2.6) converges to $\rho\|y\|$ as $\delta \to 0$, so that the inner iteration with perturbed right-hand side necessarily terminates either after k_n or $k_n - 1$ iterations for δ sufficiently small. (The latter can only occur if (2.6) holds with equality for $k = k_n - 1$ in the unperturbed data case, and in particular, only for $k_n - 1 \geq 1$.) According to (5.5), the final output $\hat{x}^\delta$ of the inner iteration is the unique element on the interpolating linear spline through $x_{k_n}^\delta$, $x_{k_n-1}^\delta$ and $x_{k_n-2}^\delta$ (when $k_n > 1$), for which

$$\|y_\delta - T_\delta \hat{x}^\delta\| = \rho\,\|y_\delta\|\,.$$

Like the edges of the spline, this element depends continuously on δ, and it therefore follows that $\hat{x}^\delta$ converges to $\hat{x}$ as $\delta \to 0$. This shows that $a_{n+1}^\delta \to a_{n+1}$ as $\delta \to 0$. □

It becomes clear from the proof of this lemma why Algorithm 2.1 had to be modified. Without modification it could happen (although only in very exceptional situations) that for some k and n

$$\|y_n - T_n x_k\| = \rho\,\|y_n\|\,, \tag{5.7}$$

in which case the inner iteration will terminate with $k_n = k + 1$. Given arbitrarily close perturbations u^δ of u, however, it cannot be predicted whether the perturbed inner iteration will terminate after k_n or $k_n - 1$ steps. When the latter is the case for some sequence $\delta \to 0$, a_{n+1}^δ will not converge to a_{n+1} in general.

The above results enable the application of a technique from [12] which states that *convergence* for unperturbed data and *monotonicity* and *stability* for perturbed data yield a regularization method.

THEOREM 5.3. *Fix $0 < \rho < 1$ and $\tau > 2/\rho^2$, and let u^δ and $a_{n(\delta)}^\delta$ be defined as before. If F satisfies (4.1) in some ball $\mathcal{B} \subset \mathcal{D}(F)$, and if $a_0 = a_0^\delta$ is sufficiently close to a solution of (1.1) in this ball then the iterates $a_{n(\delta)}^\delta$ of Algorithm 2.1 with the modification (5.5), (5.6), converge to a solution of (1.1) as $\delta \to 0$.*

Proof. Denote by $a^\dagger$ the limit of the iterates a_n of the modified Algorithm 2.1; $a^\dagger$ is a solution of (1.1) by Theorem 4.2. Assume first that $n(\delta_m) = n$ is constant for some subsequence $\delta_m \to 0$ (as $m \to \infty$) and corresponding right-hand sides u^{δ_m}. By Lemma 5.2, $a_{n(\delta_m)}^{\delta_m} \to a_n$ and hence, $F(a_{n(\delta_m)}^{\delta_m}) \to F(a_n)$ as $m \to \infty$. Taking the limit $\delta_m \to 0$ in (5.2) it follows that a_n is a solution of (1.1), and therefore $a_n = a^\dagger$.

For the remainder, it suffices to consider subsequences $\{n(\delta_m)\}_m$ which are monotonically increasing to infinity as $m \to \infty$ and $\delta_m \to 0$. In other words, $n(\delta_m) > n(\delta_l)$ for $m > l$, and therefore it follows from Proposition 5.1 that

$$\|a_{n(\delta_m)}^{\delta_m} - a^\dagger\| \leq \|a_{n(\delta_l)}^{\delta_m} - a^\dagger\| \leq \|a_{n(\delta_l)}^{\delta_m} - a_{n(\delta_l)}\| + \|a_{n(\delta_l)} - a^\dagger\|\,.$$

Given $\varepsilon > 0$ the last term on the right-hand side becomes smaller than $\varepsilon/2$ for *some* l sufficiently large by Theorem 4.2. For this fixed value of l, on the other hand, the other term on the right-hand side becomes smaller than $\varepsilon/2$ for *all* m sufficiently large because of the stability of the modified algorithm, cf. Lemma 5.2. This shows that

$$\|a_{n(\delta_m)}^{\delta_m} - a^\dagger\| < \varepsilon$$

for all m sufficiently large, and hence, $a^{\delta_m}_{n(\delta_m)} \to a^\dagger$ as $m \to \infty$. □

The same result would be true for the *original* Algorithm 2.1 without any further modification if it were known that (5.7) never occurs throughout the iteration with exact data u.

6. Concluding remarks. Instead of CGNE other Krylov subspace methods could be used for the inner iteration. For example, similar properties can be established for the Landweber method as inner iteration. In this case, monotonicity with stopping criterion (2.6) follows from a result of Defrise and de Mol (cf. [6, Proposition 6.3]). Of course, CGNE should outperform the Landweber iteration.

The restrictions on ρ and τ, i.e., $\rho^2\tau > 2$ are very conservative. Possible combinations of ρ and τ include, for example, $\rho = 0.9$ and $\tau = 2.5$, or $\rho = 0.8$ and $\tau = 3.2$, but smaller values of ρ and τ may work very well in applications. In fact, Theorem 3.1 only states that for *certain* x and y with $\|y - Tx\| \le \varepsilon$ the iteration error may increase in the $(k_* + 1)$st iteration. In most circumstances the iteration error will still decrease during subsequent iterations, so that the inner iteration could be continued beyond the termination index k_*. Another reason for this conservative estimate is the fact that the first factor of the upper bound ε in (5.4) is a severe overestimation in early stages of the iteration when $\delta \ll \|\tilde{u} - F(a_n^\delta)\|$.

