UNIVERSITÄT KARLSRUHE

Towards A General Convergence Theory For Inexact Newton Regularizations

A. Lechleiter
A. Rieder

Preprint Nr. 09/01

Institut für Wissenschaftliches Rechnen
und Mathematische Modellbildung

76128 Karlsruhe

Anschriften der Verfasser:

Dr. Armin Lechleiter
Centre de Mathématiques Appliqués (CMAP)
Ecole Polytechnique
91128 Palaiseau Cedex, France

Prof. Dr. Andreas Rieder
Institut für Angewandte und Numerische Mathematik und
Institut für Wissenschaftliches Rechnen und Mathematische Modellbildung
Universität Karlsruhe (TH)
D-76128 Karlsruhe

TOWARDS A GENERAL CONVERGENCE THEORY FOR INEXACT NEWTON REGULARIZATIONS

ARMIN LECHLEITER[†] AND ANDREAS RIEDER[‡]

February 4, 2009

Abstract. We develop a general convergence analysis for a class of inexact Newton-type regularizations for stably solving nonlinear ill-posed problems. Each of the methods under consideration consists of two components: the outer Newton iteration and an inner regularization scheme which, applied to the linearized system, provides the update. In this paper we give a novel and unified convergence analysis which is not confined to a specific inner regularization scheme but applies to a multitude of schemes including Landweber and steepest decent iterations, iterated Tikhonov method, and method of conjugate gradients.

Key words. Nonlinear ill-posed problems, inexact Newton iteration.

AMS subject classifications. 65J20, 65J22.

1. Introduction. During the last two decades a broad variety of Newton-like methods for regularizing nonlinear ill-posed problems have been suggested and analyzed, see, e.g., [1, 10, 15] for an overview and original references. So similar some of the methods are so different are their analyses, even when the same structural assumptions on the nonlinearity are required (for a recent exception, however, see [9]).

This situation is in contrast to the linear setting. Here, a general theory is known when the (linear) regularization scheme is generated by a regularizing filter function, see, e.g., [5, 11, 13, 15]. Properties of the scheme can be directly read off from properties of the generating filter function.

The present paper was driven by the wish to develop a similar general theory for a class of regularization schemes of inexact Newton-type for nonlinear ill-posed problems. This class has been introduced and named `REGINN` (`REG`ularization based on `IN`exact `N`ewton iteration) by the second author [14]. Each of the `REGINN`-methods consists of two components, the outer Newton iteration and the inner scheme providing the increment by regularizing the local linearization. Although the methods differ in their inner regularization schemes we are able to present a common convergence analysis. To this end we compile four features which not only guarantee convergence but are also shared by various inner regularization schemes which

[†]Centre de Mathématiques Appliqués (CMAP), Ecole Polytechnique, 91128 Palaiseau Cedex, France, `alechle@cmapx.polytechnique.fr`

[‡]Institut für Angewandte und Numerische Mathematik and Institut für Wissenschaftliches Rechnen und Mathematische Modellbildung, Universität Karlsruhe, 76128 Karlsruhe, Germany, `andreas.rieder@math.uni-karlsruhe.de`

are so different as, for instance, Landweber iteration, steepest decent iteration, implicit iteration, and method of conjugate gradients.

Let us now set the stage for `REGINN`. We like to solve the nonlinear ill-posed problem

$$F(x) = y^\delta \tag{1.1}$$

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

$$\|y - y^\delta\|_Y \le \delta. \tag{1.2}$$

The nonnegative *noise level* δ is assumed to be known. Algorithm `REGINN` for solving (1.1) is a Newton-type algorithm which updates the actual iterate x_n by adding a correction step s_n^{N} obtained from solving a linearization of (1.1):

$$x_{n+1} = x_n + s_n^{\mathrm{N}}, \quad n \in \mathbb{N}_0, \tag{1.3}$$

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

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

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

$$F'(x_n)s_n^{\mathrm{e}} = y - F(x_n) - E(x^+, x_n) =: b_n \tag{1.4}$$

where

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

is the linearization error. In the sequel we will use the notation

$$A_n = F'(x_n).$$

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

$$b_n^\varepsilon := y^\delta - F(x_n) \quad \text{with } \|b_n - b_n^\varepsilon\|_Y \le \delta + \|E(x^+, x_n)\|_Y. \tag{1.5}$$

Therefore, we determine the correction step s_n^{N} as a stable approximate solution of

$$A_n s = b_n^\varepsilon \tag{1.6}$$

by applying a regularization scheme, for instance, Landweber iteration, Showalter method, (iterated) Tikhonov regularization, method of conjugate gradients, etc. Therefore, let $\{s_{n,m}\}_m \subset X$ be the sequence of regularized approximations generated by a chosen regularization scheme applied to (1.6).

```
REGINN(x_{N(δ)}, R, {μ_n})
n := 0;  x_0 := x_{N(δ)};
while ‖F(x_n) − y^δ‖_Y > Rδ do
{   m := 0;
    repeat
       m := m + 1;
       compute s_{n,m} from (1.6);
    until ‖F'(x_n)s_{n,m} + F(x_n) − y^δ‖_Y ≤ μ_n‖F(x_n) − y^δ‖_Y
    x_{n+1} := x_n + s_{n,m};
    n := n + 1;
}
x_{N(δ)} := x_n;
```

Figure 1.1: REGINN: REGularization based on INexact Newton iteration.

We now explain how we pick the Newton step s_n^{N} out of $\{s_{n,m}\}$: For an adequately chosen tolerance $\mu_n \in]0,1[$ (see Lemma 2.3 below) define

$$m_n = \min\big\{m \in \mathbb{N} \,:\, \|A_n s_{n,m} - b_n^\varepsilon\|_Y \le \mu_n \|b_n^\varepsilon\|_Y\big\},$$

and set

$$s_n^{\mathrm{N}} := s_{n,m_n}.$$

In other words: the Newton step is the first element of $\{s_{n,m}\}$ for which the residual $\|A_n s_{n,m} - b_n^\varepsilon\|_Y$ is less or equal to $\mu_n\|b_n^\varepsilon\|_Y$.

Finally, we stop the Newton iteration (1.3) by a discrepancy principle: choose $R > 0$ and accept the iterate $x_{N(\delta)}$ as approximation to x^+ if

$$\|y^\delta - F(x_{N(\delta)})\|_Y \le R\delta < \|y^\delta - F(x_n)\|_Y, \quad n = 0, 1, \ldots, N(\delta) - 1, \tag{1.7}$$

see Figure 1.1.

The remainder of the paper is organized as follows. In the next section we present a residual and level set based analysis of REGINN requiring only three rather elementary properties of the regularizing sequence $\{s_{n,m}\}$ together with a structural restriction on the nonlinearity F. In a certain sense, this restriction is equivalent to the meanwhile well-established tangential cone condition, see, e.g., [10, 15]. Under our assumptions REGINN is well defined and terminates. Moreover, all iterates stay in the level set $\mathcal{L} = \{x \in \mathsf{D}(F) \colon \|F(x) - y^\delta\|_Y \le \|F(x_0) - y^\delta\|_Y\}$. Unfortunately, $\mathcal{L}$ cannot be assumed bounded, thus prohibiting the use of a weak-compactness argument to verify convergence.

Local convergence, however, is our topic in Section 3. Provided the regularizing sequence $\{s_{n,m}\}$ exhibits monotonic error decrease (up to a stopping index) all `REGINN`-iterates will stay in a ball about x^+. Finally, we apply the weak-compactness argument which guarantees convergence.

Several regularization methods applied to (1.6) generate sequences $\{s_{n,m}\}$ meeting our assumptions. Some of the respective proofs, which do not fit comfortably in the body of the text, are given in two appendices.

