Numer. Math. (2010) 114:521–548
DOI 10.1007/s00211-009-0256-0

Numerische Mathematik

Towards a general convergence theory for inexact Newton regularizations

Armin Lechleiter · Andreas Rieder

Received: 3 February 2009 / Revised: 8 June 2009 / Published online: 15 September 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.

Mathematics Subject Classification (2000) 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. [2,13,20] 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.

Dedicated to Alfred K. Louis on the occasion of his 60th birthday.

A. Lechleiter
Centre de Mathématiques Appliqués (CMAP), Ecole Polytechnique, 91128 Palaiseau Cedex, France
e-mail: alechle@cmapx.polytechnique.fr

A. Rieder (✉)
Institut für Angewandte und Numerische Mathematik and Institut für Wissenschaftliches Rechnen und Mathematische Modellbildung, Universität Karlsruhe, 76128 Karlsruhe, Germany
e-mail: andreas.rieder@math.uni-karlsruhe.de

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. [6,14,17,20]. 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 (REGularization based on INexact Newton iteration) by the second author [18,19,21]. 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 five features which not only guarantee convergence but are also shared by various inner regularization schemes which 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

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

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).

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

$$m_n = \min\left\{m \in \mathbb{N} \, : \, \|A_n s_{n,m} - b_n^\varepsilon\|_Y < \mu_n \|b_n^\varepsilon\|_Y\right\}, \tag{1.7}$$

and set

$$s_n^{\mathrm{N}} := s_{n,m_n}. \tag{1.8}$$

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 than $\mu_n \|b_n^\varepsilon\|_Y$.

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

$$\|y^\delta - F(x_{N(\delta)})\|_Y \leq R\delta < \|y^\delta - F(x_n)\|_Y, \quad n = 0, 1, \ldots, N(\delta) - 1, \tag{1.9}$$

see Fig. 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. [13,20]. Under our assumptions `REGINN` is well defined and terminates. Moreover, all iterates stay in the level set $\mathcal{L} = \{x \in \mathsf{D}(F)\colon \|F(x) - y^\delta\|_Y \leq \|F(x_0) - y^\delta\|_Y\}$. Unfortunately, $\mathcal{L}$ cannot be assumed bounded, thus prohibiting the use of a weak-compactness argument to verify weak convergence at least.

Local convergence, however, is our topic in Sect. 3. Provided the regularizing sequence $\{s_{n,m}\}$ exhibits a specific monotone error decrease (up to a stopping index)

```
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;
```

Fig. 1 REGINN: REGularization based on INexact Newton iteration

all REGINN-iterates will stay in a ball about x^+. Finally, we prove strong convergence of $\{x_{N(\delta)}\}_\delta$ as $\delta \to 0$.

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.

We did not include a section on numerical experiments as we do not present new methods but a new and unified convergence analysis for these methods. The interested reader should consult our paper [15] where we solved the inverse problem of impedance tomography using REGINN furnished with the conjugate gradients iteration for computing the Newton update (1.8).

2 Residual and level set based analysis

For the analysis of REGINN we require three properties of the regularizing sequence $\{s_{n,m}\}$, namely

$$\langle A_n s_{n,m}, b_n^\varepsilon\rangle_Y > 0 \;\; \forall m \geq 1 \text{ whenever } A_n^* b_n^\varepsilon \neq 0, \tag{2.1}$$

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

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

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

- Landweber iteration: $g_m(\lambda) = \lambda^{-1}(1 - (1 - \omega\lambda)^m)$ 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:

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

  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 [3]: 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 regularization schemes, which cannot be represented by filter functions, satisfy (2.1), (2.2), and (2.3):

- steepest decent method where $\Theta \leq 2$,[1] and
- method of conjugate gradients (cg-method) where $\Theta = 1$,

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

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

Remark 2.1 Recently, Jin and Tautenhahn [12] presented a subtle convergence analysis of (generalized) iteratively regularized Gauß–Newton methods,

$$x_{n+1} = x_n + g_{m_n}\left(A_n^* A_n\right) A_n^* b_n^\varepsilon + \left(I - g_{m_n}\left(A_n^* A_n\right) A_n^* A_n\right)(x_0 - x_n), \quad (2.4)$$

stopped by the discrepancy principle (1.9). The iteratively regularized Gauß–Newton method (with Tikhonov regularization) has been introduced by Bakushinsky in his pioneering work [1]. The differences of (2.4) to REGINN consist in the rightmost term and 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) Jin and Tautenhahn proved deep and far-reaching convergence results. Under weaker assumptions, not covered by Theorems 1, 2 or 3 in [12], we obtain weaker convergence results. However, the technique of Jin and Tautenhahn does not apply to REGINN [12, Remark 3] and cannot be extended to other nonlinear regularization schemes in a straightforward way.

Now we present first results. By (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*

$$\langle \nabla\varphi(x_n), s_{n,m}\rangle_X < 0 \quad \textit{for } m \geq 1 \textit{ whenever } A_n^* b_n^\varepsilon \neq 0.$$

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

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

and the lemma is verified. □

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*

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

the stopping index m_n (1.7) *is well defined.*

Proof By (2.2)

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

which completes the proof. □

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 (Fig. 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

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

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:

$$\begin{aligned} &\|E(v,w)\|_Y \le L\,\|F'(w)(v-w)\|_Y \;\text{ for one } L<1 \\ &\text{and for all } v, w \in \mathcal{L}(x_0) \text{ with } v-w \in \mathsf{N}\left(F'(w)\right)^\perp. \end{aligned} \tag{2.5}$$

From (2.5) we derive that

$$\|E(v,w)\|_Y \le \omega\,\|F(w) - F(v)\|_Y \;\text{ where } \omega = \frac{L}{1-L} > L \tag{2.6}$$

which is the *tangential cone condition* introduced by Scherzer [22]. In the convergence analysis of Newton methods for ill-posed problems, both (2.5) and (2.6) 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. [13, Sect. 2.1] for a detailed explanation.

Remark 2.5 Actually, (2.5) and (2.6) are equivalent in the following sense: (2.6) for one $\omega < 1$ implies (2.5) with $L = \frac{\omega}{1-\omega}$.

