Karlsruhe Institute of Technology

Wave phenomena
analysis and numerics

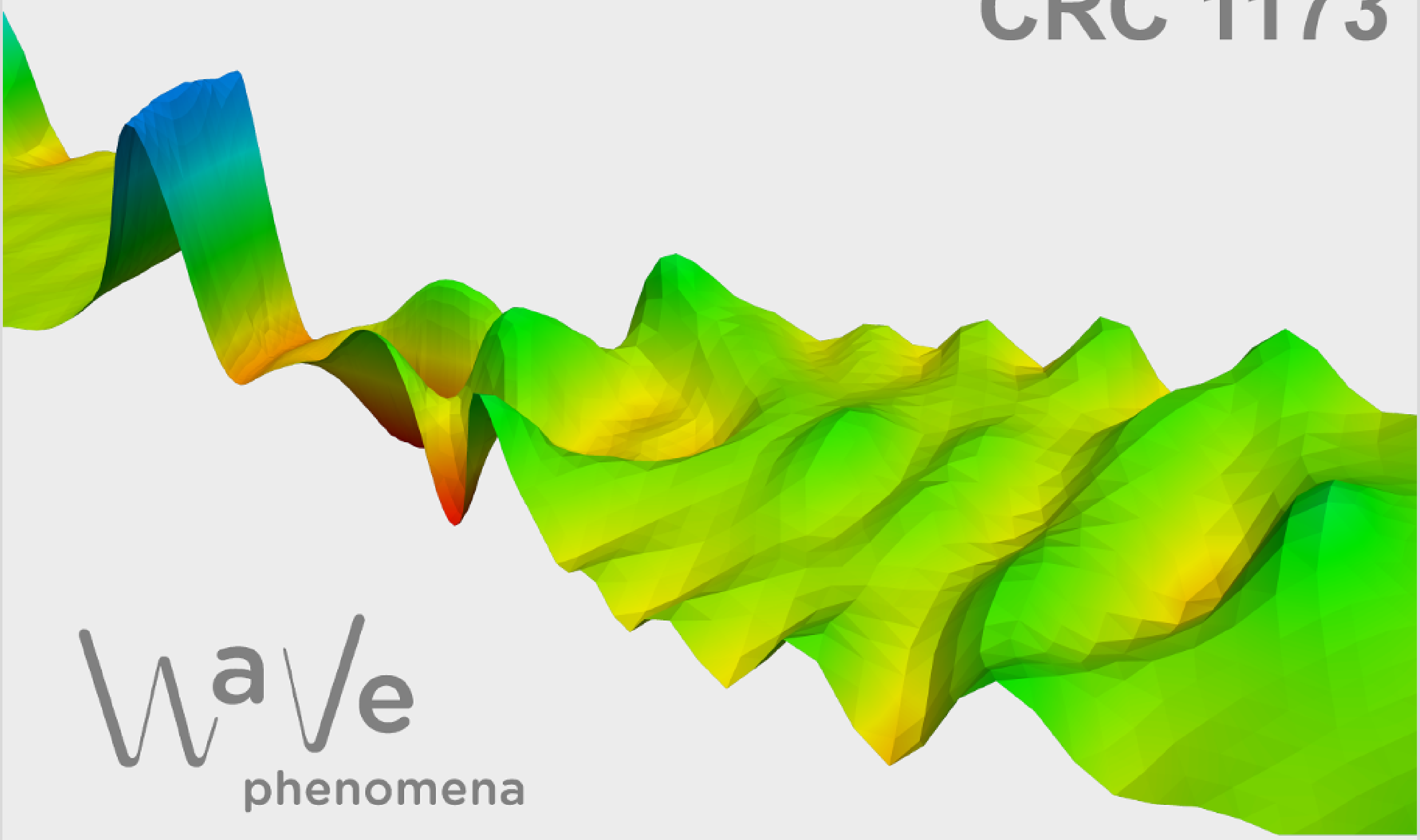
Inexact Newton regularizations with uniformly convex stability terms: a unified convergence analysis

Fábio Margotti, Marco Pauleti, Andreas Rieder

CRC Preprint 2023/12, April 2023

KARLSRUHE INSTITUTE OF TECHNOLOGY

CRC 1173



Wave phenomena

Participating universities

Universität Stuttgart

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

TU WIEN

Funded by

DFG

ISSN 2365-662X

INEXACT NEWTON REGULARIZATIONS WITH UNIFORMLY CONVEX STABILITY TERMS: A UNIFIED CONVERGENCE ANALYSIS

FÁBIO MARGOTTI, MARCO PAULETI, AND ANDREAS RIEDER

ABSTRACT. We present a unified convergence analysis of inexact Newton regularizations for nonlinear ill-posed problems in Banach spaces. These schemes consist of an outer (Newton) iteration and an inner iteration which provides the update of the current outer iterate. To this end the nonlinear problem is linearized about the current iterate and the resulting linear system is approximately (inexactly) solved by an inner regularization method. In our analysis we only rely on generic assumptions of the inner methods and we show that a variety of regularization techniques satisfies these assumptions. For instance, gradient-type and iterated-Tikhonov methods are covered. Not only the technique of proof is novel, but also the results obtained, because for the first time uniformly convex penalty terms stabilize the inner scheme.

1. INTRODUCTION

We are interested in solving the ill-posed inverse problem

$$F(x) = y\,, \tag{1}$$

where $F\colon \mathcal{D}(F) \subseteq X \longrightarrow Y$ is a continuous *nonlinear* operator acting between Banach spaces X and Y. We assume that only noisy data $y^\delta \in Y$ are available, satisfying

$$\|y - y^\delta\| \leq \delta, \tag{2}$$

where the *noise level* $\delta \geq 0$ is known.

Inexact Newton methods form a class of very efficient solvers for nonlinear inverse problems in many situations. In these methods, the forward operator F is linearized around the current iterate x_n and then the resulting linear system is (inexactly) solved to generate an update to x_n. This scheme was introduced by Dembo et al. [5] for well-posed problems, and was first adapted to ill-posed problems by Hanke [9, 10]. Rieder [26] generalized the ideas of Hanke by introducing the so-called `REGINN` algorithms (REGularization based on INexact Newton method).

The `REGINN` algorithm operates with two iterations. The *inner iteration* applies a regularization technique to stably approximate the solution of the linearized system and generate the update, which is added to the current iterate in the *outer iteration*. In addition to linear regularization methods, such as the stationary iterated-Tikhonov and Landweber methods, the solution of the linearized system can be approximated using nonlinear methods, such as the steepest descent or the conjugate gradient methods. Further, convergence, stability, and convergence rates have already been proven in Hilbert spaces [26, 27, 28]. In 2010, Lechleiter and Rieder [16] proposed a unified convergence analysis of

Date: April 12, 2023.

Key words and phrases. Nonlinear inverse problems; Inexact Newton regularization; Convex optimization; Bregman distances, Banach spaces.

The work of the third author was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project-ID 258734477 - SFB 1173.

the REGINN algorithm in Hilbert spaces based on generic properties of the sequence used in the inner iteration.

Many inverse problems are naturally formulated in Banach spaces, such as parameter identification tasks for partial differential equations. Moreover, sparseness of the solution can be forced by penalization in an appropriate Banach space [4]. Also, Banach spaces may be suitable for handling special types of noise in the data, for instance, impulsive noise as in [3]. On the other hand, the convergence analysis is complicated in Banach spaces, in part because the iterations usually operate in the dual space.

The first convergence results of REGINN in Banach spaces were proved by Jin [11], where a Landweber-like method was used in the inner iteration. Stability and regularization properties were achieved for the first time by Margotti et al. [23] for a Kaczmarz version of REGINN-Landweber. After that, different versions, using, e.g., the non-stationary iterated-Tikhonov and gradient-like methods appeared in [20, 22, 24].

All versions of the inexact Newton method cited above work in reflexive spaces and employ the subdifferential of the norm, called *duality mapping* (see (12) below), to link the primal to the dual spaces, but see [25] for a version of REGINN in L^∞. The properties of the duality mapping strongly depend on the smoothness and convexity properties of the space. Generally speaking, the smoother and more convex the norm of the space, the better the duality operator will behave. In [12], Jin investigated a version of REGINN-Landweber, replacing the duality mapping by the subdifferential of a p-convex functional. This interesting idea replaces the smoothness and convexity properties of the Banach space with the properties of the functional used. In addition, it allows us the use of more appropriate functionals than the norm of the Banach space, such as the combination of the norm with the seminorm of Total Variation or with the L^1-norm. Gu and Han [8] employed the same idea to define a p-convex two-points gradient method as inner iteration.

Inexact Newton methods penalized with more general functionals are very rare. In [13] and [21], variations of the Levenberg-Marquardt method are investigated which are penalized by the Bregman distance with respect to uniformly convex functional. To the best of our knowledge, there is no version of REGINN that uses regularization techniques penalized by uniformly convex functionals in the inner iteration. We will do exactly this in the present paper.

Main contributions of this work. We present a convergence analysis of the algorithm REGINN in reflexive Banach spaces with general uniformly convex penalty terms. Moreover, inspired by [16], we assume that the sequence generated in the inner iteration satisfies some generic conditions and use them to prove convergence and stability properties. Further, we validate that several regularization techniques satisfy the required conditions and are therefore included in the convergence analysis. Among the regularization techniques used to generate the inner iteration we will consider versions of gradient-type methods, iterated-Tikhonov methods (stationary and non-stationary, with *a priori* and *a posteriori* choices of regularization parameters), mixed gradient-Tikhonov methods, and the Tikhonov-Phillips method.