2. Residual and level set based analysis. For the analysis of `REGINN` we require three properties of the regularizing sequence $\{s_{n,m}\}$, namely

$$\langle A_n s_{n,m}, b_n^\varepsilon \rangle_Y > 0 \quad \forall m, \tag{2.1}$$

and

$$\lim_{m\to\infty} A_n s_{n,m} = P_{\overline{\mathsf{R}(A_n)}} b_n^\varepsilon. \tag{2.2}$$

The latter convergence guarantees existence of a number $\vartheta_n \geq 1$ such that $\|A_n s_{n,m}\|_Y \leq \vartheta_n \|b_n^\varepsilon\|_Y$ for all m. We, however, require also uniformity in n: There is a $\Theta \geq 1$ with

$$\|A_n s_{n,m}\|_Y \leq \Theta \|b_n^\varepsilon\|_Y \quad \forall m, n. \tag{2.3}$$

Typically, $\{s_{n,m}\}$ is generated by

$$s_{n,m} = g_m(A_n^* A_n) A_n^* b_n^\varepsilon$$

where $g_m \colon [0, \|A_n\|^2] \to \mathbb{R}$ is a so-called filter function. If

$$0 < \lambda g_m(\lambda) \leq C_g, \; \lambda > 0, \quad \text{and} \quad \lim_{m\to\infty} g_m(\lambda) = 1/\lambda, \; \lambda > 0,$$

and $b_n^\varepsilon \neq 0$ then all requirements (2.1), (2.2), and (2.3) are fulfilled where $\Theta \leq C_g$. Here are some concrete examples:

- Landweber iteration: $g_m(\lambda) = \lambda^{-1}\big(1 - (1 - \omega\lambda)^m\big)$ where $\omega \in]0, \|A_n\|^{-2}[$ and $C_g = 1$.
- Tikhonov regularization: $g_m(\lambda) = 1/(\lambda + \alpha_m)$ where $\{\alpha_m\}_m$ is a positive sequence converging strongly monotone to zero. Thus, $C_g = 1$.
- Iterated Tikhonov regularization (implicit iteration): $g_m(\lambda) = \lambda^{-1}\big(1 - \prod_{k=1}^m (1 + \alpha_k\lambda)^{-1}\big)$ where the positive sequence $\{\alpha_k\}_k$ is bounded away from zero, typically $\{\alpha_k\} \subset [\alpha_{\min}, \alpha_{\max}]$ where $0 < \alpha_{\min} < \alpha_{\max}$. Here, $C_g = 1$.
- Showalter regularization:

$$g_m(\lambda) = \begin{cases} \lambda^{-1}\big(1 - \exp\big(-\alpha_m^{-1}\lambda\big)\big) & : \lambda > 0, \\ \alpha_m^{-1} & : \lambda = 0, \end{cases}$$

  where the positive sequence $\{\alpha_m\}_m$ converges strongly monotone to zero. Again, $C_g = 1$.

- Semi-iterative ν-methods ($\nu > 0$) due to Brakhage [2]: Let A_n be scaled, that is, $\|A_n\| \leq 1$. Then, $g_m(\lambda) = \big(1 - \widetilde{P}_m^{(\nu)}(\lambda)\big)/\lambda$ where $\widetilde{P}_m^{(\nu)}(\lambda) = P_m^{(2\nu-1/2,-1/2)}(1-2\lambda)/P_m^{(2\nu-1/2,-1/2)}(1)$ with $P_m^{(\alpha,\beta)}$ denoting the Jacobi-polynomials. As $\widetilde{P}_m^{(\nu)}$ attains negative values in $]0,1[$ (all roots are within this interval) we have $C_g > 1$. Sharp estimates for C_g or Θ are hard to obtain.

Also nonlinear regularizations schemes, which cannot be represented by filter functions, satisfy (2.1), (2.2), and (2.3):

- steepest decent method where $\Theta < 2^*$, and
- method of conjugate gradients (cg-method) where $\Theta = 1$,

provided the staring iterate is 0, see Appendix A for the respective proofs.

Remark 2.1. *Recently, Jin and Tautenhahn* [9] *presented a subtle convergence analysis of (generalized) iteratively regularized Gauß-Newton methods,*

$$x_{n+1} = x_n + g_{m_n}\big(A_n^* A_n\big)A_n^* b_n^\varepsilon + \big(I - g_{m_n}\big(A_n^* A_n\big)A_n^* A_n\big)(x_0 - x_n),$$

stopped by the discrepancy principle (1.7)*. Except for the rightmost term the striking difference to* `REGINN` *consists in the a priori choice of the sequence* $\{m_n\}_n$ *which is assumed to be monotonically increasing by a certain rate.*

For a large class of filter functions (including Landweber and Showalter filters) they proved deep and far-reaching convergence results. Under weaker assumptions, not covered by Theorems 1, 2 *or* 3 *in* [9]*, we obtain weaker convergence results. However, the technique of Jin and Tautenhahn does not apply to* `REGINN` [9, Remark 3] *and cannot be extended to other nonlinear regularization schemes in a straightforward way.*

Now we present first results. By the first property of (2.1) any direction $s_{n,m}$ is a descent direction for the functional $\varphi(\cdot) = \frac{1}{2}\|y^\delta - F(\cdot)\|_Y^2$.

Lemma 2.2. *We have that*

$$\langle \nabla\varphi(x_n), s_{n,m}\rangle_X < 0.$$

Proof. By $\nabla\varphi(\cdot) = -F'(\cdot)^*\big(y^\delta - F(\cdot)\big)$ we find that

$$\langle \nabla\varphi(x_n), s_{n,m}\rangle_X = -\langle b_n^\varepsilon, A_n s_{n,m}\rangle_Y \overset{(2.1)}{<} 0$$

*We strongly conjecture that $\Theta = 1$ for the steepest decent method, see Remark A.1 below.

and the lemma is verified. $\square$

If μ_n is not too small then the Newton step $s_n^{\mathrm{N}} = s_{n,m_n}$ is well defined indeed.

LEMMA 2.3. *Assume* (2.2) *and* $\|P_{\mathsf{R}(A_n)^\perp} b_n^\varepsilon\|_Y < \|b_n^\varepsilon\|_Y$. *Then, for any tolerance*

$$\mu_n \in \left] \frac{\|P_{\mathsf{R}(A_n)^\perp} b_n^\varepsilon\|_Y}{\|b_n^\varepsilon\|_Y}, 1 \right]$$

there is an $m^* \in \mathbb{N}$ *such that* $\|A_n s_{n,m} - b_n^\varepsilon\|_Y \le \mu_n \|b_n^\varepsilon\|_Y$ *for all* $m \ge m^*$.

Proof. By (2.2)

$$\lim_{m\to\infty} \frac{\|A_n s_{n,m} - b_n^\varepsilon\|_Y}{\|b_n^\varepsilon\|_Y} = \frac{\|P_{\mathsf{R}(A_n)^\perp} b_n^\varepsilon\|_Y}{\|b_n^\varepsilon\|_Y}.$$

which completes the proof. $\square$

REMARK 2.4. *If the assumption in above lemma is violated then* `REGINN` *fails (as well as other Newton schemes): under* $\|P_{\mathsf{R}(A_n)^\perp} b_n^\varepsilon\|_Y = \|b_n^\varepsilon\|_Y$ *we have that* $s_{n,m} = 0$ *for all* m.

Now we provide a framework that guarantees termination of `REGINN` (Figure 1.1), that is, we prove existence of $x_{N(\delta)}$.

For $x_0 \in \mathsf{D}(F)$ such that $\|F(x_0) - y^\delta\|_Y > \delta$ define the level set

$$\mathcal{L}(x_0) := \big\{ x \in \mathsf{D}(F) \colon \|F(x) - y^\delta\|_Y \le \|F(x_0) - y^\delta\| \big\}.$$

Note that $x^+ \in \mathcal{L}(x_0)$.

Further, we restrict the structure of nonlinearity. Throughout we work with the following bound for the linearization error:

$$\begin{gathered} \|E(v,w)\|_Y \le L\,\|F'(w)(v-w)\|_Y \quad \text{for one } L < 1 \\ \text{and for all } v, w \in \mathcal{L}(x_0) \text{ with } v - w \in \mathsf{N}\big(F'(w)\big)^\perp. \end{gathered} \tag{2.4}$$

From (2.4) we derive that

$$\|E(v,w)\|_Y \le \omega\,\|F(w) - F(v)\|_Y \quad \text{where } \omega = \frac{L}{1-L} > L \tag{2.5}$$

which is the *tangential cone condition* introduced by Scherzer [16]. In the convergence analysis of Newton methods for ill-posed problems, both (2.4) and (2.5) are adequate to replace the Lipschitz continuity of the Fréchet derivative which is typically used to bound the linearization error in the framework of well-posed problems, see, e.g., [10, Section 2.1] for a detailed explanation.