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

$$\begin{aligned} &\left\|P_{\mathsf{R}(F'(u))^\perp}\left(F(x^+) - F(u)\right)\right\|_Y \le \varrho\|F(x^+) - F(u)\|_Y \\ &\text{for all } u \in \mathcal{L}(x_0). \end{aligned} \tag{2.7}$$

Assumption (2.7) is quite natural as it characterizes those nonlinear problems which can be tackled by local linearization (compare Remark 2.4): As (2.7) is equivalent to

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

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 (2.5) and (2.7) are satisfied globally in the domain of definition.

Example 2.6 Let $f: \mathbb{R} \to \mathbb{R}$ 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

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

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

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

with range $\mathsf{R}(F'(x)) = \left\{y \in Y \ : \ \{n\langle y, u_n\rangle_Y\}_n \in \ell^2\right\}$. Clearly, $\overline{\mathsf{R}(F'(x))} = Y$. Hence, (2.7) 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 $f'_{\max} < 2 f'_{\min}$. 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

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

For L small enough, (2.5) implies (2.7).

Lemma 2.7 *Assume* (2.5) *to hold with* $L < 1/2$. *Then,* (2.7) *holds for*

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

Proof We have that

$$\begin{aligned} &\left\| P_{\mathsf{R}(F'(u))^\perp} \left(F(x^+) - F(u)\right) \right\|_Y \\ &\quad = \left\| P_{\mathsf{R}(F'(u))^\perp} \left(F(x^+) - F(u) - F'(u)(x^+ - u)\right) \right\|_Y \\ &\quad \leq L \, \|F'(u)(x^+ - u)\|_Y . \end{aligned}$$

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. □

Theorem 2.8 *Let* $\mathsf{D}(F)$ *be open and choose* $x_0 \in \mathsf{D}(F)$ *such that* $\mathcal{L}(x_0) \subset \mathsf{D}(F)$. *Assume* (2.1), (2.2), (2.3), (2.5), *and* (2.7) *to hold true with* Θ, L, *and* ϱ *satisfying*

$$\Theta L + \varrho < \Lambda \quad \textit{for one } \Lambda < 1. \tag{2.8}$$

Further, choose

$$R > \frac{1+\varrho}{\Lambda - \Theta L - \varrho}. \tag{2.9}$$

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

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

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

$$\|y^\delta - F(x_{N(\delta)})\|_Y \le R\delta, \tag{2.10}$$

and the nonlinear residuals decrease linearly at an estimated rate

$$\frac{\|y^\delta - F(x_{n+1})\|_Y}{\|y^\delta - F(x_n)\|_Y} < \mu_n + \theta_n L \le \Lambda, \quad n = 0, \ldots, N(\delta) - 1, \tag{2.11}$$

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 (2.8) 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

$$\mu_{\min,n} = \frac{(1+\varrho)\delta}{\|b_n^\varepsilon\|_Y} + \varrho < \frac{1+\varrho}{R} + \varrho \overset{(2.9)}{<} \Lambda - \Theta L. \tag{2.12}$$

We will argue inductively and therefore assume the iterates $\{x_1, \dots, x_n\}$ 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}:

$$\begin{aligned}\frac{\|P_{\mathsf{R}(A_n)^\perp} b_n^\varepsilon\|_Y}{\|b_n^\varepsilon\|_Y} &\le \frac{\delta + \|P_{\mathsf{R}(A_n)^\perp}(F(x^+) - F(x_n))\|_Y}{\|b_n^\varepsilon\|_Y} \\ &\overset{(2.7)}{\le} \frac{\delta + \varrho\|F(x^+) - F(x_n)\|_Y}{\|b_n^\varepsilon\|_Y} \\ &\le \frac{(1+\varrho)\delta + \varrho\|b_n^\varepsilon\|_Y}{\|b_n^\varepsilon\|_Y} = \mu_{\min,n} < \mu_n. \end{aligned} \tag{2.13}$$

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 $\mathcal{L}(x_0)$. First, s_n^{N} is a decent direction: Indeed $A_n^* b_n^\varepsilon = 0$ gives $b_n^\varepsilon \in \mathsf{R}(A_n)^\perp$ contradicting (2.13). Hence, $A_n^* b_n^\varepsilon \neq 0$ and Lemma 2.2 applies. As $\mathsf{D}(F)$ 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

$$\begin{aligned}\|y^\delta - F(x_{n,\lambda})\|_Y &= \|y^\delta - F(x_n) - \lambda A_n s_n^{\mathrm{N}} - (F(x_{n,\lambda}) - F(x_n) - \lambda A_n s_n^{\mathrm{N}})\|_Y \\ &\overset{(2.5)}{\le} \|y^\delta - F(x_n) - \lambda A_n s_n^{\mathrm{N}}\|_Y + L\,\lambda\,\|A_n s_n^{\mathrm{N}}\|_Y \\ &\le \|(1-\lambda) b_n^\varepsilon + \lambda(b_n^\varepsilon - A_n s_n^{\mathrm{N}})\|_Y + L\,\lambda\,\theta_n\,\|b_n^\varepsilon\|_Y \\ &< (1-\lambda)\|b_n^\varepsilon\|_Y + \mu_n\,\lambda\,\|b_n^\varepsilon\|_Y + L\,\lambda\,\theta_n\,\|b_n^\varepsilon\|_Y \\ &\le (1 - \lambda(1-\Lambda))\;\|b_n^\varepsilon\|_Y. \end{aligned} \tag{2.14}$$

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 (2.14) that

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

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, $\|b_{n+1}^\varepsilon\|_Y < (\mu_n + \theta_n L)\|b_n^\varepsilon\|_Y$ by plugging $\lambda = 1$ into (2.14). □

A few comments are in order.

Remark 2.9 Deuflhard et al. [5, formula (2.11)] have basically introduced the following Newton–Mysovskikh-like condition

$$\left\| \big(F'(v) - F'(w)\big)\, F'(w)^{+} \right\| \le L \quad \text{for all } v, w \in \mathcal{L}(x_0) \tag{2.15}$$

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 [5, Lemma 2.3].