This paper is organized as follows: In the next section, we survey results concerning the theory of convex analysis and geometry of Banach spaces, which will be important for the full comprehension of the development of the following sections. Section 3 presents the version of algorithm REGINN whose convergence analysis is then given in Section 4, where the generic assumptions on the inner iteration are introduced and the main theoretical results are proved. Section 5, in turn, presents concrete examples of regularization methods, which satisfy the required assumptions, and therefore can be implemented as inner iteration. Section 6 is devoted to the analysis of the Tikhonov-Phillips method which, unfortunately, requires a separate treatment as it cannot be shown to satisfy the generic assumptions of Section 4. Finally, in Appendix A we simultaneously prove the stability of gradient and mixed-gradient Tikhonov methods.

2. Convex functionals and Bregman distances

For details on the material discussed in this section, we refer the reader to the textbooks [2, 6, 30]. Unless the contrary is explicitly stated, we always consider X a *real* Banach space.

The *effective domain* of the convex functional $f\colon X \longrightarrow \overline{\mathbb{R}} := (-\infty, \infty]$ is defined as $\mathrm{Dom}\,(f) := \{x \in X \mid f(x) < \infty\}$. The set $\mathrm{Dom}\,(f)$ is always convex, and we call f *proper* if $\mathrm{Dom}\,(f) \neq \emptyset$.

The functional f is said to be *uniformly convex* if there exists a non-decreasing function $\Theta\colon\ [0,\infty) \longrightarrow [0,\infty]$, vanishing only at zero, such that

$$f(\lambda x + (1-\lambda)y) + \lambda(1-\lambda)\Theta\left(\|x-y\|\right) \leq \lambda f(x) + (1-\lambda) f(y)\,, \tag{3}$$

for all $\lambda \in (0,1)$ and all $x, y \in \mathrm{Dom}\,(f)$. If f is uniformly convex, its *modulus of convexity* is the function $\varphi\colon\ [0,\infty) \longrightarrow [0,\infty]$ defined by

$$\varphi(t) := \inf\left\{\frac{\lambda f(x) + (1-\lambda) f(y) - f(\lambda x + (1-\lambda)y)}{\lambda(1-\lambda)} \;\middle|\; \begin{array}{c} x, y \in \mathrm{Dom}\,(f)\,, \|x-y\| \geq t, \\ \lambda \in (0,1) \end{array}\right\}.$$

Note that $\varphi \geq \Theta$. Further, the functional f is $p-$*convex*, $p \geq 2$, if there exists a constant $\beta > 0$ such that $\varphi(t) \geq \beta t^p$. In case $p = 2$, the functional f is called *strongly convex*.

The functional f is (sequentially) *lower semi-continuous* (l.s.c.) if, for any sequence $(x_k) \subseteq X$ satisfying $x_k \longrightarrow x$, the inequality $f\,(x) \leq \liminf_{k\to\infty} f\,(x_k)$ holds. It is called *weakly lower semi-continuous* (w.l.s.c.) if the last property holds true with $x_k \longrightarrow x$ replaced by $x_k \rightharpoonup x$. The concepts of l.s.c. and w.l.s.c. coincide if f is convex.

Let f be a proper, l.s.c., and uniformly convex functional with modulus of convexity φ. As $\inf(\emptyset) = \infty$, it may happen that $\varphi(t_0) = \infty$. Throughout this text, we will adopt the assumption that all considered proper, l.s.c., and uniformly convex functionals f are such that $\mathrm{Dom}\,(f)$ is an unbounded subset of X. In that case, $\varphi(\,\cdot\,) < \infty$. Moreover, it is possible to prove (see [29, 30]) that φ has the following properties: 1. $\varphi(t) = 0 \Leftrightarrow t = 0$; 2. φ is convex, continuous, and increasing; 3. $t \mapsto \varphi(t)/t^2$ is non-decreasing on $(0,\infty)$ and $\liminf_{t\to\infty}(\varphi(t)/t^2) > 0$. Most of the results presented here can be obtained without the extra assumption adopted on $\mathrm{Dom}\,(f)$, but we will make this assumption for simplicity. Also, the main functionals f that we consider during applications share this property.

The *subdifferential* of a functional $f\colon X \longrightarrow \overline{\mathbb{R}}$ is the point-to-set mapping $\partial f\colon X \rightrightarrows X^*$ defined by

$$\partial f\,(x) := \{x^* \in X^* \mid f\,(x) + \langle x^*, y - x\rangle \leq f\,(y) \;\text{ for all } y \in X\}\,. \tag{4}$$

The effective domain of ∂f is the set $\mathrm{Dom}\,(\partial f) := \{x \in X \mid \partial f\,(x) \neq \emptyset\}$. It is clear that the inclusion $\mathrm{Dom}\,(\partial f) \subseteq \mathrm{Dom}\,(f)$ holds whenever f is proper.

The definition of the subdifferential readily yields that $0 \in \partial f\,(x)$ if and only if x minimizes f. Further, if $f, g\colon X \longrightarrow \overline{\mathbb{R}}$ are convex and there is a point $x \in \mathrm{Dom}\,(f) \cap \mathrm{Dom}\,(g)$ where either f or g is continuous, then $\partial\,(f+g) = \partial f + \partial g$. It is evident that $\partial(\lambda f) = \lambda \partial f$ for all $\lambda > 0$.

Assume f is proper. Then, choosing elements $x, y \in X$ with $y \in \mathrm{Dom}\,(\partial f)\,$, we define the *Bregman distance* between x and y in the direction of $\xi \in \partial f(y)$ as

$$\Delta_\xi f(x,y) := f(x) - f(y) - \langle \xi, x - y\rangle\ . \tag{5}$$

Obviously, $\Delta_\xi f(y,y) = 0$ and, since $\xi \in \partial f(y)$, it additionally holds that $\Delta_\xi f(x,y) \geq 0$. Moreover, it is straightforward to prove the *three points identity*:

$$\Delta_{\xi_1} f(x_3, x_1) - \Delta_{\xi_2} f(x_3, x_2) = -\Delta_{\xi_2} f(x_1, x_2) + \langle \xi_2 - \xi_1, x_3 - x_1 \rangle \tag{6}$$

for all $x_3 \in \mathrm{Dom}(f)$, $x_1, x_2 \in \mathrm{Dom}(\partial f)$, $\xi_1 \in \partial f(x_1)$, and $\xi_2 \in \partial f(x_2)$. Further, the functional $\Delta_\xi f(\,\cdot\,, y) : X \longrightarrow \overline{\mathbb{R}}$ is convex and it becomes strictly convex whenever f is strictly convex. In this last case, $\Delta_\xi f(x,y) = 0$ if and only if $x = y$.

The *conjugate* of $f\colon X \longrightarrow \overline{\mathbb{R}}$ is the functional $f^*\colon X^* \longrightarrow [-\infty, \infty]$ defined by

$$f^*(x^*) = \sup_{x \in X} \{\langle x^*, x\rangle - f(x)\}.$$

If f is proper, convex, and l.s.c., so is f^*. From the definition follows that $f \leq g$ implies $f^* \geq g^*$. If $g(x) = p^{-1}\|x\|^p$, with $p > 1$, then $g^*(x^*) = (p^*)^{-1}\|x^*\|^{p^*}$, where $1/p + 1/p^* = 1$.

Similarly, for $h\colon [0,\infty) \longrightarrow [0,\infty]$ satisfying $h(0) = 0$, we define its *pseudo-conjugate* $h^\#\colon [0,\infty) \longrightarrow [0,\infty]$ by

$$h^\#(t) = \sup_{s \geq 0} \{ts - h(s)\}.$$

If φ is the modulus of convexity of f, then $\varphi^\#$ is called *modulus of smoothness* of f^*, see [30, Cor. 3.5.4].

Proposition 2.1. *Let f be a proper, l.s.c., and uniformly convex functional. If φ is the modulus of convexity of f, then the function $g(t) = \varphi^\#(t)/t$ defined on $(0,\infty)$ is positive, non-decreasing, and continuous with*

$$\lim_{t \to 0^+} \frac{\varphi^\#(t)}{t} = 0 \quad \textit{and} \quad \lim_{t \to \infty} \frac{\varphi^\#(t)}{t} = \infty. \tag{7}$$

Proof. Since f is uniformly convex, its conjugate f^* is uniformly smooth, [30, Thm. 3.5.5] and the left limit in (7) follows from the definition of uniform smoothness. Moreover, since $\varphi^\#$ is convex and $\varphi^\#(0) = 0$, the function $t \mapsto \varphi^\#(t)/t$ is non-decreasing on $(0,\infty)$. Thus, the limit $L \in [0,\infty]$ of $\varphi^\#(t)/t$ as $t \longrightarrow \infty$ exists. If $L < \infty$, there is some $\overline{s} \in \mathbb{R}$ such that $\overline{s} > L$. From the definition of $\varphi^\#$ we have that

$$\varphi^\#(t) \geq t\,\overline{s} - \varphi(\overline{s}), \text{ for all } t > 0.$$

Dividing by $t > 0$, and letting $t \longrightarrow \infty$, we get

$$L = \lim_{t \to \infty} \frac{\varphi^\#(t)}{t} \geq \overline{s} > L.$$

The above contradiction yields the limit on the right in (7). Now, $\varphi^\#(t) = 0$ if and only if $t = 0$, see [30, Lem. 3.3.1 (iii)], which implies $g(t) > 0$ for $t > 0$. Finally, $\varphi^\#(t) < \infty$ for $t \geq 0$, see [30, Cor. 3.5.4], and therefore $\mathrm{Dom}(\varphi^\#) = [0,\infty)$. Since $\varphi^\#$ is a proper, l.s.c., and convex function, $\varphi^\#$ is continuous on $\mathrm{int}(\mathrm{Dom}(\varphi^\#)) = (0,\infty)$. □

Let X be reflexive and $f\colon X \longrightarrow \overline{\mathbb{R}}$ be proper, convex, and l.s.c. For $x \in X$ and $x^* \in X^*$ the following equivalences hold

$$x^* \in \partial f(x) \Leftrightarrow \langle x^*, x\rangle = f(x) + f^*(x^*) \Leftrightarrow x \in \partial f^*(x^*) \Leftrightarrow x \in \underset{z \in X}{\mathrm{argmin}}\, \{f(z) - \langle x^*, z\rangle\}. \tag{8}$$

Proposition 2.2. *If $f\colon X \longrightarrow \overline{\mathbb{R}}$ is uniformly convex with modulus φ, then,*

$$\varphi(\|x_1 - x_2\|) \leq \Delta_{\xi_2} f(x_1, x_2) \tag{9}$$

for all $x_1 \in X$, $x_2 \in \mathrm{Dom}\,(\partial f)$ *and* $\xi_2 \in \partial f\,(x_2)$. *Additionally, if the space* X *is reflexive,* $x_1 \in \mathrm{Dom}\,(\partial f)$ *and* $\xi_1 \in \partial f\,(x_1)$, *then*

$$\Delta_{\xi_2} f\,(x_1, x_2) \le \varphi^{\#}\,(\|\xi_1 - \xi_2\|) \tag{10}$$

and

$$2\varphi\,(\|x_1 - x_2\|) \le \langle \xi_1 - \xi_2, x_1 - x_2\rangle\,. \tag{11}$$

Proof. For a proof of inequality (9), see, e.g., [18, ineq. (7)].