Remark 2.5. *Actually,* (2.4) *and* (2.5) *are equivalent in the following sense:* (2.5) *for one* $\omega < 1$ *implies* (2.4) *with* $L = \frac{\omega}{1-\omega}$.

Moreover, we assume the existence of a $\varrho \in [0,1[$ such that

$$\begin{aligned} \big\| P_{\mathsf{R}(F'(u))^\perp}\big(F(x^+) - F(u)\big)\big\|_Y \leq \varrho \|F(x^+) - F(u)\|_Y \\ \text{for all } u \in \mathcal{L}(x_0). \end{aligned} \tag{2.6}$$

Assumption (2.6) is quite natural as it characterizes those nonlinear problems which can be tackled by local linearization (compare Remark 2.4): As (2.6) is equivalent to

$$\sqrt{1-\varrho^2}\,\|F(x^+) - F(u)\|_Y \leq \big\| P_{\overline{\mathsf{R}(F'(u))}}\big(F(x^+) - F(u)\big)\big\|_Y,$$

the right hand side of the linearized system (1.6) has a component in the closure of the range of A_n and the magnitude of this component is uniformly bounded from below by $\sqrt{1-\varrho^2}$.

We give an example of a nonlinear operator where both (2.4) and (2.6) are satisfied globally in the domain of definition.

Example 2.6. *Let* $f\colon \mathbb{R} \to \mathbb{R}$ *be a continuously differentiable function with a derivative bounded from below:* $f'(t) \geq f'_{\min} > 0$. *We define the operator* $F\colon X \to Y$ *by*

$$F(x) = \sum_{n=1}^{\infty} \frac{1}{n} f\big(\langle x, v_n\rangle_X\big) u_n$$

where $\{v_n\}$ *and* $\{u_n\}$ *are orthonormal bases in the separable Hilbert spaces* X *and* Y, *respectively. The Fréchet derivative of* F *is the compact operator*

$$F'(x)h = \sum_{n=1}^{\infty} \frac{1}{n} f'\big(\langle x, v_n\rangle_X\big)\langle h, v_n\rangle_X u_n$$

with range $\mathsf{R}(F'(x)) = \big\{ y \in Y \,:\, \{n\langle y, u_n\rangle_Y\}_n \in \ell^2\big\}$. *Clearly,* $\overline{\mathsf{R}(F'(x))} = Y$. *Hence,* (2.6) *holds true with* $\varrho = 0$.

Now we further restrict the nonlinearity by imposing a bound from above on the derivative of f: $f'(t) \leq f'_{\max}$ *with* $f'_{\max} < 2f'_{\min}$. *For instance,* $f(t) = t + 0.25\arctan(t) + 1$ *or* $f(t) = 6t + \cos(t)$. *By the mean value theorem there is a* $\xi \in]s,t[$ *such that* $f(t) - f(s) = f'(\xi)(t-s)$. *Therefore, for all* $s, t \in \mathbb{R}$

$$\begin{aligned} |f(t) - f(s) - f'(s)(t-s)| &= \frac{|f'(\xi) - f'(s)|}{f'(s)} |f'(s)(t-s)| \\ &\leq \underbrace{\frac{f'_{\max} - f'_{\min}}{f'_{\min}}}_{=:\, L < 1} |f'(s)(t-s)| \end{aligned}$$

implying

$$\|E(v,w)\|_Y \le L\,\|F'(w)(v-w)\|_Y \quad \textit{for all } v,w \in X.$$

For L small enough, (2.4) implies (2.6).

LEMMA 2.7. *Assume* (2.4) *to hold with* $L < 1/2$. *Then,* (2.6) *holds for*

$$\varrho = \frac{L}{1-L} < 1.$$

Proof. We have that

$$\begin{aligned}
&\big\|P_{\mathsf{R}(F'(u))^\perp}\big(F(x^+)-F(u)\big)\big\|_Y \\
&\qquad\qquad = \big\|P_{\mathsf{R}(F'(u))^\perp}\big(F(x^+)-F(u)-F'(u)(x^+-u)\big)\big\|_Y \\
&\qquad\qquad \le L\,\|F'(u)(x^+-u)\|_Y.
\end{aligned}$$

Further,

$$\begin{aligned}
\|F'(u)(x^+-u)\|_Y &\le \|E(x^+,u)\|_Y + \|F(x^+)-F(u)\|_Y \\
&\le L\,\|F'(u)(x^+-u)\|_Y + \|F(x^+)-F(u)\|_Y
\end{aligned}$$

yielding first

$$\|F'(u)(x^+-u)\|_Y \le \frac{1}{1-L}\,\|F(x^+)-F(u)\|_Y$$

and then the assertion. $\square$

THEOREM 2.8. *Let* $\mathsf{D}(F)$ *be open and choose* $x_0 \in \mathsf{D}(F)$ *such that* $\overline{\mathcal{L}(x_0)} \subset \mathsf{D}(F)$. *Assume* (2.1), (2.2), (2.3), (2.4), *and* (2.6) *to hold true with* Θ, L, *and* ϱ *satisfying*

$$\Theta L + \varrho < \Lambda \quad \textit{for one } \Lambda < 1. \tag{2.7}$$

Further, choose

$$R > \frac{1+\varrho}{\Lambda - \Theta L - \varrho}. \tag{2.8}$$

Finally, select all tolerances $\{\mu_n\}$ *such that*

$$\mu_n \in \big]\mu_{\min,n},\, \Lambda - \Theta L\big], \quad \textit{with } \mu_{\min,n} := \frac{(1+\varrho)\delta}{\|b_n^\varepsilon\|_Y} + \varrho.$$

Then, there exists an $N(\delta)$ *such that all iterates* $\{x_1,\dots,x_{N(\delta)}\}$ *of* `REGINN` *are well defined and stay in* $\mathcal{L}(x_0)$. *Moreover, only final iterate satisfies the discrepancy principle, that is,*

$$\|y^\delta - F(x_{N(\delta)})\|_Y \le R\delta, \tag{2.9}$$

and the nonlinear residuals decrease linearly at an estimated rate

$$\frac{\|y^\delta - F(x_{n+1})\|_Y}{\|y^\delta - F(x_n)\|_Y} \le \mu_n + \theta_n L \le \Lambda, \quad n = 0, \ldots, N(\delta) - 1, \tag{2.10}$$

where $\theta_n = \|A_n s_n^{\mathrm{N}}\|_Y / \|b_n^\varepsilon\|_Y \le \Theta$.

Proof. Before we start with the proof let us discuss our assumptions on L, ϱ, Λ, and R. Condition (2.7) guarantees that the denominator of the lower bound on R is positive. The lower bound on R is needed to have a well-defined nonempty interval for selecting μ_n. Indeed, as long as $\|b_n^\varepsilon\|_Y > R\delta$ we get

$$\mu_{\min,n} = \frac{(1+\varrho)\delta}{\|b_n^\varepsilon\|_Y} + \varrho < \frac{1+\varrho}{R} + \varrho \overset{(2.8)}{<} \Lambda - \Theta L. \tag{2.11}$$

We will argue inductively and therefore assume the iterates $\{x_1, \ldots, x_n\}$ to be well defined in $\mathcal{L}(x_0)$. If $\|b_n^\varepsilon\|_Y \le R\delta$ then `REGINN` will terminate with $N(\delta) = n$. Otherwise, $\|b_n^\varepsilon\|_Y > R\delta$ and $\mu_n \in]\mu_{\min,n}, \Lambda - \Theta L]$ will provide Newton step s_n^{N}:

$$\begin{aligned} \frac{\|P_{\mathsf{R}(A_n)^\perp} b_n^\varepsilon\|_Y}{\|b_n^\varepsilon\|_Y} &\le \frac{\delta + \|P_{\mathsf{R}(A_n)^\perp}(F(x^+) - F(x_n))\|_Y}{\|b_n^\varepsilon\|_Y} \\ &\overset{(2.6)}{\le} \frac{\delta + \varrho \|F(x^+) - F(x_n)\|_Y}{\|b_n^\varepsilon\|_Y} \\ &\le \frac{(1+\varrho)\delta + \varrho\|b_n^\varepsilon\|_Y}{\|b_n^\varepsilon\|_Y} = \mu_{\min,n} < \mu_n. \end{aligned} \tag{2.12}$$

By Lemma 2.3 the Newton step s_n^{N} and hence $x_{n+1} = x_n + s_n^{\mathrm{N}} \in X$ are well defined.