If $\mathcal{L}(x_0)$ is convex then (2.15) implies (2.5). 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

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

Remark 2.10 An assumption similar to (2.7) is

$$\begin{aligned}
&\left\| P_{\mathsf{R}(F'(u))^{\perp}} (\eta - F(u)) \right\|_Y \le \widetilde{\varrho} \|\eta - F(u)\|_Y \quad \text{for one } \widetilde{\varrho} < 1 \\
&\quad \text{and for all } u \in \mathcal{L}(x_0) \text{ and all } \eta \in Y \text{ with } \|\eta - F(x^{+})\|_Y \le \delta_{\max}.
\end{aligned} \tag{2.16}$$

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

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

the assertion of Theorem 2.8 remains true whenever $\widetilde{\varrho} + \Theta L < \Lambda$, $\{\mu_n\} \subset]\widetilde{\varrho}, \Lambda - \Theta L]$ and $R > 0$ (no other restriction on R, compare (2.9)).

The mapping from Example 2.6 satisfies (2.16) with $\widetilde{\varrho} = 0$ for any $\delta_{\max} \ge 0$. Nevertheless, (2.16) is quite restrictive. While (2.7) holds trivially for any linear mapping (with $\varrho = 0$), (2.16) 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 (2.16) as well as $\overline{\mathsf{R}(F)} \ne Y$. 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

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

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

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

contradicts (2.16).

3 Local convergence

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

Since

$$\|y - F(x_{N(\delta)})\|_Y \overset{(2.10)}{<} (R+1)\delta \tag{3.1}$$

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 convergence of $\{x_{N(\delta)}\}$ provided the regularizing sequence $\{s_{n,m}\}_{0\le m\le m_n}$ exhibits a fourth property in addition to those from (2.1), (2.2), and (2.3). We require the following:

For any $n \in \{1, \dots, m_n\}$ there is a $v_{n,m-1} \in Y$ such that

$$s_{n,m} = s_{n,m-1} + A_n^* v_{n,m-1}.$$

Further, let there be a continuous and monotonically increasing function $\Psi: \mathbb{R} \to \mathbb{R}$ with $t \le \Psi(t)$ for $t \in [0,1]$ such that if $\gamma_n = \|b_n^\varepsilon - A_n s_n^{\mathrm{e}}\|_Y / \|b_n^\varepsilon\|_Y < 1$ then

$$\|s_{n,m} - s_n^{\mathrm{e}}\|_X^2 - \|s_{n,m-1} - s_n^{\mathrm{e}}\|_X^2 < C_{\mathrm{M}} \|b_n^\varepsilon\|_Y \|v_{n,m-1}\|_Y \left(\Psi(\gamma_n) - \frac{\|b_n^\varepsilon - A_n s_{n,m-1}\|_Y}{\|b_n^\varepsilon\|_Y}\right)$$

for $m = 1, \dots, m_n$ where $C_{\mathrm{M}} > 0$ is a constant. (3.2)

A direct consequence of (3.2) is monotonicity, i.e.,

$$\frac{\|b_n^\varepsilon - A_n s_{n,m-1}\|_Y}{\|b_n^\varepsilon\|_Y} \geq \Psi(\gamma_n) \implies \|s_{n,m} - s_n^{\mathrm{e}}\|_X < \|s_{n,m-1} - s_n^{\mathrm{e}}\|_X. \tag{3.3}$$

Examples of methods with property (3.2) are

- Landweber iteration and steepest decent: $\Psi(t) = 2t$,
- Implicit iteration: $\Psi(t) = 2\,\frac{\alpha_{\max} + s}{\alpha_{\min}}\,t$ 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.

3.1 Monotone error decay

Under (3.2) we formulate a version of Theorem 2.8 where all assumptions are related to a ball about x^+, that is, the implicitly defined, generally unbounded level set $\mathcal{L}(x_0)$ is replaced by $B_r(x^+)$. Especially, (2.5) is replaced by

$$\|E(v, w)\|_Y \leq L\,\|F'(w)(v - w)\|_Y \;\text{ for one } L < 1$$
$$\text{and for all } v, w \in B_r(x^+) \subset \mathsf{D}(F). \tag{3.4}$$

Theorem 3.1 *Assume (2.1), (2.2), (2.3), (3.2). Additionally, let (2.7) hold true in $B_r(x^+)$ and assume (3.4) with L satisfying*

$$\Psi\left(\frac{L}{1-L}\right) + \Theta L < \Lambda \quad \textit{for one } \Lambda < 1.^{2} \tag{3.5}$$

Further, define

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

and choose R so large that

$$\mu_{\min} + \Theta L < \Lambda. \tag{3.6}$$

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

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

$$\|x^+ - x_n\|_X < \|x^+ - x_{n-1}\|_X, \quad n = 1, \ldots, N(\delta). \tag{3.7}$$

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

Moreover, only the final iterate satisfies the discrepancy principle (2.10) *and the nonlinear residuals decrease linearly at the estimated rate* (2.11).

Proof Let us first discuss our assumptions. If (3.5) applies then, by continuity of Ψ, there exists a R such that $\mu_{\min}$ satisfies (3.6) and the interval for selecting the tolerances is nonempty.

As before we use an inductive argument: Assume the iterates $x_1, \ldots, x_n$ 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 (2.13) and (2.12) we have that

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

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 (3.7). We will rely on (3.3). By (1.5) and (3.4), 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

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

and then

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

Accordingly, $\|b_n^\varepsilon - A_n s_{n,m-1}\|_Y \geq \mu_{\min}\|b_n^\varepsilon\|_Y$, $m = 1, \ldots, m_n$, and we have by repeatedly applying the monotonicity (3.3)

$$\begin{aligned}\|x^+ - x_{n+1}\|_X &= \|s_n^{\mathrm{e}} - s_{n,m_n}\|_X \\ &< \|s_n^{\mathrm{e}} - s_{n,m_n-1}\|_X < \|s_n^{\mathrm{e}} - s_{n,m_n-2}\|_X \\ &< \cdots < \|s_n^{\mathrm{e}} - s_{n,0}\|_X = \|s_n^{\mathrm{e}}\|_X = \|x^+ - x_n\|_X \end{aligned} \tag{3.8}$$

which is (3.7). □

Remark 3.2 Some nonlinear ill-posed problems (such as a model in electrical impedance tomography, see [16]) satisfy a slightly stronger version of (3.4) where L is replaced by $C\|v - w\|_X$. In view of (3.7) we expect in this situation the reduction rate (2.11) to approach μ_n as the Newton iteration progresses.