To prove (10), define $g(z) := \Delta_{\xi_1} f(z, x_1)$. Then, we have that $\partial g(z) = \partial f(z) - \xi_1$, and since $\xi_2 \in \partial f(x_2)$, we have $\xi_2 - \xi_1 \in \partial g(x_2)$. From (8),

$$g^*(\xi_2 - \xi_1) = \langle \xi_2 - \xi_1, x_2\rangle - g(x_2) = \Delta_{\xi_2} f(x_1, x_2) + \langle \xi_2 - \xi_1, x_1\rangle\,.$$

Define $h(z) = \varphi(\|z - x_1\|)$. Then, from (9) we obtain $g \ge h$. Thus, $g^*(z^*) \le h^*(z^*) \le \varphi^{\#}(\|z^*\|) + \langle z^*, x_1\rangle$ for all $z^* \in X^*$. Therefore,

$$\Delta_{\xi_2} f(x_1, x_2) + \langle \xi_2 - \xi_1, x_1\rangle = g^*(\xi_2 - \xi_1) \le h^*(\xi_2 - \xi_1) \le \varphi^{\#}(\|\xi_2 - \xi_1\|) + \langle \xi_2 - \xi_1, x_1\rangle\,,$$

which proves (10).

Finally, (11) follows from (9) using $\Delta_{\xi_2} f(x_1, x_2) + \Delta_{\xi_1} f(x_2, x_1) = \langle \xi_1 - \xi_2, x_1 - x_2\rangle$. ☐

We fix $p > 1$ and define the point-to-set function $J_p \colon X \rightrightarrows X^*$ as

$$J_p\,(x) := \{x^* \in X^* \mid \langle x^*, x\rangle = \|x^*\|\,\|x\| \ \text{ and } \ \|x^*\| = \|x\|^{p-1}\}\,, \tag{12}$$

which we call *duality mapping*. The Hahn-Banach theorem shows that $J_p\,(x) \neq \emptyset$ for all $x \in X$. Moreover, one can prove that $J_p = \partial g$, where $g(x) := p^{-1}\|x\|^p$. We call the Banach space X *smooth* if the functional g is Gâteaux-differentiable in X.[1] In this case, $J_p \colon X \longrightarrow X^*$ is single-valued and $J_p(x) = \{\nabla g(x)\}$. Similarly, the Banach space X is called *locally uniformly smooth* if the functional g is Fréchet-differentiable in X, in which case, J_p is continuous. Important examples of locally uniformly smooth Banach spaces are the Lebesgue spaces $L^p\,(\Omega)$, the Sobolev spaces $W^{n,p}\,(\Omega)$, $n \in \mathbb{N}$, and the space of $p-$summable sequences $\ell^p\,(\mathbb{R})$, for $1 < p < \infty$.

Let $\Omega \subseteq X$ be a nonempty set. The *annihilator* of Ω is defined by

$$\Omega^{\perp} := \{\xi \in X^* \mid \langle \xi, x\rangle = 0\,, \ \text{for all } x \in \Omega\}\,.$$

It is easy to show that $\Omega^{\perp}$ is a subspace of X^* and that $\Omega_1 \subseteq \Omega_2 \Rightarrow \Omega_2^{\perp} \subseteq \Omega_1^{\perp}$. Moreover, if $A \colon X \longrightarrow Y$ is a bounded linear operator acting between Banach spaces and $A^* \colon Y^* \longrightarrow X^*$ is the adjoint operator, then $\mathcal{R}(A^*) \subseteq \mathcal{N}(A)^{\perp}$.

3. Inexact Newton method REGINN

Algorithm REGINN, as proposed in [26], linearizes the nonlinear equation $F(x) = y^\delta$ around the current iterate x_n and then solves the linearized system inexactly in the so-called *inner iteration* to generate an increment, which is added to the current iterate in the *outer iteration* to obtain an update. More precisely, assume that F is Fréchet-differentiable and define $A_n := F'(x_n)$ and $b_n^\delta := y^\delta - F(x_n)$. Now, the inner iteration $(s_{n,k})_k$ is generated by applying a regularization technique to approximate a solution of the linearized system $A_n s = b_n^\delta$. The inner iteration stops at the index $k = \widehat{k}_n$ defined by

$$\widehat{k}_n := \inf\,\{k \in \mathbb{N} \mid \|A_n s_{n,k} - b_n^\delta\| < \mu_n \|b_n^\delta\|\}\,, \tag{13}$$

[1]The differentiability properties of this functional are independent of the particular choice of $p > 1$.

where $\mu_n \in (0,1)$ is a pre-defined constant. The next outer iterate is then $x_{n+1} := x_n + s_{n,\widehat{k}_n}$. Finally, the outer iteration is terminated by the *discrepancy principle*, i.e., we fix $\tau > 1$ and finish the iteration at the index $n = n_\delta$ where

$$n_\delta := \inf \{ n \in \mathbb{N} \mid \|F(x_n) - y^\delta\| \le \tau\delta \} . \tag{14}$$

In what follows, we describe the version of algorithm `REGINN` we shall consider in this work. First of all, we assume that X is a reflexive Banach space and fix a proper, l.s.c., and uniformly convex functional $f \colon X \longrightarrow \overline{\mathbb{R}}$.

We start the outer iteration with initial guesses $x_0 \in \mathcal{D}(F) \cap \mathrm{Dom}\,(\partial f)$ and $\xi_0 \in \partial f(x_0)$. Assuming the iterates $x_n \in \mathcal{D}(F) \cap \mathrm{Dom}\,(\partial f)$ and $\xi_n \in \partial f(x_n)$ are already defined, we define A_n and b_n^δ as above and produce the inner iteration as follows: set $x_{n,0} := x_n$ and $\xi_{n,0} := \xi_n$. Now, assuming the $k-$th inner iterates $x_{n,k}$ and $\xi_{n,k} \in \partial f(x_{n,k})$ are already computed, choose a *direction* $\omega_{n,k} \in X^*$, a *step-size* $\lambda_{n,k} > 0$, and define $\xi_{n,k+1} := \xi_{n,k} - \lambda_{n,k}\,\omega_{n,k}$. Now we set $x_{n,k+1} := \nabla f^*(\xi_{n,k+1})$. Because f is uniformly convex, the conjugate function f^* is Fréchet-differentiable in X^* and ∇f^* is uniformly continuous, see [30, Thm. 3.5.10]. Therefore, the next inner iterate $x_{n,k+1}$ is well defined and since X is reflexive, $\xi_{n,k+1} \in \partial f(x_{n,k+1})$, see (8). Moreover, using (8) once more, one sees that $x_{n,k}$ may be computed by solving the following minimization problem

$$x_{n,k} = \operatorname*{argmin}_{x \in X} \{ f(x) - \langle \xi_{n,k}, x \rangle \} .$$

The inner iteration terminates with a generalization of the rule (13). More precisely, for $s_{n,k} := x_{n,k} - x_n$ choose $\mu_n \in (0,1)$ as well as $k_{\max,n} \in \mathbb{N} \cup \{\infty\}$. Then, $x_{n+1} := x_{n,k_n}$ (which is equivalent to $x_{n+1} = x_n + s_{n,k_n}$) and $\xi_{n+1} := \xi_{n,k_n}$, where

$$k_n := \min \{ \widehat{k}_n, k_{\max,n} \} . \tag{15}$$

Finally, the outer iteration (x_n) terminates at index $n = n_\delta$ according to the rule (14).

Remark 3.1. a) *According to rule (15) the inner iteration may stop at any index less than or equal to $\widehat{k}_n$ in (13).*

b) *The directions $\omega_{n,k}$ and the step-sizes $\lambda_{n,k}$ depend on the specific method used in the inner iteration. We give two examples. By taking $f(x) = p^{-1}\|x\|^p$, $p > 1$, and assuming that the spaces X and Y fulfill certain conditions, we have that $\partial f = J_p$ is the duality mapping, and the inner iteration of* `REGINN` *becomes*

$$\begin{aligned} \xi_{n,k+1} &= J_p(x_{n,k}) - \lambda_{n,k}\,\omega_{n,k} , \\ x_{n,k+1} &= J_{p^*}(\xi_{n,k+1}) , \end{aligned}$$

where $1/p + 1/p^* = 1$. Thus, if $\omega_{n,k} = A_n^* J_r(A_n s_{n,k} - b_n^\delta)$ for some $r > 1$, then our method reduces to the one studied in [20]. Moreover, in case $\omega_{n,k} = A_n^* J_r(A_n s_{n,k+1} - b_n^\delta)$, then, it is similar to the one analyzed in [22].

In particular, if X and Y are Hilbert spaces and $f(x) = \|x\|^2/2$, then ∂f is the identity operator and the inner iteration reduces to

$$x_{n,k+1} = x_{n,k} - \lambda_{n,k}\,\omega_{n,k} ,$$

which is essentially the original method introduced in [26].