To allow more inner iterations it is also possible to switch to the stopping criterion suggested by Theorem 3.1: terminate the inner iteration with x_k when

$$\|\tilde{y}_n - T_n x_k\|^2 + \|\tilde{y}_n - T_n x_{k+1}\|^2 \le \rho^2 \|\tilde{y}_n\| \frac{\|w_k\|}{\sigma_k} . \tag{6.1}$$

Here, ρ is the same parameter as before. It follows from Lemma 3.2 that the corresponding stopping index is never smaller than the one determined by (2.6).

In practical applications the inner iteration will also be terminated when a maximum number of $k_{\max}$ iterations has been made. As mentioned before, all theoretical results in this paper allow such an additional constraint. It should be mentioned that $k_{\max} = 1$ leads to the *method of steepest descent* which has been analyzed by Scherzer [19].

A final remark on condition (2.5) may be appropriate. It is an important aspect of the present work that all results can be formulated under very general conditions on the nonlinearity of F. Of course, (2.5) is nevertheless restrictive; still, conditions of this type seem necessary to deduce local convergence for *every* element u from the range of F. For comparison, with weaker assumptions on F the papers [1, 2, 17] did only establish convergence for *certain* right-hand sides u of (1.1). Blaschke et al. [4, 3] deduce convergence of their Newton type schemes for *all* possible elements in the range, however, only under more restrictive assumptions on F than (2.5). For example, for the aforementioned problem of identifying the coefficient a of (1.2) inequality (2.5) is fulfilled but the assumptions from [4, 3] are not.

Algorithm 2.1 is currently tested on an inverse problem in electrical impedance tomography. The corresponding numerical results shall be published elsewhere.

REFERENCES

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

[2] ——, *Nonsaturating iteration methods for solving degenerate nonlinear operator equations*, Dokl. Math., 52 (1995), pp. 159–160.

[3] B. BLASCHKE, *Some Newton type methods for the regularization of nonlinear ill-posed problems*, in preparation.

[4] B. BLASCHKE, A. NEUBAUER, AND O. SCHERZER, *On convergence rates for the iteratively regularized Gauss-Newton method*, submitted.

[5] R. S. DEMBO, S. C. EISENSTAT, AND T. STEIHAUG, *Inexact Newton methods*, SIAM J. Numer. Anal., 19 (1982), pp. 400–408.

[6] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer Academic Publ., Dordrecht, 1996.

[7] G. ERIKSSON AND G. DAHLQUIST, *On an inverse non-linear diffusion problem*, in Numerical Treatment of Inverse Problems in Differential and Integral Equations, P. Deuflhard and E. Hairer, eds., Basel, 1983, Birkhäuser, pp. 238–245.

[8] C. W. GROETSCH, *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Pitman, Boston, 1984.

[9] S. GUTMAN AND M. KLIBANOV, *Iterative methods for multi-dimensional inverse scattering problems at fixed frequency*, Inverse Problems, 10 (1994), pp. 573–599.

[10] M. HANKE, *A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems*, submitted.

[11] ——, *Conjugate Gradient Type Methods for Ill-Posed Problems*, Longman Scientific & Technical, Harlow, Essex, 1995.

[12] 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.

[13] C. KRAVARIS AND J. H. SEINFELD, *Identification of parameters in distributed parameter systems by regularization*, SIAM J. Control Optim., 23 (1985), pp. 217–241.

[14] R. KRESS AND W. RUNDELL, *A quasi-Newton method in inverse obstacle scattering*, Inverse Problems, 10 (1994), pp. 1145–1157.

[15] A. K. LOUIS, *Inverse und schlecht gestellte Probleme*, Teubner, Stuttgart, 1989.

[16] S. G. NASH AND A. SOFER, *Linear and Nonlinear Programming*, McGraw-Hill, New York, 1996.

[17] M. Z. NASHED AND X. CHEN, *Convergence of Newton-like methods for singular operator equations using outer inverses*, Numer. Math., 66 (1993), pp. 235–257.

[18] K. PAULSON, W. LIONHEART, AND M. PIDCOCK, POMPUS*: an optimized* EIT *reconstruction algorithm*, Inverse Problems, 11 (1995), pp. 425–437.

[19] O. SCHERZER, *A convergence analysis of a method of steepest descent and a two-step algorithm for nonlinear ill-posed problems*, Numer. Funct. Anal. Optim., to appear.

[20] C. R. VOGEL AND G. WADE, *A modified Levenberg-Marquardt algorithm for large scale inverse problems*, in Computation and Control III. Progress in Systems and Control, K. Bowers and J. Lund, eds., Basel, 1993, Birkhäuser, pp. 367–378.

[21] T. J. YORKEY, *Electrical impedance tomography with piecewise polynomial conductivities*, J. Comp. Phys., 91 (1990), pp. 344–360.