Remark 3.3 A stronger assumption than (3.4) is

$$\|E(v,w)\|_Y \le \widetilde{L}\,\|F'(w)(v-w)\|_Y^{1+\kappa} \quad \text{for one } \kappa > 0$$
$$\text{and for all } v, w \in B_r(x^+). \tag{3.9}$$

Here, $\widetilde{L}$ is allowed to be arbitrarily large. If r is sufficiently small we have (3.4) 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 $\|y^\delta - F(x_0)\|_Y^\kappa \le L/\widetilde{L}$.[3] Then, all assertions of Theorem 3.1 remain valid with the stronger rate

$$\frac{\|y^\delta - F(x_{n+1})\|_Y}{\|y^\delta - F(x_n)\|_Y} \le \mu_n + \theta_n^{1+\kappa}\, \Lambda^{\kappa n}\, L \le \Lambda, \quad n = 0, \ldots, N(\delta) - 1. \tag{3.10}$$

We only need to verify the rate. We have

$$\begin{aligned}\left\|b_{n+1}^\varepsilon\right\|_Y &= \left\|b_n^\varepsilon - A_n s_n^{\mathrm{N}} + E(x_{n+1}, x_n)\right\|_Y \overset{(3.9)}{\le} \mu_n \left\|b_n^\varepsilon\right\|_Y + \widetilde{L}\left\|A_n s_n^{\mathrm{N}}\right\|_Y^{1+\kappa} \\ &\le \left(\mu_n + \widetilde{L}\theta_n^{1+\kappa}\left\|b_n^\varepsilon\right\|_Y^\kappa\right)\left\|b_n^\varepsilon\right\|_Y\end{aligned}$$

which inductively implies (3.10).

Remark 3.4 Both bounds (3.4) and (3.9) for the linearization error may be derived from the following *affine contravariant* Lipschitz condition:

$$\left\|\left(F'(v) - F'(w)\right)(v-w)\right\|_Y \le L_\kappa\, \|F'(w)(v-w)\|_Y^{1+\kappa}$$
$$\text{for one } \kappa \in [0,1] \text{ and for all } v, w \in B_r(x^+) \tag{3.11}$$

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

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

For a general discussion of the importance of affine contravariance for Newton-like algorithms we refer to Sect. 1.2.2 of Deuflhard's book [4]. In particular, Sect. 4.2 of the same book treats Gauß–Newton methods for (well-posed) finite dimensional least squares problems under (3.11) globally in $\mathsf{D}(F)$ and with $\kappa = 1$.

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

3.2 Convergence

We now turn to the convergence of $\{x_{N(\delta)}\}_{0<\delta\le\delta_{\max}}$ as $\delta \to 0$.

Corollary 3.5 *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 [24]. 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 $0 \ne v \in \mathsf{N}(A)$ then

$$\|F(x^+ + tv) - y\|_Y = \|F(x^+ + tv) - F(x^+)\|_Y \overset{(3.4)}{\le} (L+1)\,|t|\,\|Av\|_Y = 0$$

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

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

Corollary 3.6 *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 (3.4) we obtain that

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

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.

However, strong convergence in X can be verified under an additional stability assumption on the regularization scheme applied to the locally linearized system (1.6). To formulate this assumption we introduce new notation: We subsequently need to differ clearly between the noisy ($\delta > 0$) and the noiseless situation ($\delta = 0$). From now on in this section, quantities referring to the noisy setting (i.e., $y^\delta \ne y$) will be marked by a superscript δ. Quantities without superscript indicate exact data (i.e., $y = y^0$).

We require the following stability of the regularizing sequence $\{s_{n,m}\}_{0\le m\le m_n}$:

$$\lim_{\delta\to 0} s^{\delta}_{n,m} = s_{n,m} \quad \text{for any fixed } m \le m_n. \tag{3.12}$$

All four examples satisfying (3.2) share also stability:

- Landweber and implicit iteration are linear iterative schemes. Thus, stability (3.12) can be shown straightforwardly by an inductive argument which we shortly present for the implicit iteration: Let $T, T_\gamma \in \mathcal{L}(X,Y)$ with $\lim_{\gamma\to 0}\|T - T_\gamma\| = 0$. Further, let $g, g_\gamma \in Y$ with $\lim_{\gamma\to 0}\|g - g_\gamma\|_Y = 0$. Define implicit iteration with respect to (T, g) and (T_γ, g_γ) by

$$f_{m+1} = (\alpha_m I + T^*T)^{-1}(\alpha_m f_m + T^* g)$$

and

$$f^{\gamma}_{m+1} = (\alpha_m I + T^*_\gamma T_\gamma)^{-1}(\alpha_m f^{\gamma}_m + T^*_\gamma g_\gamma),$$

respectively, where $f_0 = f^{\gamma}_0$. Since

$$\lim_{\gamma\to 0}\|(\alpha_m I + T^*_\gamma T_\gamma)^{-1} - (\alpha_m I + T^*T)^{-1}\| = 0$$

convergence of f^{γ}_m to f_m as $\gamma \to 0$ follows inductively for any m.
- Proof of stability is more complicated for the nonlinear steepest decent and cg iterations. Fortunately, we only need to refer to previous work of Scherzer [23, Lemma 3.2] and Hanke [10, Lemma 3.4] for steepest decent and cg, respectively. Both lemmas apply to our setting because early termination of both iterations does not occur before reaching the stopping index m_n.

Theorem 3.7 *Assume* (3.12) *and adopt all assumptions and notations of Theorem* 3.1*, however, restrict all tolerances* $\{\mu_n\}$ *to* $[\underline{\mu}, \Lambda - \Theta L]$ *for a* $\underline{\mu} > \mu_{\min}$*. Additionally let* x^+ *be the only solution of* $F(x) = y$ *in* $B_r(x^+)$*. Then,*

$$\lim_{\delta\to 0}\|x^+ - x^{\delta}_{N(\delta)}\|_X = 0.$$

For the proof of above theorem we basically generalize results of Hanke [10, Sects. 4 and 5] and [9] who, in turn, generalized ideas of Hanke et al. [11].

We start with some preparatory lemmas and validate first convergence of `REGINN` in the noiseless setting. In the remainder of this section we tacitly presume the assumptions of Theorem 3.7.