Algorithm 1 describes the method in pseudocode.

```
INPUT: x_0, ξ_0, F, f, y^δ, δ, τ (as specified in the text)
OUTPUT: x_n satisfying ‖F(x_n) − y^δ‖ ≤ τδ .
n = 0;
While ‖F(x_n) − y^δ‖ > τδ Do
    [1] A_n = F'(x_n), b_n^δ = y^δ − F(x_n);
    [2] k = 0, x_{n,0} = x_n, ξ_{n,0} = ξ_n;
    [3] Choose k_{max,n} ∈ ℕ ∪ {∞} and μ_n ∈ (0,1);
    [4] Repeat
        [4.1] Choose ω_{n,k} ∈ X* and λ_{n,k} > 0;
        [4.2] ξ_{n,k+1} = ξ_{n,k} − λ_{n,k} ω_{n,k};
        [4.3] x_{n,k+1} = ∇f*(ξ_{n,k+1});
        [4.4] k = k + 1;
        Until ‖A_n(x_{n,k} − x_n) − b_n^δ‖ < μ_n ‖b_n^δ‖ or k = k_{max,n}
    [5] x_{n+1} = x_{n,k}, ξ_{n+1} = ξ_{n,k};
    [6] n = n + 1;
End While
```

ALGORITHM 1. Algorithm REGINN with convex penalization terms.

4. CONVERGENCE ANALYSIS

In the current section, we adopt some assumptions concerning the inner iteration and derive a unified convergence analysis of Algorithm 1. Later, in Section 5, we will show that many regularization methods satisfy these assumptions.

For the remaining of this article, we assume the following conditions related to the inverse problem (1), the spaces X and Y and the functional $f\colon X \longrightarrow \overline{\mathbb{R}}$:

Assumption 1. *The initial guesses $x_0 \in \mathcal{D}(F) \cap \mathrm{Dom}\,(\partial f)$ and $\xi_0 \in \partial f(x_0)$ are fixed and independent of the noise level $\delta \geq 0$. Let the following hold.*

(A.1) *There is a radius $\rho > 0$ such that $B_\rho(x_0) := \{x \in X \mid \|x - x_0\| < \rho\} \subseteq \mathcal{D}(F)$.*
(A.2) *f is a proper, l.s.c., and uniformly convex functional with modulus of convexity φ, see (3). Moreover, the effective domain $\mathrm{Dom}(f)$ is unbounded.*
(A.3) *There is some $x^\star \in \mathcal{D}(F) \cap \mathrm{Dom}\,(f)$ such that $F(x^\star) = y$ and $\Delta_{\xi_0} f(x^\star, x_0) < \varphi(\rho/2) < \infty$.*
(A.4) *Tangential cone condition (TCC): F is continuously Fréchet-differentiable in $B_\rho(x_0)$ and there is a constant $\eta \in [0, 1)$ such that*
$$\|F(x) - F(z) - F'(z)(x - z)\| \leq \eta \|F(x) - F(z)\| \quad \text{for all } x, z \in B_\rho(x_0).$$
(A.5) *There is a constant $M > 0$ such that $\|F'(x)\| \leq M$ for all $x \in B_\rho(x_0)$.*
(A.6) *The Banach spaces X and Y are reflexive and locally uniformly smooth, respectively.*

The critical of the above assumptions is the tangential cone condition (A.4) because it significantly restricts the structure of admissible nonlinearities. Yet, in different versions it is a notorious prerequisite for the convergence analysis of many regularization schemes, see, e.g., the corresponding papers cited in Section 1. Examples of nonlinear inverse problems which satisfy the TCC are reported in [7, 14, 15]. We emphasize that, by (A.6), the duality mapping $J_p\colon Y \longrightarrow Y^*$, $p > 1$, is single-valued and continuous, see text below (12).

We now define $A_n := F'(x_n)$, $b_n^\delta := y^\delta - F(x_n)$, $s_{n,k} := x_{n,k} - x_n$, $e_n := x^\star - x_n$, and fix some $r > 1$. Next, we formulate an assumption on the inner iteration of `REGINN`:

Assumption 2. *If the n-th iterate $x_n \in \mathcal{D}(F) \cap \mathrm{Dom}(\partial f)$ is well defined, let there exist a constant $c_0 \in (0,1)$ such that, for $k = 0, \dots, k_n - 1$, there is a nonzero vector $v_{n,k} \in X$ satisfying*

$$\Delta_{\xi_{n,k+1}}(x^\star, x_{n,k+1}) - \Delta_{\xi_{n,k}}(x^\star, x_{n,k}) \\ \leq \lambda_{n,k} \|v_{n,k}\|^{r-1} \left[\|A_n e_n - b_n^\delta\| - c_0 \|A_n s_{n,k} - b_n^\delta\| \right] . \tag{16}$$

From here on, if the context permits, we will write $\Delta_\xi\,(v,w)$ or even $\Delta\,(v,w)$ instead of $\Delta_\xi f\,(v,w)$. In particular, $\Delta(\cdot, x_{n,k}) = \Delta_{\xi_{n,k}}(\cdot, x_{n,k})$ and $\Delta(\cdot, x_n) = \Delta_{\xi_n}(\cdot, x_n)$.

Assumption 2 implies monotonicity of the inner iterations, as we now prove.

Theorem 4.1. *Let Assumptions 1 and 2 hold with $\eta < c_0$. Let $\tau > (1+\eta)/(c_0 - \eta)$. Assume that the iterates $x_0, x_1, \dots, x_n$ are well defined in $\mathcal{D}(F)$. Moreover, assume that $\|F(x_n) - y^\delta\| > \tau\delta$. Define*

$$\overline{\mu}_n := \eta + \frac{(1+\eta)\delta}{\|b_n^\delta\|} . \tag{17}$$

If $\mu_n \in (c_0^{-1}\overline{\mu}_n, 1)$ then, for all $k < k_n$,

$$(c_0 - \overline{\mu}_n/\mu_n)\lambda_{n,k}\|v_{n,k}\|^{r-1}\|A_n s_{n,k} - b_n^\delta\| \leq \Delta(x^\star, x_{n,k}) - \Delta(x^\star, x_{n,k+1}) , \tag{18}$$

and, consequently, the Bregman distances of the inner iterates are strictly decreasing, i.e.,

$$\Delta(x^\star, x_{n,k+1}) < \Delta(x^\star, x_{n,k}) , \ \textit{for all} \ \ k = 0, \dots, k_n - 1 . \tag{19}$$

Proof. First, observe that the lower bound on τ together with $\|b_n^\delta\| > \tau\delta$ implies that the interval for selecting μ_n is nonempty. Now, the TCC (Assumption 1, (A.4)) implies that $\|A_n e_n - b_n^\delta\| \leq \overline{\mu}_n \|b_n^\delta\|$. By $\|A_n s_{n,k} - b_n^\delta\| \geq \mu_n \|b_n^\delta\|$ for $k < k_n$, and by Assumption 2 we obtain (18), which in turn proves (19). ☐

Assumption 3. *Under Assumption 2 let there exist a non-decreasing function $h\colon (0,\infty) \longrightarrow (0,\infty)$ such that*

$$h(\|A_n s_{n,k} - b_n^\delta\|) \leq \lambda_{n,k}\|v_{n,k}\|^{r-1}$$

for $n = 0, \dots, n_\delta - 1$ and $k = 0, \dots, k_n - 1$.

Theorem 4.2. *Let Assumptions 1, 2, and 3 hold with $\eta < c_0$. Let $\tau > (1+\eta)/(c_0 - \eta)$. If $\mu_n \in (c_0^{-1}\overline{\mu}_n, 1)$ then, for $n = 0, \dots, n_\delta - 1$, the inner iteration terminates, i.e., the index $\widehat{k}_n$ is finite. Further, x_n is well defined for $n \leq n_\delta$ and belongs to $\mathcal{D}(F)$. Moreover, for $n < n_\delta$,*

$$(c_0 - \overline{\mu}_n/\mu_n) \sum_{k=0}^{k_n - 1} \lambda_{n,k}\|v_{n,k}\|^{r-1}\|A_n s_{n,k} - b_n^\delta\| \leq \Delta(x^\star, x_n) - \Delta(x^\star, x_{n+1}) , \tag{20}$$

and consequently, the Bregman distances of the outer iterates are strictly decreasing, i.e.,

$$\Delta(x^\star, x_{j+1}) < \Delta(x^\star, x_j) , \ \textit{for all} \ \ j = 0, \dots, n_\delta - 1 . \tag{21}$$

Especially, $x_{n,k} \in B_\rho(x_0) \subseteq \mathcal{D}(F)$, for all $n < n_\delta$ and $k \leq \widehat{k}_n$.