We next show that x_{n+1} is in $\mathcal{L}(x_0)$. As s_n^{N} is a decent direction (Lemma 2.2) and $\mathsf{D}(F)$ is assumed to be open there exists a $\lambda > 0$ such that $x_{n,\lambda} := x_n + \lambda\, s_n^{\mathrm{N}}$ is in $\mathsf{D}(F)$ and

$$\|y^\delta - F(x_{n,\lambda})\|_Y < \|y^\delta - F(x_n)\|_Y \le \|y^\delta - F(x_0)\|_Y.$$

Thus, $x_{n,\lambda} \in \mathcal{L}(x_0)$. Further, $x_{n,\lambda} - x_n = \lambda\, s_n^{\mathrm{N}} \in \mathsf{R}(A_n^*) \subset \mathsf{N}(A_n)^\perp$. Accordingly we may proceed by estimating

$$\begin{aligned} \|y^\delta - F(x_{n,\lambda})\|_Y &= \|y^\delta - F(x_n) - \lambda A_n s_n^{\mathrm{N}} - (F(x_{n,\lambda}) - F(x_n) - \lambda A_n s_n^{\mathrm{N}})\|_Y \\ &\overset{(2.4)}{\le} \|y^\delta - F(x_n) - \lambda A_n s_n^{\mathrm{N}}\|_Y + L\,\lambda\,\|A_n s_n^{\mathrm{N}}\|_Y \\ &\le \|(1-\lambda) b_n^\varepsilon + \lambda(b_n^\varepsilon - A_n s_n^{\mathrm{N}})\|_Y + L\,\lambda\,\theta_n\,\|b_n^\varepsilon\|_Y \\ &\le (1-\lambda)\|b_n^\varepsilon\|_Y + \mu_n\,\lambda\,\|b_n^\varepsilon\|_Y + L\,\lambda\,\theta_n\,\|b_n^\varepsilon\|_Y \\ &\le \big(1 - \lambda(1-\Lambda)\big)\,\|b_n^\varepsilon\|_Y. \end{aligned} \tag{2.13}$$

Define

$$\lambda_{\max} := \sup\big\{\lambda \in [0,1] \,:\, x_{n,\lambda} \in \mathcal{L}(x_0)\big\}.$$

Assume $\lambda_{\max} < 1$, that is, $x_{n,\lambda_{\max}} \in \partial\mathcal{L}(x_0) \subset \mathsf{D}(F)$. By continuity we obtain from (2.13) that

$$\|y^\delta - F(x_{n,\lambda_{\max}})\|_Y \leq \big(1 - \lambda_{\max}(1-\Lambda)\big)\, \|b_n^\varepsilon\|_Y < \|b_n^\varepsilon\|_Y \leq \|b_0^\varepsilon\|_Y$$

contradicting $x_{n,\lambda_{\max}} \in \partial\mathcal{L}(x_0)$. Hence, $\lambda_{\max} = 1$ and $x_{n+1} = x_{n,\lambda_{\max}} \in \mathcal{L}(x_0)$. Finally, $\|b_{n+1}^\varepsilon\|_Y \leq (\mu_n + \theta_n L)\|b_n^\varepsilon\|_Y$ by plugging $\lambda = 1$ into (2.13). $\square$

A few comments are in order.

Remark 2.9. *Deuflhard, Engl, and Scherzer* [4, formula (2.11)] *have basically introduced the following Newton-Mysovskikh-like condition*

$$\big\|\big(F'(v) - F'(w)\big)F'(w)^+\big\| \leq L \quad \text{for all } v, w \in \mathcal{L}(x_0) \tag{2.14}$$

where $F'(w)^+$ denotes the Moore-Penrose inverse of $F'(w)$. They discovered interesting relations to other structural assumptions used in the convergence analysis of iterative methods for the solution of nonlinear ill-posed problems [4, Lemma 2.3].

If $\mathcal{L}(x_0)$ is convex then (2.14) *implies* (2.4)*. Indeed, for $v, w \in \mathcal{L}(x_0)$ with $v - w \in \mathsf{N}(F'(w))^\perp$ we have*

$$F'(w)^+F'(w)(v-w) = P_{\mathsf{N}(F'(w))^\perp}(v-w) = v - w$$

resulting in

$$\begin{aligned}
\|E(v,w)\|_Y &\leq \int_0^1 \big\|\big(F'(w + t(v-w)) - F'(w)\big)(v-w)\big\|_Y \,\mathrm{d}t \\
&= \int_0^1 \big\|\big(F'(w + t(v-w)) - F'(w)\big)F'(w)^+F'(w)(v-w)\big\|_Y \,\mathrm{d}t \\
&\leq L\,\big\|F'(w)(v-w)\big\|_Y.
\end{aligned}$$

Remark 2.10. *An assumption similar to* (2.6) *is*

$$\begin{gathered}
\big\|P_{\mathsf{R}(F'(u))^\perp}\big(\eta - F(u)\big)\big\|_Y \leq \widetilde{\varrho}\|\eta - F(u)\|_Y \quad \text{for one } \widetilde{\varrho} < 1 \\
\text{and for all } u \in \mathcal{L}(x_0) \text{ and all } \eta \in Y \text{ with } \|\eta - F(x^+)\|_Y \leq \delta_{\max}.
\end{gathered} \tag{2.15}$$

Under above property the hypotheses of Theorem 2.8 *can be relaxed: Let $\delta \leq \delta_{\max}$. Since*

$$\frac{\|P_{\mathsf{R}(A_n)^\perp} b_n^\varepsilon\|_Y}{\|b_n^\varepsilon\|_Y} \leq \widetilde{\varrho}$$

the assertion of Theorem 2.8 *remains true whenever $\widetilde{\varrho} + L < \Lambda$, $\{\mu_n\} \subset]\widetilde{\varrho}, \Lambda - L]$ and $R > 0$ (no other restriction on R, compare* (2.8)*).*

The mapping from Example 2.6 *satisfies* (2.15) *with* $\widetilde{\varrho} = 0$ *for any* $\delta_{\max} \geq 0$. *Nevertheless,* (2.15) *is quite restrictive. While* (2.6) *holds trivially for any linear mapping (with* $\varrho = 0$*),* (2.15) *can only hold for a linear mapping with a dense range. Indeed, let* $F\colon X \to Y$ *be a linear and bounded mapping with a non-closed range. Assume* (2.15) *as well as* $\overline{\mathsf{R}(F)} \neq Y$. *Let* $y^\delta \notin \overline{\mathsf{R}(F)}$ *(a natural assumption for noisy data). There is a sequence* $\{u_n\} \subset X$ *such that* $\lim_{n\to\infty} \|Fu_n - P_{\overline{\mathsf{R}(F)}} y^\delta\|_Y = 0$. *Since*

$$\begin{aligned}\lim_{n\to\infty} \|Fu_n - y^\delta\|_Y &= \|P_{\mathsf{R}(F)^\perp} y^\delta\|_Y = \|P_{\mathsf{R}(F)^\perp}(Fx^+ - y^\delta)\|_Y \leq \delta \\ &< \|Fx_0 - y^\delta\|_Y\end{aligned}$$

we may assume the whole sequence $\{u_n\}$ *is in* $\mathcal{L}(x_0)$. *Now,*

$$\frac{\big\|P_{\mathsf{R}(F)^\perp}(y^\delta - Fu_n)\big\|_Y}{\|y^\delta - Fu_n\|_Y} = \frac{\big\|P_{\mathsf{R}(F)^\perp} y^\delta\big\|_Y}{\|y^\delta - Fu_n\|_Y} \xrightarrow{n\to\infty} 1$$

contradicts (2.15).

3. Local convergence. After establishing termination of `REGINN` the next question to answer is: Does the family $\{x_{N(\delta)}\}_{0<\delta\leq\delta_{\max}}$ converge to a solution of $F(\cdot) = y$ as the noise level δ approaches 0 ?