Lemma 3.8 *In the noiseless situation, that is,* $\delta = 0$*, we have that*

$$\lim_{n\to\infty}\|x^+ - x_n\|_X = 0.$$

Proof Please note that REGINN is well-defined for $\delta = 0$ under the hypotheses of Theorem 3.1, that is, if $x_{n-1} \in B_r(x^+)$ does not solve $F(x) = y$ then $x_n \in B_r(x_+)$ exists and satisfies (3.7) as well as (2.11). Therefore, if REGINN terminates early with x_N then $F(x_N) = y$ implying $x_N = x^+$ due to the uniqueness of x^+ in $B_r(x_+)$. Otherwise, $\{x_n\}_{n\in\mathbb{N}_0}$ is a Cauchy sequence which we will show now.

Let $l, p \in \mathbb{N}$ with $l > p$. We observe that

$$\|x_l - x_p\|_X^2 = \|s_l^{\mathrm{e}} - s_p^{\mathrm{e}}\|_X^2 = 2\langle s_l^{\mathrm{e}} - s_p^{\mathrm{e}}, s_l^{\mathrm{e}}\rangle_X + \|s_p^{\mathrm{e}}\|_X^2 - \|s_l^{\mathrm{e}}\|_X^2 \tag{3.13}$$

and

$$s_l^{\mathrm{e}} - s_p^{\mathrm{e}} = -\sum_{i=p}^{l-1} s_{i,m_i} \overset{(3.2)}{=} \sum_{i=p}^{l-1} A_i^* \widetilde{v}_i \quad \text{where } \widetilde{v}_i := -\sum_{k=1}^{m_i} v_{i,k-1}.$$

Hence,

$$\langle s_l^{\mathrm{e}} - s_p^{\mathrm{e}}, s_l^{\mathrm{e}}\rangle_X = \sum_{i=p}^{l-1} \langle \widetilde{v}_i, A_i s_l^{\mathrm{e}}\rangle_Y \le \sum_{i=p}^{l-1} \|\widetilde{v}_i\|_Y \|A_i s_l^{\mathrm{e}}\|_Y.$$

We proceed with

$$\begin{aligned}
\|A_i s_l^{\mathrm{e}}\|_Y &= \|F'(x_i)(x^+ - x_l)\|_Y \le \|F'(x_i)(x^+ - x_i)\|_Y + \|F'(x_i)(x_l - x_i)\|_Y \\
&\overset{(3.4)}{\le} \frac{1}{1-L}\left(\|y - F(x_i)\|_Y + \|F(x_i) - F(x_l)\|_Y\right) \\
&\le \frac{1}{1-L}\left(2\|y - F(x_i)\|_Y + \|y - F(x_l)\|_Y\right) \\
&\le \frac{3}{1-L}\|y - F(x_i)\|_Y
\end{aligned}$$

where the last estimate holds true due to the monotonicity of the residuals (2.11) and yields that

$$\langle s_l^{\mathrm{e}} - s_p^{\mathrm{e}}, s_l^{\mathrm{e}}\rangle_X \le \frac{3}{1-L}\sum_{i=p}^{l-1} \|\widetilde{v}_i\|_Y \|y - F(x_i)\|_Y.$$

For bounding $\|\widetilde{v}_i\|_Y \|y - F(x_i)\|_Y$ we will apply (3.2). Setting $n = i$ in the inequality in (3.2) and summing up both sides from $m = 1$ to m_i results in

$$\|s_{i,m_i} - s_i^{\mathrm{e}}\|_X^2 - \|s_i^{\mathrm{e}}\|_X^2 < C_{\mathrm{M}} \|b_i^{\varepsilon}\|_Y \sum_{m=1}^{m_i} \|v_{i,m-1}\|_Y \left(\Psi(\gamma_i) - \mu_i\right)$$

where we have taken into account that $\|b_i^\varepsilon - A_i s_{i,m-1}\|_Y \geq \mu_i \|b_i^\varepsilon\|_Y$. Recall that $b_i^\varepsilon = y - F(x_i)$, $s_{i+1}^{\mathrm{e}} = s_i^{\mathrm{e}} - s_{i,m_i}$, and $\mu_i \geq \underline{\mu} > \mu_{\min} \geq \Psi(\gamma_i)$. Thus,

$$\|\widetilde{v}_i\|_Y \|y - F(x_i)\|_Y \leq \sum_{m=1}^{m_i} \|v_{i,m-1}\|_Y \|y - F(x_i)\|_Y \leq \frac{\|s_i^{\mathrm{e}}\|_X^2 - \|s_{i+1}^{\mathrm{e}}\|_X^2}{C_{\mathrm{M}}(\underline{\mu} - \mu_{\min})}$$

and

$$\langle s_l^{\mathrm{e}} - s_p^{\mathrm{e}}, s_l^{\mathrm{e}} \rangle_X \leq \frac{3}{(1-L)C_{\mathrm{M}}(\underline{\mu} - \mu_{\min})} \left(\|s_p^{\mathrm{e}}\|_X^2 - \|s_l^{\mathrm{e}}\|_X^2 \right).$$

We plug the latter bound into (3.13) to obtain

$$\|x_l - x_p\|_X^2 \leq \left(\frac{6}{(1-L)C_{\mathrm{M}}(\underline{\mu} - \mu_{\min})} + 1 \right) \left(\|s_p^{\mathrm{e}}\|_X^2 - \|s_l^{\mathrm{e}}\|_X^2 \right).$$

The monotonicity $\|s_{n+1}^{\mathrm{e}}\|_X < \|s_n^{\mathrm{e}}\|_X$ forces the convergence of $\|s_n^{\mathrm{e}}\|_X$ as $n \to \infty$. So, $\|x_l - x_p\|_X$ can be made arbitrarily small by increasing p. Due to the uniqueness of x^+ in $B_r(x^+)$ the limit of $\{x_n\}_n$, which is in $B_r(x^+)$, has to be x^+. □

Below we will prove kind of stability of x_n^δ, $n \leq N(\delta)$, as $\delta \to 0$. To this end we introduce sets $\{\mathcal{X}_n\}_{n \in \mathbb{N}_0}$ defined recursively from `REGINN`-iterates for exact data y.