Proof. We argue by induction on n. We start with $x_0 \in \mathcal{D}(F)$ and suppose that the iterates $x_1, \dots, x_n \in \mathcal{D}(F)$ have already been computed for some $n < n_\delta$ satisfying (21) for $j = 0, \dots, n-1$. Let $\ell \in \mathbb{N}_0$ with $\ell \le \widehat{k}_n$. In view of (18),

$$(c_0 - \overline{\mu}_n/\mu_n) \sum_{k=0}^{\ell-1} \lambda_{n,k} \|v_{n,k}\|^{r-1} \|A_n s_{n,k} - b_n^\delta\| \\ \le \sum_{k=0}^{\ell-1} [\Delta(x^\star, x_{n,k}) - \Delta(x^\star, x_{n,k+1})] = \Delta(x^\star, x_{n,0}) - \Delta(x^\star, x_{n,\ell}) \le \Delta(x^\star, x_n)\,. \tag{22}$$

Since $\|A_n s_{n,k} - b_n^\delta\| \ge \mu_n \|b_n^\delta\|$ for $k < \widehat{k}_n$ and since the function h from Assumption 3 is non-decreasing, we infer that

$$h(\mu_n \|b_n^\delta\|)\, \mu_n\, \|b_n^\delta\| \ell \le \sum_{k=0}^{\ell-1} h(\|A_n s_{n,k} - b_n^\delta\|) \|A_n s_{n,k} - b_n^\delta\| \\ \le \sum_{k=0}^{\ell-1} \lambda_{n,k} \|v_{n,k}\|^{r-1} \|A_n s_{n,k} - b_n^\delta\|. \tag{23}$$

This bound together with (22) leads to

$$\ell \le \frac{\Delta(x^\star, x_n)}{(c_0 \mu_n - \overline{\mu}_n) h(\mu_n \|b_n^\delta\|) \|b_n^\delta\|} < \infty$$

and hence $\widehat{k}_n < \infty$. Plugging $\ell = k_n \le \widehat{k}_n$ into (22) we obtain (20), which, in turn, proves (21) for $j = 0, \dots, n$. By (19) we have, for $k \le k_n$,

$$\Delta(x^\star, x_{n,k}) < \Delta(x^\star, x_{n,k-1}) < \cdots < \Delta(x^\star, x_{n,0}) = \Delta(x^\star, x_n)\,. \tag{24}$$

But, from Assumption 1 (A.3) we get

$$\varphi(\|x^\star - x_0\|) \le \Delta(x^\star, x_0) < \varphi(\rho/2)\,,$$

which results in $\|x^\star - x_0\| < \rho/2$. From (24) and (21) we conclude that

$$\varphi(\|x^\star - x_{n,k}\|) \le \Delta(x^\star, x_{n,k}) < \Delta(x^\star, x_n) < \cdots < \Delta(x^\star, x_0) < \varphi(\rho/2).$$

Thus, $\|x^\star - x_{n,k}\| < \rho/2$ and $x_{n,k} \in B_\rho(x_0) \subseteq \mathcal{D}(F)$ for all $k \le k_n$. In particular, $x_{n+1} = x_{n,k_n} \in \mathcal{D}(F)$ and the inductive step is complete. □

From (20) we obtain for all $i \le n_\delta$ that

$$\sum_{n=0}^{i-1} \sum_{k=0}^{k_n-1} (c_0 - \overline{\mu}_n/\mu_n) \lambda_{n,k} \|v_{n,k}\|^{r-1} \|A_n s_{n,k} - b_n^\delta\| \le \sum_{n=0}^{i-1} [\Delta(x^\star, x_n) - \Delta(x^\star, x_{n+1})] \\ = \Delta(x^\star, x_0) - \Delta(x^\star, x_i)\,. \tag{25}$$

We now prove that the outer iteration terminates for noisy data ($\delta > 0$), i.e., $n_\delta < \infty$. To this end we must control the term $c_0\mu_n - \overline{\mu}_n$. Note that the restrictions on η and τ in the last theorem ensure $c_0^{-1}(\eta + (1+\eta)\tau^{-1}) < 1$. We thus fix a number $\mu_{\min}$ satisfying

$$c_0^{-1}(\eta + (1+\eta)\tau^{-1}) < \mu_{\min} < 1 \tag{26}$$

and restrict μ_n to the interval $(\mu_{\min}, 1)$. As $\mu_{\min} > c_0^{-1}\overline{\mu}_n$ for $n < n_\delta$, all the results of the previous theorems hold in this new situation.

Corollary 4.3. *Let $\delta > 0$ be fixed and adopt all the hypotheses of Theorem 4.2. Moreover, assume that $\mu_n \in (\mu_{\min}, 1)$ for all $n < n_\delta$ with $\mu_{\min}$ as in (26). Then, $n_\delta < \infty$.*

Proof. Let $i \le n_\delta$. For $n < n_\delta$ we have that $c_0 - \overline{\mu}_n/\mu_n > c_0 - (\eta + (1+\eta)\tau^{-1})\mu_{\min}^{-1} =: C_0 > 0$. Hence, from (25),

$$C_0 \sum_{n=0}^{i-1} \sum_{k=0}^{k_n-1} \lambda_{n,k} \|v_{n,k}\|^{r-1} \|A_n s_{n,k} - b_n^\delta\| \le \Delta(x^\star, x_0)\,. \tag{27}$$

Note that $\|b_n^\delta\| > \tau\delta$ and $k_n \ge 1$ for $n < n_\delta$. Since h is positive and non-decreasing, it follows from (23) with $\ell = k_n$ that

$$\begin{aligned} i\, C_0 h(\mu_{\min}\tau\delta)\mu_{\min}\tau\delta &\le i\, C_0 h(\mu_n \|b_n^\delta\|)\mu_n \|b_n^\delta\| k_n \\ &\le C_0 \sum_{n=0}^{i-1}\sum_{k=0}^{k_n-1} \lambda_{n,k}\|v_{n,k}\|^{r-1}\|A_n s_{n,k} - b_n^\delta\| \le \Delta(x^\star, x_0) < \infty\,, \end{aligned}$$

so that $i < \infty$ and, consequently, $n_\delta < \infty$. □

Corollary 4.4. *For those outer iterations satisfying $k_n = \widehat{k}_n$, we have*

$$\|b_{n+1}^\delta\| \le \Lambda \|b_n^\delta\|, \quad \textit{with} \quad 0 < \Lambda := \frac{\mu_{\max} + \eta}{1 - \eta},$$

where η is the constant in Assumption 1 (A.4), $\mu_{\min} < \mu_{\max} < 1$ and $\mu_n \in (\mu_{\min}, \mu_{\max})$. Additionally, if $\eta < c_0/(1 + 2c_0)$, then by choosing $\tau > (1 + \eta)/((1 - 2\eta)c_0 - \eta)$ we can further restrict $\mu_{\min}$ and $\mu_{\max}$ such that

$$c_0^{-1}\left(\eta + \frac{1+\eta}{\tau}\right) < \mu_{\min} < \mu_{\max} < 1 - 2\eta.$$

Consequently, $\Lambda < 1$, and accordingly

$$n_\delta \le \log_\Lambda\left(\frac{\tau\delta}{\|F(x_0) - y^\delta\|}\right) + 1,$$

whenever $k_n = \widehat{k}_n$ for $n = 0, \dots, n_\delta - 1$.

Proof. See, e.g., [20, Rem.6]. □

4.1. **Convergence without noise.** In the noise-free situation ($\delta = 0$) we cannot guarantee n_δ to be finite. However, we can prove summability of some important series, which yield then convergence of (x_n) to a solution of the inverse problem (1). In the sequel, we use the notation $b_n := y - F(x_n)$.

Corollary 4.5. *Under the hypotheses of Theorem 4.2 assume that $\delta = 0$, $n_\delta = \infty$, and $\mu_n \in (\mu_{\min}, 1)$ for all $n \in \mathbb{N}$, with $\mu_{\min}$ as in (26). Then,*

$$\sum_{n=0}^{\infty}\sum_{k=0}^{k_n-1} \lambda_{n,k}\|v_{n,k}\|^{r-1}\|A_n s_{n,k} - b_n\| < \infty \quad \textit{and} \quad \sum_{n=0}^{\infty} h(\mu_{\min}\|b_n\|)\mu_{\min}\|b_n\| < \infty\,. \tag{28}$$

Consequently, $F(x_n) \longrightarrow y$ as $n \longrightarrow \infty$.

Proof. Because $n_\delta = \infty$, inequality (27) holds true for all $i \in \mathbb{N}$. Letting $i \longrightarrow \infty$ we obtain the convergence of the first series in (28). In view of (23), $\mu_n > \mu_{\min}$, and $k_n \ge 1$, convergence of the second one follows from the first. Hence, $h(\mu_{\min}\|b_n\|)\mu_{\min}\|b_n\| \longrightarrow 0$ as $n \longrightarrow \infty$ which implies $\|F(x_n) - y\| = \|b_n\| \longrightarrow 0$ since h is positive and non-decreasing. □

For the proof of the above announced convergence result, we need a further assumption on the inner iteration.

Assumption 4. *Under Assumption 2 let there exist a sequence* $(\gamma_{n,k}) \subseteq X^*$ *such that for all* $n < n_\delta$ *and* $k < k_n$,

$$\xi_{n,k+1} = \xi_{n,k} - \lambda_{n,k}\left[A_n^* J_r(v_{n,k}) + \alpha_{n,k}\gamma_{n,k}\right],$$

where $\alpha_{n,k}\|\gamma_{n,k}\| \le c_1 \|v_{n,k}\|^{r-1}\|A_n s_{n,k} - b_n^\delta\|$, *with* $(\alpha_{n,k}) \subseteq [0,\infty)$ *and* $c_1 \ge 0$.

Theorem 4.6 (Convergence with exact data)**.** *Let* $\delta = 0$ *and adopt all the hypotheses of Theorem 4.2. Further, assume that* $\mu_n \in (\mu_{\min}, 1)$ *for* $n \in \mathbb{N}$, *with* $\mu_{\min}$ *as in (26). Under Assumption 4 there is some* $x_\infty \in \mathcal{D}(F)$ *satisfying* $F(x_\infty) = y$ *such that* `REGINN` *either terminates after* $m \in \mathbb{N}$ *iterations with* $x_m = x_\infty$ *or the sequence* (x_n) *converges:*

$$\lim_{n\to\infty} \Delta(x_\infty, x_n) = 0 \qquad \textit{and} \qquad \lim_{n\to\infty} \|x_\infty - x_n\| = 0\,. \tag{29}$$

Proof. If there is some m with $\|F(x_m) - y\| \le \tau\delta = 0$, then $F(x_m) = y$ and the outer iteration terminates with a solution of (1).