Since

$$\|y - F(x_{N(\delta)})\|_Y \overset{(2.9)}{<} (R+1)\delta \tag{3.1}$$

the images of $\{x_{N(\delta)}\}$ under F converge to y. This, however, implies by no means convergence of $\{x_{N(\delta)}\}$. Indeed, $\{x_{N(\delta)}\}$ might explode as $\delta \to 0$. There is no reason to suppose compactness or boundedness of the level set $\mathcal{L}(x_0)$. Contrary, for an ill-posed problem $\mathcal{L}(x_0)$ is expected to be unbounded.

In this section we will show boundedness and then some kind of convergence of $\{x_{N(\delta)}\}$ provided the regularizing sequence $\{s_{n,m}\}$ exhibits a fourth property in addition to those from (2.1), (2.2), and (2.3). We require the following monotonicity:

$$\left.\begin{array}{l}\text{Let there be a continuous and monotonically increasing function } \Psi\colon \mathbb{R} \to \mathbb{R} \text{ with } t \leq \Psi(t) \text{ for } t \in [0,1] \text{ such that if} \\ \gamma_n = \|b_n^\varepsilon - A_n s_n^{\mathrm{e}}\|_Y / \|b_n^\varepsilon\|_Y < 1 \text{ and } \|b_n^\varepsilon - A_n s_{n,m-1}\|_Y \geq \Psi(\gamma_n)\|b_n^\varepsilon\|_Y \text{ then} \\ \qquad \|s_{n,m} - s_n^{\mathrm{e}}\|_X < \|s_{n,m-1} - s_n^{\mathrm{e}}\|_X.\end{array}\right\} \tag{3.2}$$

Examples of methods with monotonicity are

- Landweber iteration and steepest decent: $\Psi(t) = 2t$,
- Implicit iteration: $\Psi(t) = 2\,\dfrac{\alpha_{\max} + s}{\alpha_{\min}}\, t$ where $s = \sup_n \|A_n\|^2$ and $\{\alpha_k\}_k \subset [\alpha_{\min}, \alpha_{\max}]$,

- cg-method: $\Psi(t) = \sqrt{2t}$,

the respective proofs are given in Appendix B.

Under (3.2) we formulate a version of Theorem 2.8 where all assumptions are related to a ball about x^+, that is, the implicitely defined, generally unbounded level set $\mathcal{L}(x_0)$ is replaced by $B_r(x^+)$. Especially, (2.4) is replaced by

$$\begin{gathered}\|E(v,w)\|_Y \leq L\,\|F'(w)(v-w)\|_Y \quad \text{for one } L<1\\ \text{and for all } v,w \in B_r(x^+) \subset \mathsf{D}(F).\end{gathered} \tag{3.3}$$

THEOREM 3.1. *Assume* (2.1), (2.2), (2.3), (3.2). *Additionally, let* (2.6) *hold true in* $B_r(x^+)$ *and assume* (3.3) *with* L *satisfying*

$$\Psi\Big(\frac{L}{1-L}\Big) + \Theta L < \Lambda \quad \textit{for one } \Lambda < 1.^{\dagger} \tag{3.4}$$

Further, define

$$\mu_{\min} := \Psi\left(\Big(\frac{1}{R} + L\Big)\frac{1}{1-L}\right)$$

and choose R *so large that*

$$\mu_{\min} + \Theta L < \Lambda. \tag{3.5}$$

Restrict all tolerances $\{\mu_n\}$ *to* $\mu_n \in [\mu_{\min}, \Lambda - \Theta L]$ *and start with* $x_0 \in B_r(x^+)$.

Then, there exists an $N(\delta)$ *such that all iterates* $\{x_1,\dots,x_{N(\delta)}\}$ *of* `REGINN` *are well defined and stay in* $B_r(x^+)$. *We even have a strictly monotone error reduction:*

$$\|x^+ - x_n\|_X < \|x^+ - x_{n-1}\|_X, \quad n = 1,\dots,N(\delta). \tag{3.6}$$

Moreover, only the final iterate satisfies the discrepancy principle (2.9) *and the nonlinear residuals decrease linearly at the estimated rate* (2.10).

Proof. Let us first discuss our assumptions. If (3.4) applies then, by continuity of Ψ, there exists a R such that $\mu_{\min}$ satisfies (3.5) and the interval for selecting the tolerances is nonempty.

As before we use an inductive argument: Assume the iterates $x_1,\dots,x_n$ to be well defined in $B_\rho(x^+)$. If $\|b_n^\varepsilon\|_Y < R\delta$ `REGINN` will be stopped with $N(\delta) = n$. Otherwise, $\|b_n^\varepsilon\|_Y \geq R\delta$ and $\mu_n \in [\mu_{\min}, \Lambda - \Theta L]$ will provide a new Newton step. Indeed, in view of (2.12) and (2.11) we have that

$$\frac{\|P_{\mathsf{R}(A_n)^\perp} b_n^\varepsilon\|_Y}{\|b_n^\varepsilon\|_Y} < \frac{1+\varrho}{R} + \varrho \overset{(3.2)}{\leq} \Psi\Big(\frac{1+\varrho}{R} + \varrho\Big) \leq \mu_{\min}$$

[†] As $\frac{L}{1-L} + L \leq \frac{L}{1-L} + \Theta L \leq \Psi\big(\frac{L}{1-L}\big) + \Theta L < 1$ we have the necessary condition $L < (3-\sqrt{5})/2 \approx 0.38$.

where the latter estimate holds true due to $\varrho \leq L/(1-L)$ (Lemma 2.7) and the monotonicity of Ψ. By Lemma 2.3 the Newton step s_n^{N} and hence $x_{n+1} = x_n + s_n^{\mathrm{N}} \in X$ are well defined.

It remains to verify the strictly monotone error reduction (3.6). We will rely on (3.2). By (1.5) and (3.3), we have

$$\|b_n - b_n^\varepsilon\|_Y \leq \delta + L\|b_n\|_Y \leq \frac{1}{R}\|b_n^\varepsilon\|_Y + L(\|b_n - b_n^\varepsilon\|_Y + \|b_n^\varepsilon\|_Y)$$

yielding first

$$\gamma_n = \frac{\|b_n - b_n^\varepsilon\|_Y}{\|b_n^\varepsilon\|_Y} \leq \Big(\frac{1}{R} + L\Big)\frac{1}{1-L}$$

and then

$$\Psi(\gamma_n) \leq \mu_{\min} \leq \mu_n.$$

Accordingly, $\|b_n^\varepsilon - A_n s_{n,m-1}\|_Y \geq \mu_{\min}\|b_n^\varepsilon\|_Y$, $m = 1, \dots, m_n$, and we have by repeatedly applying the monotonicity (3.2)

$$\begin{aligned}\|x^+ - x_{n+1}\|_X &= \|s_n^{\mathrm{e}} - s_{n,m_n}\|_X \\ &< \|s_n^{\mathrm{e}} - s_{n,m_n-1}\|_X < \|s_n^{\mathrm{e}} - s_{n,m_n-2}\|_X \\ &< \cdots < \|s_n^{\mathrm{e}} - s_{n,0}\|_X = \|s_n^{\mathrm{e}}\|_X = \|x^+ - x_n\|_X\end{aligned}$$

which is (3.6). ◻

Remark 3.2. *Some nonlinear ill-posed problems (such as a model in electrical impedance tomography, see* [12]*) satisfy a slightly stronger version of* (3.3) *where L is replaced by $C\|v-w\|_X$. In view of* (3.6) *we expect in this situation the reduction rate* (2.10) *to approach μ_n as the Newton iteration progresses.*

Corollary 3.3. *Adopt all assumptions and notations of Theorem* 3.1. *Additionally let F be weakly sequentially closed and let $\{\delta_j\}_{j\in\mathbb{N}}$ be a positive zero sequence.*

Then, any subsequence of $\{x_{N(\delta_j)}\}_{j\in\mathbb{N}}$ contains a subsequence which converges weakly to a solution of $F(x) = y$.