Definition 3.9 Set $\mathcal{X}_0 := \{x_0\}$ and determine $\mathcal{X}_{n+1}$ from $\mathcal{X}_n$ in the following way: for any $\xi_n \in \mathcal{X}_n$ compute the Newton step $s_n^{\mathrm{N}} = s_{n,m_n}$ as explained in (1.8) and (1.7) where, however, A_n is replaced by $F'(\xi_n)$ and b_n^ε by $y - F(\xi_n)$. Then, $\xi_n + s_n^{\mathrm{N}}$ belongs to $\mathcal{X}_{n+1}$. If

$$\left\| F'(\xi_n) s_{n,m_n-i} - (y - F(\xi_n)) \right\|_Y = \mu_n \|y - F(\xi_n)\|_Y \quad \text{for } i = 1, \ldots, k_n < m_n, \tag{3.14}$$

then $\xi_n + s_{n,m_n-i}$, $i = 1, \ldots, k_n$, are also elements of $\mathcal{X}_{n+1}$.

We call ξ_n the predecessor of $\xi_n + s_{n,m_n-i}$, $i = 0, \ldots, k_n$, and, in turn, call the latter successors of ξ_n.

Obviously, $x_n \in \mathcal{X}_n$ and, generically, $\mathcal{X}_n$ contains only this element. Assume that $\mathcal{X}_j = \{x_j\}$ for $j = 0, \ldots, n$ and that s_{n,m_n-1} as well as s_{n,m_n-2} satisfy (3.14) where $\xi_n = x_n$ of course. Then, $\mathcal{X}_{n+1} = \{x_{n+1}, x_n + s_{n,m_n-1}, x_n + s_{n,m_n-2}\}$ and from now on all sets $\mathcal{X}_j$ with $j \geq n+1$ will have three elements at least. By monotonicity (3.3), see also (3.8), we have that

$$\|x^+ - \xi_{n+1}\|_X < \|x^+ - \xi_n\|_X \quad \text{whenever } \xi_{n+1} \in \mathcal{X}_{n+1} \text{ is a successor of } \xi_n \in \mathcal{X}_n. \tag{3.15}$$

Remark 3.10 For the four iterative methods Landweber, implicit iteration, steepest decent, and cg the number k_n in (3.14) cannot exceed 1 for the following reason: the residuals of these methods decrease strictly monotonically up to the iteration index m_n.

The need for introducing the sets $\{\mathcal{X}_n\}_{n\in\mathbb{N}_0}$ becomes clear in the proof of the next lemma.

Lemma 3.11 *Algorithm* `REGINN` *is stable in the following sense: for fixed* $n \in \mathbb{N}_0$ *with* $n \leq N(\delta)$ *for all* δ *sufficiently small, the iterate* x_n^δ *converges to* $\mathcal{X}_n$ *as* $\delta \to 0$, *that is, for any zero sequence* $\{\delta_j\}_{j\in\mathbb{N}}$ *the sequence* $\{x_n^{\delta_j}\}_{j\in\mathbb{N}}$ *splits into convergent subsequences, all of which converge to elements of* $\mathcal{X}_n$.

Proof Before we start let us emphasize the difference of m_n and m_n^δ. Both are defined by (1.7), however, the former with respect to exact data and the latter with respect to noisy data.

We employ an inductive argument. For $n = 0$ the only element in $\mathcal{X}_0$ is x_0 independent of $\delta \geq 0$. Hence, the statement is true for $n = 0$. Assume now that x_n^δ converges to $\mathcal{X}_n$ as $\delta \to 0$ and that $n+1 \leq N(\delta)$ for δ sufficiently small. Let $\{\delta_i\}$ be a subsequence with $\lim_{i\to\infty} x_n^{\delta_i} = \xi_n$ for one $\xi_n \in \mathcal{X}_n$. By (3.12) and continuity the following limit holds true:

$$\lim_{i\to\infty} \left\| A_n^{\delta_i} s_{n,k}^{\delta_i} - b_n^{\varepsilon_i} \right\|_Y = \left\| F'(\xi_n) s_{n,k} - \widetilde{b}_n \right\|_Y, \quad k \in \{0, \ldots, m_n\},$$

where $b_n^{\varepsilon_i} = y^{\delta_i} - F(x_n^{\delta_i})$ and $\widetilde{b}_n = y - F(\xi_n)$. Since $\|F'(\xi_n)s_{n,m_n} - \widetilde{b}_n\|_Y < \mu_n \|\widetilde{b}_n\|_Y$ we conclude for i sufficiently large that

$$\left\| A_n^{\delta_i} s_{n,m_n}^{\delta_i} - b_n^{\varepsilon_i} \right\|_Y < \mu_n \left\| b_n^{\varepsilon_i} \right\|_Y$$

yielding $m_n^{\delta_i} \leq m_n$ for large enough i. In case $\mu_n \|\widetilde{b}_n\|_Y < \|F'(\xi_n)s_{n,m_n-1} - \widetilde{b}_n\|_Y$ we also have

$$\mu_n \left\| b_n^{\varepsilon_i} \right\|_Y < \left\| A_n^{\delta_i} s_{n,m_n-1}^{\delta_i} - b_n^{\varepsilon_i} \right\|_Y$$

provided large enough i. Thus, $m_n - 1 < m_n^{\delta_i}$ and $m_n^{\delta_i} = m_n$ for sufficiently large i, i.e., $\lim_{i\to\infty} s^{\delta_i}_{n,m_n^{\delta_i}} = s_{n,m_n}$ which gives $x_n^{\delta_i} + s^{\delta_i}_{n,m_n^{\delta_i}} \to \xi_n + s_{n,m_n} \in \mathcal{X}_{n+1}$ as $i \to \infty$.

In case (3.14) applies, we have $\mu_n \|\widetilde{b}_n\|_Y < \|F'(\xi_n)s_{n,m_n-(k_n+1)} - \widetilde{b}_n\|_Y$. Arguing as above we obtain $m_n - (k_n + 1) < m_n^{\delta_i}$ implying the inclusion $m_n - k_n \leq m_n^{\delta_i} \leq m_n$ for i sufficiently large. Accordingly, $\{m_n^{\delta_i}\}$ might have the $k_n + 1$ limit points $m_n - k_n, \ldots, m_n$. In any case, all possible limit points of $x_n^{\delta_i} + s^{\delta_i}_{n,m_n^{\delta_i}}$ are in $\mathcal{X}_{n+1}$ by construction of this set. □

The sets $\mathcal{X}_n$ converge uniformly to x^+.