We now assume that $n_\delta = \infty$ and prove (29). Indeed, as $b_n \longrightarrow 0$ by Corollary 4.5, we can select a subsequence (x_{n_j}) such that, for each $j \in \mathbb{N}$,

$$\|b_{n_j}\| \le \|b_m\| \quad \text{for all } m \le n_j\,. \tag{30}$$

We first validate that (x_{n_j}) is a Cauchy sequence. In fact, fix natural numbers $m < \ell$. From the three points identity (6),

$$\varphi(\|x_{n_\ell} - x_{n_m}\|) \le \Delta(x_{n_\ell}, x_{n_m}) = [\Delta(x^\star, x_{n_m}) - \Delta(x^\star, x_{n_\ell})] + \langle \xi_{n_\ell} - \xi_{n_m}, x_{n_\ell} - x^\star\rangle\,. \tag{31}$$

We now estimate the last term on the right, relying on Assumption 4:

$$\begin{aligned}
|\langle \xi_{n_\ell} - \xi_{n_m}, x_{n_\ell} - x^\star\rangle| &= \left|\sum_{n=n_m}^{n_\ell-1} \langle \xi_{n+1} - \xi_n, e_{n_\ell}\rangle\right| = \left|\sum_{n=n_m}^{n_\ell-1} \langle \xi_{n,k_n} - \xi_{n,0}, e_{n_\ell}\rangle\right| \\
&= \left|\sum_{n=n_m}^{n_\ell-1}\sum_{k=0}^{k_n-1} \langle \xi_{n,k+1} - \xi_{n,k}, e_{n_\ell}\rangle\right| \\
&= \left|\sum_{n=n_m}^{n_\ell-1}\sum_{k=0}^{k_n-1} -\lambda_{n,k}\left[\langle J_r(v_{n,k}), A_n e_{n_\ell}\rangle + \alpha_{n,k}\langle \gamma_{n,k}, e_{n_\ell}\rangle\right]\right| \\
&\le \sum_{n=n_m}^{n_\ell-1}\sum_{k=0}^{k_n-1} \lambda_{n,k}\left[\|v_{n,k}\|^{r-1}\|A_n e_{n_\ell}\| + \alpha_{n,k}\|\gamma_{n,k}\|\|e_{n_\ell}\|\right].
\end{aligned}$$

Further, in view of (30), for $n < n_\ell$ and $k < k_n$,

$$\begin{aligned}
\|A_n e_{n_\ell}\| &\le \|F(x_{n_\ell}) - F(x_n) - F'(x_n)(x_{n_\ell} - x_n)\| \\
&\qquad\qquad + \|F(x^\star) - F(x_n) - F'(x_n)(x^\star - x_n)\| + \|b_{n_\ell}\| \\
&\le \eta\|F(x_{n_\ell}) - F(x_n)\| + \eta\|F(x^\star) - F(x_n)\| + \|b_{n_\ell}\| \\
&\le 2\eta\|b_n\| + (\eta+1)\|b_{n_\ell}\| \le (3\eta+1)\|b_n\| \le \frac{3\eta+1}{\mu_{\min}}\|A_n s_{n,k} - b_n\|\,.
\end{aligned} \tag{32}$$

According to Theorem 4.2, the sequence $(x_{n,k})$ is uniformly bounded in n and k and so is (e_n). From Assumption 4 follows the existence of a constant $C_1 \ge 0$ such that

$$\alpha_{n,k}\|\gamma_{n,k}\|\|e_{n_\ell}\| \le C_1\|v_{n,k}\|^{r-1}\|A_n s_{n,k} - b_n\|\,.$$

Putting all together, we get

$$|\langle \xi_{n_\ell} - \xi_{n_m}, x_{n_\ell} - x^\star\rangle| \le \left[C_1 + \frac{(3\eta+1)}{\mu_{\min}}\right] \sum_{n=n_m}^{n_\ell - 1} \sum_{k=0}^{k_n-1} \lambda_{n,k} \|v_{n,k}\|^{r-1} \|A_n s_{n,k} - b_n\| . \tag{33}$$

From (28), we conclude that the last term on the right of (31) converges to zero as $m, \ell \longrightarrow \infty$. Moreover, the term in the brackets in (31) also tends to zero as $m, \ell \longrightarrow \infty$ for the following reason: the sequence $(\Delta(x^\star, x_n))_n$ converges, since it is monotonically decreasing, see (21), and bounded from below by zero. As a result, $\varphi(\|x_{n_m} - x_{n_\ell}\|) \longrightarrow 0$ and, consequently, $\|x_{n_m} - x_{n_\ell}\| \longrightarrow 0$ as $m, \ell \longrightarrow \infty$, proving that (x_{n_j}) is a Cauchy sequence. Since X is a complete space, there is some $x_\infty \in X$ such that $x_{n_j} \longrightarrow x_\infty$. By monotonicity of $(\Delta(x^\star, x_n))_n$,

$$\varphi(\|x^\star - x_{n_j}\|) \le \Delta(x^\star, x_{n_j}) \le \cdots \le \Delta(x^\star, x_0) < \varphi(\rho/2) ,$$

giving $\|x^\star - x_\infty\| \le \rho/2$. Since $\|x^\star - x_0\| < \rho/2$, we have $x_\infty \in B_\rho(x_0) \subseteq \mathcal{D}(F)$. From $\|b_{n_j}\| \longrightarrow 0$, we conclude that $F(x_\infty) = y$.

Finally, for every $j \in \mathbb{N}$, the functional $\Delta(\,\cdot\,, x_{n_j})$ is l.s.c. and by (31) and (33),

$$\lim_{j\to\infty} \Delta(x_\infty, x_{n_j}) \le \lim_{j\to\infty} \left[\liminf_{m\to\infty} \Delta(x_{n_m}, x_{n_j})\right] = 0 .$$

Using (21) again, we see that the sequence $(\Delta(x_\infty, x_n))_n$ must converge, and since the subsequence $(\Delta(x_\infty, x_{n_j}))_j$ converges to zero, we conclude that $\Delta(x_\infty, x_n) \longrightarrow 0$, which in turn proves that $x_n \longrightarrow x_\infty$, and the proof is complete. □

Let $S \subseteq X$ be the set of all solutions of (1) in the closed ball $B_{\rho/2}[x_0] := \{x \in X \mid \|x - x_0\| \le \rho/2\} \subseteq \mathcal{D}(F)$, i.e., $S := \{x \in B_{\rho/2}[x_0] \mid F(x) = y\}$. Because the operator F is continuous, S is closed and by using the TCC (Assumption 1 (A.4)), we obtain that S is convex. Further, Assumption 1 (A.3) guarantees that S is nonempty because $F(x^\star) = y$ and $\|x^\star - x_0\| \le \varphi^{-1}(\Delta(x^\star, x_0)) < \rho/2$.

Now, since the functional f is proper, l.s.c., and uniformly convex (Assumption 1 (A.2)), the functional $g(\cdot) = \Delta_{\xi_0}(\cdot, x_0)$ is proper, l.s.c., strictly convex, and coercive. Because the space X is reflexive (Assumption 1 (A.6)), there exists a unique $x^\dagger \in S$ such that $x^\dagger = \operatorname{argmin}\{\Delta_{\xi_0}(\,\cdot\,, x_0) \mid x \in S\}$, see e.g., [6, Prop. II 1.2].

We call $x^\dagger$ the x_0-*minimum distance solution.* It is uniquely characterized by $x^\dagger \in S$ and

$$\Delta_{\xi_0}(x^\dagger, x_0) \le \Delta_{\xi_0}(z, x_0) \quad \text{for all} \quad z \in S .$$

In [21] it is proven that, if $x^\dagger \in \operatorname{int}(\operatorname{Dom}(f))$ and $z \in \mathcal{N}(F'(x^\dagger))$, then f is subdifferentiable at $x^\dagger$ and there exists an element $\xi_z^\dagger \in \partial f(x^\dagger)$ such that

$$\langle \xi_z^\dagger - \xi_0, z\rangle \ge 0 . \tag{34}$$

Corollary 4.7. *Let all assumptions from Theorem 4.6 hold true. Additionally, assume that $c_1 = 0$ in Assumption 4, $x^\dagger \in \operatorname{int}(\operatorname{Dom}(f))$, $n_\delta = \infty$, and*

$$\mathcal{N}(F'(x^\dagger)) \subseteq \mathcal{N}(F'(x)) \quad \text{for all} \quad x \in B_\rho(x_0) . \tag{35}$$

Then, $x_n \longrightarrow x^\dagger$ as $n \longrightarrow \infty$.

Proof. From Assumption 4 with $c_1 = 0$ follows that, for $n \in \mathbb{N}$ and $k < k_n$,

$$\xi_{n,k+1} - \xi_{n,k} = -\lambda_{n,k} A_n^* J_r(v_{n,k}) \in \mathcal{R}(F'(x_n)^*) \subseteq \mathcal{N}(F'(x_n))^\perp \subseteq \mathcal{N}(F'(x^\dagger))^\perp ,$$

so that, for $m \in \mathbb{N}$,

$$\xi_m - \xi_0 = \sum_{n=0}^{m-1} \sum_{k=0}^{k_n-1} (\xi_{n,k+1} - \xi_{n,k}) \in \mathcal{N}(F'(x^\dagger))^\perp .$$

Define $z := x_\infty - x^\dagger$ and recall from (29) that $x_n \longrightarrow x_\infty$ with $F(x_\infty) = y$. From TCC (Assumption 1 (A.4)),

$$\|F'(x^\dagger)z\| = \|F'(x^\dagger)(x_\infty - x^\dagger)\| \leq (1+\eta)\|F(x_\infty) - F(x^\dagger)\| = 0 ,$$

which proves that $z \in \mathcal{N}(F'(x^\dagger))$, and therefore, $\langle \xi_m - \xi_0, z\rangle = 0$, for all $m \in \mathbb{N}$. On the other hand, since $x^\dagger \in \operatorname{int}(\operatorname{Dom}(f))$, there exists a vector $\xi_z^\dagger \in \partial f(x^\dagger)$ such that inequality (34) holds. Thus,

$$\langle \xi_z^\dagger - \xi_m, z\rangle \geq 0 \text{ for every } m \in \mathbb{N} . \tag{36}$$