Proof. Any subsequence of the bounded family $\{x_{N(\delta_j)}\}_{j\in\mathbb{N}} \subset B_r(x^+)$ is bounded and, therefore, has a weakly convergent subsequence. Let ξ be its weak limit. By (3.1) the images under F of this weakly convergent subsequence converge (weakly) to y. Due to the weak closedness of F we have that $y = F(\xi)$. ◻

The whole family $\{x_{N(\delta_j)}\}_{j\in\mathbb{N}}$ converges weakly to x^+ if x^+ is the unique solution of $F(x) = y$ in $B_r(x^+)$. This follows, for instance, from Proposition 10.13 (2) in [17]. However, under the assumptions of Theorem 3.1, the latter can only happen if $\mathsf{N}(A)$, the null space of $A = F'(x^+)$, is trivial. In fact, if $0 \neq v \in \mathsf{N}(A)$ then

$$\|F(x^+ + tv) - y\|_Y = \|F(x^+ + tv) - F(x^+)\|_Y \overset{(3.3)}{\leq} (L+1)\,|t|\,\|Av\|_Y = 0$$

for any $t \in [0, r/\|v\|_X]$.

On the other hand, if $\mathsf{N}(A)$ is trivial we even have a norm convergence.

Corollary 3.4. *Under the assumptions of Theorem* 3.1 *we have that*

$$\|x^+ - x_{N(\delta)}\|_A < \frac{1+R}{1-L}\,\delta$$

where $\|\cdot\|_A = \|A\cdot\|_Y$ *is a semi-norm in general.*

Proof. From (3.3) we obtain that

$$\|x^+ - x_{N(\delta)}\|_A \leq \frac{1}{1-L}\,\|y - F(x_{N(\delta)})\|_Y$$

which, in view of (3.1), implies the assertion. □

The above corollary yields norm convergence whenever $\mathsf{N}(A) = \{0\}$. In general, this norm is weaker than the standard norm in X.

We finish with two remarks.

Remark 3.5. *A stronger assumption than* (3.3) *is*

$$\begin{aligned} \|E(v,w)\|_Y \leq \widetilde{L}\,\|F'(w)(v-w)\|_Y^{1+\kappa} \quad &\text{for one } \kappa > 0 \\ \text{and for all } v, w \in B_r(x^+). \end{aligned} \tag{3.7}$$

Here, $\widetilde{L}$ *is allowed to be arbitrarily large. If* r *is sufficiently small we have* (3.3) *with*

$$L := 2^\kappa r^\kappa\,\widetilde{L} \max_{u \in B_r(x^+)} \|F'(u)\|^\kappa < 1.$$

Now, let r *be so small that all assumptions of Theorem* 3.1 *apply with* L *as above. Additionally, choose* $x_0 \in B_r(x^+)$ *satisfying* $\|y^\delta - F(x_0)\|_Y^\kappa \leq L/\widetilde{L}$.[‡] *Then, all assertions of Theorem* 3.1 *remain valid with the stronger rate*

$$\frac{\|y^\delta - F(x_{n+1})\|_Y}{\|y^\delta - F(x_n)\|_Y} \leq \mu_n + \theta_n^{1+\kappa}\,\Lambda^{\kappa n}\,L \leq \Lambda, \quad n = 0, \ldots, N(\delta) - 1. \tag{3.8}$$

We only need to verify the rate. We have

$$\begin{aligned} \big\|b_{n+1}^\varepsilon\big\|_Y &= \big\|b_n^\varepsilon - A_n s_n^{\mathrm{N}} + E(x_{n+1}, x_n)\big\|_Y \overset{(3.7)}{\leq} \mu_n \big\|b_n^\varepsilon\big\|_Y + \widetilde{L}\big\|A_n s_n^{\mathrm{N}}\big\|_Y^{1+\kappa} \\ &\leq \Big(\mu_n + \widetilde{L}\theta_n^{1+\kappa}\big\|b_n^\varepsilon\big\|_Y^\kappa\Big)\,\big\|b_n^\varepsilon\big\|_Y \end{aligned}$$

[‡] This bound implicitly forces $\|y - y^\delta\|_Y^\kappa < L/\widetilde{L}$.

which inductively implies (3.8).

REMARK 3.6. *Both bounds* (3.3) *and* (3.7) *for the linearization error may be derived from the following* affine contravariant *Lipschitz condition:*

$$\begin{gathered}\big\|\big(F'(v)-F'(w)\big)(v-w)\big\|_Y \le L_\kappa\, \|F'(w)(v-w)\|_Y^{1+\kappa}\\ \text{for one } \kappa\in[0,1] \text{ and for all } v,\, w\in B_r(x^+)\end{gathered} \tag{3.9}$$

where $L_\kappa > 0$ and in case $\kappa = 0$ we require $L_0 < 1$. Indeed,

$$\begin{aligned}\|E(v,w)\|_Y &= \Big\|\int_0^1 \big(F'\big(w+t(v-w)\big)-F'(w)\big)(v-w)\,\mathrm{d}t\Big\|_Y\\ &\le \frac{L_\kappa}{1+\kappa}\,\|F'(w)(v-w)\|_Y^{1+\kappa}.\end{aligned}$$

For a general discussion of the importance of affine contravariance for Newton-like algorithms we refer to Section 1.2.2 of Deuflhard's book [3]. *In particular, Section 4.2 of the same book treats Gauß-Newton methods for (well-posed) finite dimensional least squares problems under* (3.9) *globally in* $\mathsf{D}(F)$ *and with $\kappa = 1$.*

Appendix A. Proof of (2.1) **and** (2.2) **for cg and steepest decent.** Let $T \in \mathcal{L}(X, Y)$ and $0 \neq g \in Y$. The cg-method is an iteration for solving the normal equation $T^*Tf = T^*g$. Starting with $f_0 \in X$ the cg-method produces a sequence $\{f_m\}_{m\in\mathbb{N}_0}$ with the following minimization property

$$\|g - Tf_m\|_Y = \min\big\{\|g-Tf\|_Y \,\big|\, f\in X,\ f-f_0\in U_m\big\},\quad m\ge 1,$$

where U_m is the m-th Krylov space,

$$U_m := \operatorname{span}\big\{T^*r^0, (T^*T)T^*r^0, (T^*T)^2T^*r^0, \ldots, (T^*T)^{m-1}T^*r^0\big\} \subset \mathsf{N}(T)^\perp$$

with $r^0 := g - Tf_0$. Here, $\mathsf{N}(T)^\perp$ denotes the orthogonal complement of the null space $\mathsf{N}(T)$ of T. Since

$$\langle g - Tf_m, Tu\rangle_Y = 0 \quad \text{for all } u \in U_m, \tag{A.1}$$

see formula (5.19) in [15], we have that

$$\langle g - Tf_m, Tf_m\rangle_Y = 0 \quad \text{for all } m \in \mathbb{N}_0 \text{ provided } f_0 = 0. \tag{A.2}$$

Therefore,

$$0 \le \|g - Tf_m\|_Y^2 = \|g\|_Y^2 - \|Tf_m\|_Y^2$$

which is (2.3) with $\Theta = 1$. Further,

$$\langle g, Tf_m\rangle_Y \overset{(A.2)}{=} \|Tf_m\|_Y^2 > 0 \quad \text{as } f_m \in U_m \subset \mathsf{N}(T)^\perp,$$

that is, we have established (2.1). It is a well-known property of cg-iteration that

$$\lim_{m\to\infty} Tf_m = P_{\overline{\mathsf{R}(T)}}g$$

whenever $f_0 \in \mathsf{N}(T)^\perp$, see, e.g., page 135 ff. in [15]. Hence, (2.2) holds for cg-iteration.

Let us now consider steepest decent. Starting with $f_0 \in X$ steepest decent produces the sequence $\{f_m\}_{m\in\mathbb{N}_0}$ by

$$f_{m+1} = f_m + \lambda_m T^* r_m \quad \text{where } r_m = g - Tf_m \text{ and } \lambda_m = \frac{\|T^* r_m\|_X^2}{\|TT^* r_m\|_Y^2}.$$

We first validate monotonicity of the residuals:

$$\|r_{m+1}\|_Y < \|r_m\|_Y. \tag{A.3}$$

Define $f_{m+1}^L := f_m + \omega T^* r_m$ with $0 < \omega < 2/\|T\|^2$ and observe

$$g - Tf_{m+1}^L = (I - \omega TT^*) r_m.$$

Due to the optimality of the step size λ_m we have

$$\|r_{m+1}\|_Y \le \|g - Tf_{m+1}^L\|_Y = \|(I - \omega TT^*) r_m\|_Y < \|r_m\|_Y.$$

Whence (A.3) holds true.