Lemma 3.12 *For any $\eta > 0$ there is an $M(\eta) \in \mathbb{N}_0$ such that*

$$\|x^+ - \xi_n\|_X < \eta \quad \textit{for all } n \geq M(\eta) \textit{ and all } \xi_n \in \mathcal{X}_n.$$

Proof Assume the contrary. Then, there exists an $\eta > 0$ and a strictly increasing sequence $\{j_n\}_{n\in\mathbb{N}} \subset \mathbb{N}_0$ such that for any j_n there is a $\zeta_{j_n} \in \mathcal{X}_{j_n}$ with $\|x^+ - \zeta_{j_n}\|_X \geq \eta$. Without loss of generality we may assume that $j_n = n$. Indeed, if this not true then there is one j_n such that $j_{n-1} \neq j_n - 1$. However, the predecessor $\zeta_{j_n-1} \in \mathcal{X}_{j_n-1}$ of ζ_{j_n} satisfies also $\|x^+ - \zeta_{j_n-1}\|_X > \|x^+ - \zeta_{j_n}\|_X \geq \eta$, see (3.15), and we can add $j_n - 1$ to $\{j_n\}_{n\in\mathbb{N}}$.

So, for any $n \in \mathbb{N}_0$ we can assume existence of $\zeta_n \in \mathcal{X}_n$ with $\|x^+ - \zeta_n\|_X \geq \eta$. Without loss of generality we may moreover assume, for any n, that ζ_{n+1} is a predecessor of ζ_n. Otherwise consider $\widetilde{\zeta}_n$, the actual predecessor of ζ_{n+1}. By (3.15), $\|x^+ - \widetilde{\zeta}_n\|_Y > \|x^+ - \zeta_{n+1}\|_Y \geq \eta$ and we can replace ζ_n by $\widetilde{\zeta}_n$ and even $\zeta_0, \ldots, \zeta_{n-1}$ by the respective predecessors of $\widetilde{\zeta}_n$.

Thus, the sequence $\{\zeta_n\}_n$ originates from a run of `REGINN` with a modified rule for picking m_n: in (1.7) replace the less-than sign by the less-than-or-equal sign. Since this modification of `REGINN` does not alter its convergence, Lemma 3.8 applies and $\{\zeta_n\}_n$ converges to x^+ contradicting $\|x^+ - \zeta_n\|_X \geq \eta$ for all n. □

Now we are able to verify strong convergence.

Proof of Theorem 3.7 Let $\{\delta_j\}_{j\in\mathbb{N}}$ be a zero sequence. Assume first that $N(\delta_j) = n$ as $j \to \infty$. By Lemma 3.11 we may, without loss of generality, assume that $x_n^{\delta_j}$ converges to an element ξ_n of $\mathcal{X}_n$ as $j \to \infty$. In view of (3.1) we conclude that $F(\xi_n) = y$. Since $\mathcal{X}_n \subset B_r(x^+)$ we have $\xi_n = x^+$ due to the uniqueness of x^+ in $B_r(x^+)$.

Next, let $N(\delta_j)$ be bounded as $j \to \infty$. Then, $\{N(\delta_j)\}_j$ splits into convergent subsequences and we can argue as in the case of constant $N(\delta_j)$.

Finally, we consider $N(\delta_j) \to \infty$ as $j \to \infty$. For any $\eta > 0$ there is an $n = n(\eta)$ such that for every $\xi_n \in \mathcal{X}_n$ we have $\|x^+ - \xi_n\|_X \leq \eta/2$ (Lemma 3.12). Further, according to Lemma 3.11 and due to the finiteness of $\mathcal{X}_n$ there is a $J(\eta) \in \mathbb{N}$ such that for any $j \geq J(\eta)$ we can find a $\xi_n(j) \in \mathcal{X}_n$ satisfying $\|\xi_n(j) - x_n^{\delta_j}\|_X < \eta/2$. Thus, for $j \geq J(\eta)$ so large that $N(\delta_j) > n(\eta)$ we obtain from monotonicity the bound

$$\|x^+ - x_{N(\delta_j)}^{\delta_j}\|_X < \|x^+ - x_n^{\delta_j}\|_X \leq \|x^+ - \xi_n(j)\|_X + \|\xi_n(j) - x_n^{\delta_j}\|_X < \eta$$

for a suitably chosen $\xi_n(j) \in \mathcal{X}_n$. □

Remark 3.13 Without uniqueness of x^+ one can prove that $\{x_{N(\delta)}^{\delta}\}$ splits into convergent subsequences as $\delta \to 0$, each of which converges to a solution of $F(x) = y$. However, the arguments are more involved, compare [10].

Remark 3.14 Convergence of $\{x_{N(\delta)}^{\delta}\}$ with (sub-optimal) rates has been shown under the usual abstract smoothness (source) conditions and under restrictions on the nonlinearity being stronger than (3.4), see [19] for some linear inner regularizations and [21] for the cg-method as inner regularization. The next challenge, of course, is to explore how far convergence rates results can be obtained in our general setting of this paper. To master this challenge fresh ideas are needed.

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

$$\|g - Tf_m\|_Y = \min\left\{\|g - Tf\|_Y \,\middle|\, f \in X,\ f - f_0 \in U_m\right\}, \quad m \geq 1,$$

where U_m is the mth Krylov space,

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

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 [20], 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,

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

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

$$\langle g, Tf_m\rangle_Y \overset{(A.2)}{=} \|Tf_m\|_Y^2.$$

To establish (2.1) we validate that $Tf_m \neq 0$ under $T^*g \neq 0$. Assume $Tf_m = 0$ then $\|g\|_Y = \|r_m\|_Y \leq \|r_k\|_Y \leq \|g\|$ for any $k = 0, \ldots, m$. So, $f_k = 0$, $k = 0, \ldots, m$, but f_1 is a non-zero multiple of T^*g and cannot be zero. Thus, (2.1) holds true for cg.

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 [20]. 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

$$f_{m+1} = f_m + \lambda_m T^* r_m \quad \text{where} \qquad r_m = g - Tf_m \text{ and}$$
$$\lambda_m = \begin{cases} \frac{\|T^* r_m\|_X^2}{\|TT^* r_m\|_Y^2} & : T^* r_m \neq 0, \\ \|T\|^{-2} & : \text{otherwise.} \end{cases}$$

We first validate monotonicity of the residuals:

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

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

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

Whence (A.3) holds true.