Next, let (x_{n_j}) be a subsequence satisfying (30). We first prove that $x_{n_j} \longrightarrow x^\dagger$. Fix $\epsilon > 0$. Choose $M_1 \in \mathbb{N}$ such that

$$\sum_{n=n_{M_1}}^{\infty} \sum_{k=0}^{k_n-1} \lambda_{n,k}\|v_{n,k}\|^{r-1}\|A_n s_{n,k} - b_n\| < \frac{\mu_{\min}}{3\eta+1}\,\epsilon , \tag{37}$$

which is possible because this series converges, see (28). Now choose $M_2 \in \mathbb{N}$ such that

$$j \geq M_2 \Longrightarrow \|x_{n_j} - x_\infty\| < \frac{\epsilon}{\|\xi_z^\dagger - \xi_{n_{M_1}}\|} .$$

Then, for $j > \max\{M_1, M_2\}$, it follows from (36), (11), and (33) (with $C_1 = 0$, since $c_1 = 0$) that

$$\begin{aligned} 2\varphi(\|x^\dagger - x_{n_j}\|) &\leq \langle \xi_z^\dagger - \xi_{n_j}, x^\dagger - x_{n_j}\rangle \\ &= -\langle \xi_z^\dagger - \xi_{n_j}, z\rangle + \langle \xi_z^\dagger - \xi_{n_{M_1}}, x_\infty - x_{n_j}\rangle + \langle \xi_{n_j} - \xi_{n_{M_1}}, x_{n_j} - x_\infty\rangle \\ &\leq 0 + \|\xi_z^\dagger - \xi_{n_{M_1}}\|\|x_\infty - x_{n_j}\| \\ &\qquad + \frac{3\eta+1}{\mu_{\min}} \sum_{n=n_{M_1}}^{n_j-1} \sum_{k=0}^{k_n-1} \lambda_{n,k}\|v_{n,k}\|^{r-1}\|A_n s_{n,k} - b_n\| \\ &< 2\epsilon . \end{aligned}$$

This proves that $x_{n_j} \longrightarrow x^\dagger$. Since the sequence (x_n) converges to x_∞, see (29), and the subsequence (x_{n_j}) converges to $x^\dagger$, we conclude that $x_\infty = x^\dagger$ and $x_n \longrightarrow x^\dagger$. □

4.2. **Regularization property.** We start with a regularization result in the weak topology. Then, with an additional assumption on the inner iteration, we prove norm convergence.

Theorem 4.8. *Let (δ_j) be a sequence of positive numbers converging to zero. Assume all the hypotheses of Theorem 4.6 and that F is weakly (sequentially) closed*[2]*. Then, any subsequence of $(x_{n_{\delta_j}}^{\delta_j})_j$ has itself a subsequence which converges weakly to a solution of (1). Additionally, if $x^\star$ is the unique solution of the inverse problem in $B_\rho(x_0)$, then the sequence $(x_{n_{\delta_j}}^{\delta_j})_j$ converges weakly to $x^\star$.*

[2] F weakly closed means that, if $(z_k) \subseteq \mathcal{D}(F)$ converges weakly to $z \in X$ and $(F(z_k))$ converges weakly to $y \in Y$, then $z \in \mathcal{D}(F)$ and $y = F(z)$.

Proof. According to Theorem 4.2, the sequence $(x^{\delta_j}_{n_{\delta_j}})_j$ stays within $B_\rho(x_0) \subseteq \mathcal{D}(F)$. Hence, $(x^{\delta_j}_{n_{\delta_j}})_j$ is bounded and, since X is reflexive, there is a subsequence $z_m = x^{\delta_{j_m}}_{n_{\delta_{j_m}}}$ which converges weakly to some $z \in X$. Now, $\|F(z_m)-y\| \leq \|F(z_m)-y^{\delta_{j_m}}\|+\|y-y^{\delta_{j_m}}\| \leq (\tau+1)\delta_{j_m}$ implies that $F(z_m) \longrightarrow y$ as $m \longrightarrow \infty$. As F is weakly closed, we have $z \in \mathcal{D}(F)$ and $y = F(z)$. Applying the same reasoning to any subsequence of $(x^{\delta_j}_{n_{\delta_j}})_j$, we obtain the first result. As at the end of the proof of Theorem 4.6, one can show that $z \in B_\rho(x_0)$.

Now, if $x^\star$ is the unique solution of the inverse problem in $B_\rho(x_0)$ then, any subsequence of $(x^{\delta_j}_{n_{\delta_j}})_j$ has a further subsequence which converges weakly to $x^\star$, and consequently, $x^{\delta_j}_{n_{\delta_j}} \rightharpoonup x^\star$, as $j \longrightarrow \infty$. □

In order to obtain norm convergence, we assume a kind of stability of the inner iteration. For this purpose, we fix a sequence (δ_j) of positive noise levels converging to zero and, for each $j \in \mathbb{N}$, we fix the data vector $y^{\delta_j} \in Y$ satisfying $\|y - y^{\delta_j}\| \leq \delta_j$. Also, for each pair (δ_j, y^{δ_j}), we fix the (finite) sequence $(x^{\delta_j}_n)^{n_{\delta_j}}_{n=0}$. Moreover, we define

$$\overline{n} := \limsup_{j\to\infty} n_{\delta_j},$$

which might be infinity. The arguments we will use consist of successively extracting a subsequence from the current sequence, and to avoid a notation overload, we denote a subsequence of (δ_j) again by (δ_j).

Assumption 5. *Let $n, k \in \mathbb{N}$ be given with $n < \overline{n}$ and $k < \limsup_{j\to\infty} k^{\delta_j}_n$.*
If, for some subsequence of (δ_j), we have that $\xi^{\delta_j}_{n,m} \longrightarrow \xi_{n,m}$ as $j \longrightarrow \infty$, $m = 0, \dots, k$, then $\xi^{\delta_j}_{n,k+1} \longrightarrow \xi_{n,k+1}$ as $j \longrightarrow \infty$ for a further subsequence of (δ_j) (depending on k).

The condition $k < \limsup_{j\to\infty} k^{\delta_j}_n$ in Assumption 5 has the following reason: at least for one subsequence the next inner iteration step should be possible, i.e., $\xi^{\delta_j}_{n,k+1}$ is well defined.

We are now in the position to establish stability of Algorithm 1.

Theorem 4.9 (Stability)**.** *For any $n \leq \overline{n}$ there are a subsequence of (δ_j) (depending on n) and sequences (x_m) and (ξ_m) generated by a run of* `REGINN` *with a suitable choice of $(k_{\max,m})$, such that*

$$x^{\delta_j}_m \longrightarrow x_m \quad \textit{and} \quad \xi^{\delta_j}_m \longrightarrow \xi_m \quad \textit{as} \quad j \longrightarrow \infty\,, \quad \textit{for} \quad m = 0, \dots, n\,. \tag{38}$$

Proof. We argue by induction on n. For $n = 0$ the claim is clearly true because $x^{\delta_j}_0 = x_0$ and $\xi^{\delta_j}_0 = \xi_0$ for all $j \in \mathbb{N}$. Assume that (38) holds true for some $n < \overline{n}$ and some subsequence (δ_j). Our task is to find a $\overline{k} \leq \widehat{k}_n$, see (13) and (15), such that $\xi^{\delta_j}_{n+1} \longrightarrow \xi_{n,\overline{k}}$ for some subsequence. Then, by setting $k_{\max,n} = \overline{k}$, we get $\xi^{\delta_j}_{n+1} \longrightarrow \xi_{n+1}$, and using the continuity of ∇f^*, see, e.g., [30, Thm. 3.5.10], we obtain $x^{\delta_j}_{n+1} = \nabla f^*(\xi^{\delta_j}_{n+1}) \longrightarrow \nabla f^*(\xi_{n+1}) = x_{n+1}$, which will complete the proof.

First, we prove that $\limsup_{j\to\infty} k^{\delta_j}_n \leq \widehat{k}_n < \infty$. Indeed, assume the contrary. By (38), $\xi^{\delta_j}_{n,0} = \xi^{\delta_j}_n \longrightarrow \xi_n = \xi_{n,0}$ as $j \longrightarrow \infty$. Then, by using induction on k, together with Assumption 5, we conclude that, for each $k \leq \widehat{k}_n$, there is a subsequence (depending on k) satisfying

$$\xi^{\delta_j}_{n,m} \longrightarrow \xi_{n,m} \quad \text{as} \quad j \longrightarrow \infty\,, \quad \text{for} \quad m = 0, \dots, k\,.$$

Consequently, after passing to a subsequence, $x_{n,m}^{\delta_j} = \nabla f^*(\xi_{n,m}^{\delta_j}) \longrightarrow \nabla f^*(\xi_{n,m}) = x_{n,m}$, for all $m \leq \widehat{k}_n$. Hence,

$$\|A_n s_{n,\widehat{k}_n} - b_n\| < \mu_n \|b_n\| \Longrightarrow \|A_n^{\delta_j} s_{n,\widehat{k}_n}^{\delta_j} - b_n^{\delta_j}\| < \mu_n \|b_n^{\delta_j}\|$$

for j large enough. We then conclude that $k_n^{\delta_j} \leq \widehat{k}_n^{\delta_j} \leq \widehat{k}_n$ for j large enough. This contradicts the assumption that $\limsup_{j\to\infty} k_n^{\delta_j} > \widehat{k}_n$. Therefore, there is a number $\overline{k} \leq \widehat{k}_n$ and a subsequence of (δ_j) such that $k_n^{\delta_j} = \overline{k}$ for all j.