Let $f_0 = 0$. Then,

$$\|Tf_m\|_Y^2 - 2\langle Tf_m, g\rangle_Y + \|g\|_Y^2 = \|r_m\|_Y^2 \overset{\text{(A.3)}}{<} \|r_0\|_Y^2 = \|g\|_Y^2$$

leading to

$$\|Tf_m\|_Y^2 < 2\langle Tf_m, g\rangle_Y$$

which yields (2.1) for $m \ge 1$. By Cauchy-Schwarz inequality we deduce that

$$\|Tf_m\|_Y < 2\|g\|_Y$$

which yields (2.3) with $\Theta < 2$.

REMARK A.1. *We strongly suspect that $\Theta = 1$. Indeed,*

$$\|Tf_1\|_Y = \frac{\|T^* g\|_X^2}{\|TT^* g\|_Y^2}\|TT^* g\|_Y = \frac{\|T^* g\|_X^2}{\|TT^* g\|_Y} = \frac{\langle TT^* g, g\rangle_Y}{\|TT^* g\|_Y} \le \|g\|_Y.$$

Further, from (A.3) *we obtain*

$$\|Tf_2\|_Y^2 < 2\langle T^* g, \lambda_1 T^* r_1\rangle_Y + \|Tf_1\|_Y^2 = \|Tf_1\|_Y^2.$$

Thus,

$$\|Tf_2\| < \|Tf_1\| \le \|g\|.$$

However, we are not able to give a complete proof of our conjecture.

As we do not know an adequate reference for the convergence

$$\lim_{m\to\infty} Tf_m = P_{\overline{\mathsf{R}(T)}}g$$

we give a short proof. First we replace $g \in Y$ by $P_{\overline{\mathsf{R}(T)}}g$ which does not change the steepest decent method. The monotonicity (A.3) now reads

$$\|P_{\overline{\mathsf{R}(T)}}g - Tf_{m+1}\|_Y < \|P_{\overline{\mathsf{R}(T)}}g - Tf_m\|_Y.$$

Thus,

$$\lim_{m\to\infty} \|P_{\overline{\mathsf{R}(T)}}g - Tf_m\|_Y = \varepsilon.$$

It remains to confirm that $\varepsilon = 0$. Assume the contrary: $\varepsilon > 0$. Then, there exists an $f^\varepsilon \in X$ with

$$\|P_{\overline{\mathsf{R}(T)}}g - Tf^\varepsilon\|_Y < \frac{\varepsilon}{4}.$$

Straightforward calculations yield

$$\begin{aligned}\|f_{m+1} - f^\varepsilon\|_X^2 - \|f_m - f^\varepsilon\|_X^2 = 2\lambda_m \big\langle r_m, P_{\overline{\mathsf{R}(T)}}g - Tf^\varepsilon\big\rangle_Y \\ - 2\lambda_m \|r_m\|_Y^2 + \lambda_m^2 \|T^* r_m\|_X^2.\end{aligned}$$

We proceed with

$$\begin{aligned}\lambda_m \|T^* r_m\|_X^2 &= \lambda_m \big\langle TT^* r_m, r_m\big\rangle_Y \\ &\le \lambda_m \|TT^* r_m\|_Y \|r_m\|_Y = \frac{\langle TT^* r_m, r_m\rangle_Y}{\|TT^* r_m\|_Y}\, \|r_m\|_Y \le \|r_m\|_Y^2\end{aligned}$$

implying

$$\begin{aligned}\|f_{m+1} - f^\varepsilon\|_X^2 - \|f_m - f^\varepsilon\|_X^2 &< 2\,\lambda_m \|r_m\|_Y\, \frac{\varepsilon}{4} - \lambda_m \|r_m\|_Y^2 \\ &= \lambda_m \|r_m\|_Y \Big(\frac{\varepsilon}{2} - \|r_m\|_Y\Big).\end{aligned}$$

As $\|r_m\|_Y > \varepsilon$ for all m we have

$$\|f_{m+1} - f^\varepsilon\|_X^2 - \|f_m - f^\varepsilon\|_X^2 < -\frac{\varepsilon}{2}\,\lambda_m \|r_m\|_Y.$$

Adding both sides of the above inequality from $m = 0$ to $m = k - 1$ gives

$$\|f_k - f^\varepsilon\|_X^2 - \|f^\varepsilon\|_X^2 < -\frac{\varepsilon}{2} \sum_{m=0}^{k-1} \lambda_m \|r_m\|_Y.$$

Since $\lambda_m \geq \|T\|^{-2}$ we end up with

$$\sum_{m=0}^{k-1} \|r_m\|_Y < \frac{2\|T\|^2}{\varepsilon}\big(\|f^\varepsilon\|_X^2 - \|f_k - f^\varepsilon\|_X^2\big) \leq \frac{2\|T\|^2}{\varepsilon}\|f^\varepsilon\|_X^2.$$

The upper bound does not depend on k contradicting $\|r_m\|_Y > \varepsilon > 0$.

Appendix B. Proof of monotonicity (3.2) **for Landweber, steepest decent, implicit iteration and cg.** We profit from results of Hämarik and Tautenhahn [6].

Applied to the normal equation $T^*Tf = T^*g$ (notation as in Appendix A) the four methods under consideration produce iterates $\{f_m\}_{m\in\mathbb{N}}$ by

$$f_{m+1} = f_m + T^* z_m, \quad f_0 = 0,$$

where

- Landweber: $z_m = \omega r_m$, $\omega \in]0, \|T\|^{-2}[$,
- steepest decent: $z_m = \lambda_m r_m$,
- implicit iteration: $z_m = (\alpha_m I + TT^*)^{-1} r_m$, and
- cg: $z_m = w_{m+1}(TT^*)g$ for a polynomial w_{m+1} of degree $m+1$, see Hanke [7, formula (2.7)].

For any $\widetilde{f} \in X$ we have that

$$\|f_m - \widetilde{f}\|_X^2 - \|f_{m-1} - \widetilde{f}\|_X^2 = 2\langle g - T\widetilde{f}, z_{m-1}\rangle_Y - \langle r_{m-1} + r_m, z_{m-1}\rangle_Y, \tag{B.1}$$

see [6, formula (3.2)].

Let $\gamma = \|g - T\widetilde{f}\|_Y / \|g\|_Y$ denote the relative residual of $\widetilde{f}$.

B.1. Landweber and steepest decent. Plugging in $z_m = \beta_m r_m$ with $\beta_m \in \{\omega, \lambda_m\}$ we obtain from (B.1)

$$\begin{aligned}\|f_m - \widetilde{f}\|_X^2 - \|f_{m-1} - \widetilde{f}\|_X^2 = \beta_{m-1}\big(2\langle g - T\widetilde{f}, r_{m-1}\rangle_Y \\ - \|r_{m-1}\|_Y^2 - \langle r_m, r_{m-1}\rangle_Y\big).\end{aligned}$$

By

$$\langle r_m, r_{m-1}\rangle_Y = \langle (I - \beta_{m-1}TT^*) r_{m-1}, r_{m-1}\rangle_Y > 0$$

we end up with

$$\begin{aligned}\|f_m - \widetilde{f}\|_X^2 - \|f_{m-1} - \widetilde{f}\|_X^2 &< \beta_{m-1}\|r_{m-1}\|_Y\big(2\|g - T\widetilde{f}\|_Y - \|r_{m-1}\|_Y\big) \\ &= \beta_{m-1}\|r_{m-1}\|_Y\|g\|_Y\left(\Psi(\gamma) - \frac{\|r_{m-1}\|_Y}{\|g\|_Y}\right)\end{aligned}$$

where $\Psi(t) = 2t$. Thus, we have established (3.2) for Landweber as well as steepest decent.