Let $f_0 = 0$. Then,

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

leading to

$$\|Tf_m\|_Y^2 \leq 2\langle Tf_m, g\rangle_Y. \tag{A.4}$$

By Cauchy–Schwarz inequality we deduce that

$$\|Tf_m\|_Y \leq 2\|g\|_Y$$

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

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

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

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

$$\lim_{m\to\infty} Tf_m = P_{\overline{\mathsf{R}(T)}}g \tag{A.5}$$

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

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

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

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

Let $T^*r_m \ne 0$. Then,

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

The latter inequality remains true for $T^*r_m = 0$ and, hence, implies

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

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

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

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

Property (2.1) follows from (A.4) as soon as we have verified that $Tf_m \neq 0, m \geq 1$. Assume $Tf_m = 0$. Then, $Tf_m = 0$ and $Tf_{m+1} = Tf_1$ which yields $Tf_{m+i} = Tf_i$, $i \in \mathbb{N}_0$. Especially, $Tf_{km} = 0, k \in \mathbb{N}$, contradicting (A.5) under the requirement $T^*g \neq 0$.

Appendix B: Proof of (3.2) for Landweber, steepest decent, implicit iteration and cg

We profit from results of Hämarik and Tautenhahn [7].

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

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

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 [8, formula (2.7)].

Observe that (B.1) proves the first part of assumption (3.2).

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

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

see [7, 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.2)

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

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

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

where $\Psi(t) = 2t$. Thus, we have established (3.2) with $C_{\mathrm{M}} = 1$ 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.2),

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

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) with $C_{\mathrm{M}} \leq \frac{\alpha_{\min}}{\alpha_{\max}}$ follows for implicit iteration.

B.3. cg-Method

We follow arguments by Hanke [10, Theorem 3.1]. Here (B.2) 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,[4]

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

The normalized polynomial $w_m/w_m(0)$ is denoted $p_m^{[2]}$ by Hanke [8]. 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

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

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

$$C_{\mathrm{M}} = \Psi(\gamma) + \frac{\|r_{m-1}\|_Y}{\|g\|_Y} \leq \Psi(1) + 1.$$

References

1. Bakushinsky, A.B.: The problem of the convergence of the iteratively regularized Gauss-Newton method. Comput. Math. Phys. **32**, 1353–1359 (1992)
2. Bakushinsky, A.B., Kokurin, M.Y.: Iterative Methods for Approximate Solution of Inverse Problems. Vol. 577 of Mathematics and its Applications. Springer, Dordrecht, The Netherlands (2004)
3. Brakhage, H.: On ill-posed problems and the method of conjugate gradients. In: Engl, H.W., Groetsch, C.W. (eds.) Inverse and Ill-Posed Problems. Vol. 4 of Notes and Reports in Mathematics in Science and Engineering, pp. 165–175, Boston. Academic Press, New York (1987)
4. Deuflhard, P.: Newton Methods for Nonlinear Problems. Vol. 35 of Springer Series in Computational Mathematics. Springer, Berlin (2004)
5. Deuflhard, P., Engl, H.W., Scherzer, O.: A convergence analysis of iterative methods for the solution of nonlinear ill-posed problems under affinely invariant conditions. Inverse Problems **14**, 1081–1106 (1998)
6. Engl, H.W., Hanke, M., Neubauer, A.: Regularization of Inverse Problems. Vol. 375 of Mathematics and its Applications. Kluwer Academic Publishers, Dordrecht (1996)
7. Hämarik, U., Tautenhahn, U.: On the monotone error rule for parameter choice in iterative and continuous regularization methods. BIT **41**, 1029–1038 (2001)
8. Hanke, M.: Conjugate Gradient Type Methods for Ill-Posed Problems. Vol. 327 of Pitman Research Notes in Mathematics. Longman Scientific & Technical, Harlow, UK (1995)

[4] As m is less or equal to the ultimate stopping index of cg, see, e.g. [8, Sect. 2.1] or [20, Sect. 5.3], we have that $w_m(TT^*)g \neq 0$.

9. Hanke, M.: A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems. Inverse Problems **13**, 79–95 (1997)
10. Hanke, M.: Regularizing properties of a truncated Newton-CG algorithm for nonlinear inverse problems. Numer. Funct. Anal. Optim. **18**, 971–993 (1998)
11. Hanke, M., Neubauer, A., Scherzer, O.: A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. Numer. Math. **72**, 21–37 (1995)
12. Jin, Q., Tautenhahn, U.: On the discrepancy principle for some Newton type methods for solving nonlinear ill-posed problems. Numer. Math **111**, 509–558 (2009)
13. Kaltenbacher, B., Neubauer, A., Scherzer, O.: Iterative Regularization Methods for Nonlinear Ill-Posed Problems. Radon Series on Computational and Applied Mathematics. de Gruyter, Berlin (2008)
14. Kirsch, A.: An Introduction to the Mathematical Theory of Inverse Problems. Vol. 120 of Applied Mathematical Sciences. Springer-Verlag, New York (1996)
15. Lechleiter, A., Rieder, A.: Newton regularizations for impedance tomography: a numerical study. Inverse Problems **22**, 1887–1967 (2006)
16. Lechleiter, A., Rieder, A.: Newton regularizations for impedance tomography: convergence by local injectivity. Inverse Problems **24**, 065009 (2008)
17. Louis, A.K.: Inverse und schlecht gestellte Probleme. Studienbücher Mathematik, B.G. Teubner, Stuttgart, Germany (1989)
18. Rieder, A.: On the regularization of nonlinear ill-posed problems via inexact Newton iterations. Inverse Problems **15**, 309–327 (1999)
19. Rieder, A.: On convergence rates of inexact Newton regularizations. Numer. Math. **88**, 347–365 (2001)
20. Rieder, A.: Keine Probleme mit Inversen Problemen. Vieweg, Wiesbaden (2003)
21. Rieder, A.: Inexact Newton regularization using conjugate gradients as inner iteration. SIAM J. Numer. Anal. **43**, 604–622 (2005)
22. Scherzer, O.: The use of Morozov's discrepancy principle for Tikhonov regularization for solving nonlinear ill-posed problems. Computing **51**, 45–60 (1993)
23. Scherzer, O.: A convergence analysis of a method of steepest descent and a two-step algorithm for nonlinear ill-posed problems. Numer. Funct. Anal. Optim. **17**, 197–214 (1996)
24. Zeidler, E.: Nonlinear Functional Analysis and its Applications I: Fixed-Point Theorems. Springer-Verlag, New York (1993)