B.2. Implicit iteration. Next we address implicit iteration. Since $z_{m-1} = \alpha_{m-1}^{-1} r_m$ we deduce $\langle r_m, z_{m-1}\rangle_Y > 0$. Further, $\langle r_{m-1}, z_{m-1}\rangle_Y \geq \alpha_{\min}\|z_{m-1}\|_Y^2$. By (B.1),

$$\|f_m - \widetilde{f}\|_X^2 - \|f_{m-1} - \widetilde{f}\|_X^2 < \|z_{m-1}\|_Y \big(2\|g - T\widetilde{f}\|_Y - \alpha_{\min}\|z_{m-1}\|_Y\big).$$

The lower bound $\|z_{m-1}\|_Y \geq (\alpha_{\max} + \|T\|^2)^{-1}\|r_{m-1}\|_Y$ yields

$$\|f_m - \widetilde{f}\|_X^2 - \|f_{m-1} - \widetilde{f}\|_X^2$$
$$< \|z_{m-1}\|_Y \|g\|_Y \frac{\alpha_{\min}}{\alpha_{\max} + \|T\|^2}\left(\Psi(\gamma) - \frac{\|r_{m-1}\|_Y}{\|g\|_Y}\right)$$

with $\Psi(t) = 2\,\frac{\alpha_{\max}+\|T\|^2}{\alpha_{\min}}\,t$ and (3.2) follows for the implicit iteration.

B.3. cg-method. We follow arguments by Hanke [8, Theorem 3.1]. Here (B.1) reads

$$\|f_m - \widetilde{f}\|_X^2 - \|f_{m-1} - \widetilde{f}\|_X^2$$
$$= 2\langle g - T\widetilde{f}, w_m(TT^*)g\rangle_Y - \langle r_{m-1} + r_m, w_m(TT^*)g\rangle_Y,$$

To proceed we rewrite w_m as $w_m(t) = w_m(0) + tq(t)$ where $q \in \Pi_{m-1}$ and $w_m(0) > 0$. Hence, $w_m(TT^*)g = w_m(0)g + Tu$ with $u = T^*q(TT^*)g \in U_{m-1}$. Applying (A.1) and (A.2) we obtain

$$\begin{aligned}\langle r_{m-1}, w_m(TT^*)g\rangle_Y &= w_m(0)\langle g - Tf_{m-1}, g\rangle_Y + \langle g - Tf_{m-1}, Tu\rangle_Y\\ &= w_m(0)\|r_{m-1}\|_Y^2.\end{aligned}$$

Analogously,

$$\langle r_m, w_m(TT^*)g\rangle_Y = w_m(0)\|r_m\|_Y^2.$$

Thus,

$$\|f_m - \widetilde{f}\|_X^2 - \|f_{m-1} - \widetilde{f}\|_X^2$$
$$\leq w_m(0)\left(2\,\frac{\|w_m(TT^*)g\|_Y}{w_m(0)}\,\|g - T\widetilde{f}\|_Y - \|r_{m-1}\|_Y^2\right).$$

The normalized polynomial $w_m/w_m(0)$ is denoted $p_m^{[2]}$ by Hanke [7]. By his Theorem 3.2 we have

$$\frac{\|w_m(TT^*)g\|_Y}{w_m(0)} < \frac{\|w_0(TT^*)g\|_Y}{w_0(0)} = \|g\|_Y,$$

so that

$$\|f_m - \widetilde{f}\|_X^2 - \|f_{m-1} - \widetilde{f}\|_X^2 < w_m(0)\|g\|_Y^2\left(\Psi(\gamma)^2 - \frac{\|r_{m-1}\|_Y^2}{\|g\|_Y^2}\right)$$

with $\Psi(t) = \sqrt{2t}$ and we have established (3.2) for the cg-method.

REFERENCES

[1] A. B. Bakushinsky and M. Y. Kokurin, *Iterative Methods for Approximate Solution of Inverse Problems*, vol. 577 of Mathematics and Its Applications, Springer, Dordrecht, The Netherlands, 2004.

[2] H. Brakhage, *On ill-posed problems and the method of conjugate gradients*, in Inverse and Ill-Posed Problems, H. W. Engl and C. W. Groetsch, eds., vol. 4 of Notes and Reports in Mathematics in Science and Engineering, Boston, 1987, Academic Press, pp. 165–175.

[3] P. Deuflhard, *Newton Methods for Nonlinear Problems*, vol. 35 of Springer Series in Computational Mathematics, Springer, Berlin, 2004.

[4] P. Deuflhard, H. W. Engl, and O. Scherzer, *A convergence analysis of iterative methods for the solution of nonlinear ill-posed problems under affinely invariant conditions*, Inverse Problems, 14 (1998), pp. 1081–1106.

[5] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, vol. 375 of Mathematics and Its Applications, Kluwer Academic Publishers, Dordrecht, 1996.

[6] U. Hämarik and U. Tautenhahn, *On the monotone error rule for parameter choice in iterative and continuous regularization methods*, BIT, 41 (2001), pp. 1029–1038.

[7] M. Hanke, *Conjugate Gradient Type Methods for Ill-Posed Problems*, vol. 327 of Pitman Research Notes in Mathematics, Longman Scientific & Technical, Harlow, UK, 1995.

[8] ———, *Regularizing properties of a truncated Newton-CG algorithm for nonlinear inverse problems*, Numer. Funct. Anal. Optim., 18 (1998), pp. 971–993.

[9] Q. Jin and U. Tautenhahn, *On the discrepancy principle for some Newton type methods for solving nonlinear ill-posed problems*, Numer. Math, 111 (2009), pp. 509–558.

[10] B. Kaltenbacher, A. Neubauer, and O. Scherzer, *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*, Radon Series on Computational and Applied Mathematics, de Gruyter, Berlin, 2008.

[11] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems*, vol. 120 of Applied Mathematical Sciences, Springer-Verlag, New York, 1996.

[12] A. Lechleiter and A. Rieder, *Newton regularizations for impedance tomography: convergence by local injectivity*, Inverse Problems, 24 (2008), p. 065009.

[13] A. K. Louis, *Inverse und schlecht gestellte Probleme*, Studienbücher Mathematik, B. G. Teubner, Stuttgart, Germany, 1989.

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

[15] ———, *Keine Probleme mit Inversen Problemen*, Vieweg, Wiesbaden, Germany, 2003.

[16] O. Scherzer, *The use of Morozov's discrepancy principle for Tikhonov regularization for solving nonlinear ill-posed problems*, Computing, 51 (1993), pp. 45–60.

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

IWRMM-Preprints seit 2008

Nr. 08/01	Patrizio Neff, Antje Sydow, Christian Wieners: Numerical approximation of incremental infinitesimal gradient plasticity
Nr. 08/02	Götz Alefeld, Zhengyu Wang: Error Estimation for Nonlinear Complementarity Problems via Linear Systems with Interval Data
Nr. 08/03	Ulrich Kulisch : Complete Interval Arithmetic and its Implementation on the Computer
Nr. 08/04	Armin Lechleiter, Andreas Rieder: Newton Regularizations for Impedance Tomography: Convergence by Local Injectivity
Nr. 08/05	Vu Hoang, Michael Plum, Christian Wieners: A computer-assisted proof for photonic band gaps
Nr. 08/06	Vincent Heuveline, Peter Wittwer: Adaptive boundary conditions for exterior stationary flows in three dimensions
Nr. 08/07	Jan Mayer: Parallel Algorithms for Solving Linear Systems with Sparse Triangular Matrices
Nr. 08/08	Ulrich Kulisch : Begegnungen eines Mathematikers mit Informatik
Nr. 08/09	Tomas Dohnal, Michael Plum, Wolfgang Reichel: Localized Modes of the Linear Periodic Schrödinger Operator with a Nonlocal Perturbation
Nr. 08/10	Götz Alefeld: Verified Numerical Computation for Nonlinear Equations
Nr. 08/11	Götz Alefeld, Zhengyu Wang: Error bounds for complementarity problems with tridiagonal nonlinear functions
Nr. 08/12	Tomas Dohnal, Hannes Uecker: Coupled Mode Equations and Gap Solitons for the 2D Gross-Pitaevskii equation with a non-separable periodic potential
Nr. 09/01	Armin Lechleiter, Andreas Rieder: Towards A General Convergence Theory For Inexact Newton Regularizations

Eine aktuelle Liste aller IWRMM-Preprints finden Sie auf:

www.mathematik.uni-karlsruhe.de/iwrmm/seite/preprints