Now, since $\limsup_{j\to\infty} k_n^{\delta_j} = \overline{k}$, using the same reasoning as above, we have, for a subsequence,

$$x_{n,k}^{\delta_j} \longrightarrow x_{n,k} \quad \text{and} \quad \xi_{n,k}^{\delta_j} \longrightarrow \xi_{n,k} \quad \text{as} \quad j \longrightarrow \infty \quad \text{for} \quad k = 0, \ldots, \overline{k}\,. \tag{39}$$

We thus set $k_{\max,n} = \overline{k}$, which implies that $\xi_{n+1} = \xi_{n,\overline{k}}$. Finally,

$$\lim_{j\to\infty} \xi_{n+1}^{\delta_j} = \lim_{j\to\infty} \xi_{n,k_n^{\delta_j}}^{\delta_j} = \lim_{j\to\infty} \xi_{n,\overline{k}}^{\delta_j} \overset{(39)}{=} \xi_{n,\overline{k}} = \xi_{n+1}.$$

Hence, $x_{n+1}^{\delta_j} \longrightarrow x_{n+1}$ and the inductive argument is complete. □

Theorem 4.10 (Regularization)**.** *Every subsequence of $(x_{n_{\delta_j}}^{\delta_j})_j$ has itself a subsequence which converges strongly to a solution of (1).*

Proof. It is sufficient to prove that $(x_{n_{\delta_j}}^{\delta_j})_j$ has a subsequence which converges strongly to a solution of the inverse problem. We distinguish two cases:

Case 1. The sequence $(n_{\delta_j})_j$ has an accumulation point $\widehat{n} \in \mathbb{N}$. Then, taking a subsequence if necessary, we may assume that $n_{\delta_j} = \widehat{n}$ for all $j \in \mathbb{N}$. Of course, $\widehat{n} \leq \limsup_{j\to\infty} n_{\delta_j} = \overline{n}$. By Theorem 4.9,

$$\lim_{j\to\infty} x_{n_{\delta_j}}^{\delta_j} = \lim_{j\to\infty} x_{\widehat{n}}^{\delta_j} = x_{\widehat{n}}\,.$$

Then, $x_{\widehat{n}}$ must be a solution of the inverse problem:

$$\|F(x_{\widehat{n}}) - y\| = \lim_{j\to\infty} \|F(x_{n_{\delta_j}}^{\delta_j}) - y^{\delta_j}\| \leq \lim_{j\to\infty} \tau\delta_j = 0\,.$$

Case 2. The sequence $(n_{\delta_j})_j$ has no accumulation points. Here, $n_{\delta_j} \longrightarrow \infty$, and we can follow the proof of Theorem 9 from [21]. □

Using the same arguments as in [17] we can prove the following corollary.

Corollary 4.11. *The following assertions hold true.*

1. *The sequence $(x_{k_{\delta_j}}^{\delta_j})$ splits into convergent subsequences, each one converging to a solution of (1).*
2. *If $x^\star$ in Assumption 1* (A.3) *is the unique solution of (1) in $B_\rho(x_0)$, then $x_{k_{\delta_j}}^{\delta_j} \longrightarrow x^\star$ as $j \longrightarrow \infty$;*
3. *If condition (35) holds, $c_1 = 0$ in Assumption 4, and the x_0-minimum distance solution $x^\dagger$ belongs to* $\operatorname{int}(\operatorname{Dom}(f))$*, then the sequence $(x_{k_{\delta_j}}^{\delta_j})_j$ converges to $x^\dagger$.*

Proof. See [17, Cor. 3.12]. □

5. Verification of the assumptions for some regularization schemes

We present examples of iterative regularization schemes for the linearized system $A_n s = b_n^\delta$, which satisfy the required properties to be included as the inner iteration of `REGINN`. These methods are: gradient-type, iterated-Tikhonov, and mixed gradient-Tikhonov methods. We take Assumption 1 as given and check Assumptions 2, 3, and 4 for these methods. The verification of Assumption 5 for the gradient and mixed gradient-Tikhonov methods relies on similar arguments and is therefore presented for both methods together in Appendix A; for the iterated-Tikhonov method we refer to [21, Thm. 8].

5.1. **Gradient methods.** In these methods, the iterates are updated by a multiple of the negative gradient of the functional $H(x) = r^{-1}\|A_n(x - x_n) - b_n^\delta\|^r$, $r > 1$. As Y is smooth (Assumption 1 (A.6)), the gradient is $\nabla H(x) = A_n^* J_r(A_n(x - x_n) - b_n^\delta)$ so that

$$\xi_{n,k+1} = \xi_{n,k} - \lambda_{n,k} A_n^* J_r(A_n(x_{n,k} - x_n) - b_n^\delta), \tag{40}$$

with $\lambda_{n,k} > 0$ to be defined. Thus, $\omega_{n,k} = A_n^* J_r(A_n(x_{n,k} - x_n) - b_n^\delta)$ in step [4.1] of Algorithm 1.

Remark 5.1. *If $k_{\max,n} \equiv 1$, then $k_n \equiv 1$ and* `REGINN` *becomes*

$$\begin{aligned} \xi_{n+1} &= \xi_{n,1} = \xi_n - \lambda_{n,0} F'(x_n)^* J_r(F(x_n) - y^\delta)\,, \\ x_{n+1} &= \nabla f^*(\xi_{n+1})\,, \end{aligned}$$

which is the gradient method applied to the nonlinear inverse problem $F(x) = y^\delta$.

Let $x^\star$ be as in Assumption 1 (A.3), $k < k_n$ and $n < n_\delta$. As before, $e_n := x^\star - x_n$ and $s_{n,k} := x_{n,k} - x_n$. We start with the verification of Assumption 2. By the three points identity (6),

$$\Delta(x^\star, x_{n,k+1}) - \Delta(x^\star, x_{n,k}) = \Delta(x_{n,k}, x_{n,k+1}) + \langle \xi_{n,k+1} - \xi_{n,k}, x_{n,k} - x^\star\rangle\,. \tag{41}$$

From (40) and properties of the duality mapping we obtain

$$\begin{aligned} \langle \xi_{n,k+1} - \xi_{n,k}, x_{n,k} - x^\star\rangle &= -\lambda_{n,k}\langle A_n^* J_r(A_n s_{n,k} - b_n^\delta), s_{n,k} - e_n\rangle \\ &= \lambda_{n,k}\big[\langle J_r(A_n s_{n,k} - b_n^\delta), A_n e_n - b_n^\delta\rangle \\ &\qquad - \langle J_r(A_n s_{n,k} - b_n^\delta), A_n s_{n,k} - b_n^\delta\rangle\big] \\ &\le \lambda_{n,k}\|A_n s_{n,k} - b_n^\delta\|^{r-1}\big[\|A_n e_n - b_n^\delta\| - \|A_n s_{n,k} - b_n^\delta\|\big]\,. \end{aligned} \tag{42}$$

Moreover, from (10),

$$\Delta(x_{n,k}, x_{n,k+1}) \le \varphi^\#(\|\xi_{n,k+1} - \xi_{n,k}\|) = \varphi^\#(\lambda_{n,k}\|A_n^* J_r(A_n s_{n,k} - b_n^\delta)\|) \tag{43}$$

and, from Assumption 1 (A.5),

$$\frac{1}{M}\|A_n s_{n,k} - b_n^\delta\| \le \frac{\|A_n s_{n,k} - b_n^\delta\|^r}{\|A_n^* J_r(A_n s_{n,k} - b_n^\delta)\|}\,.$$

Recalling the function $g(t) = \varphi^\#(t)/t$ from Proposition 2.1, to any choice of $c_0 \in (0,1)$ and $c_1 \in (0, (1 - c_0)/M)$ there exist $0 < \lambda_{n,k}^{\min} < \lambda_{n,k}^{\max}$ such that $\lambda_{n,k} \in \big[\lambda_{n,k}^{\min}, \lambda_{n,k}^{\max}\big]$ yields

$$c_1\|A_n s_{n,k} - b_n^\delta\| \le g(\lambda_{n,k}\|A_n^* J_r(A_n s_{n,k} - b_n^\delta)\|) \le \frac{(1 - c_0)\|A_n s_{n,k} - b_n^\delta\|^r}{\|A_n^* J_r(A_n s_{n,k} - b_n^\delta)\|}\,. \tag{44}$$

From (41), (42), (43) and the upper bound of (44) we obtain (16) for $v_{n,k} := A_n s_{n,k} - b_n^\delta$. So, Assumption 2 is verified.

Now we turn to Assumption 3. Let $w\colon (0,\infty) \longrightarrow (0,\infty)$ be a bijective and increasing function such that $w \geq g$, for instance, $w(t) := g(t) + t$ will do the job. By the lower bound of (44),

$$\frac{w^{-1}(c_1\|A_n s_{n,k} - b_n^\delta\|^r)}{M} \leq \frac{\lambda_{n,k}}{M}\|A_n^* J_r(A_n s_{n,k} - b_n^\delta)\| \leq \lambda_{n,k}\|v_{n,k}\|^{r-1}\,.$$

Thus, Assumption 3 holds for

$$h(t) := \frac{w^{-1}(c_1 t^r)}{M}\,.$$

In view of (40), Assumption 4 trivially holds with $\gamma_{n,k} \equiv 0$, $\alpha_{n,k} \equiv 0$ and $c_1 = 0$. The verification of Assumption 5 can be found in Appendix A.

Remark 5.2. *If f is p-convex, then there exists a constant $\beta > 0$ such that $\varphi^{\#}(\cdot) \leq \beta(\cdot)^{p^*}$. In this case, it is possible to obtain explicit formulas for $\lambda_{n,k}^{\min}$ and $\lambda_{n,k}^{\max}$. Indeed,*

$$\varphi^{\#}(\lambda_{n,k}\|A_n^* J_r(A_n s_{n,k} - b_n^\delta)\|) \leq (1-c_0)\lambda_{n,k}\|A_n s_{n,k} - b_n^\delta\|^r$$

whenever $\lambda_{n,k} \leq \lambda_{n,k}^{\max}$, where

$$\lambda_{n,k}^{\max} := \left(\frac{1-c_0}{\beta}\right)^{\frac{p}{p^*}} \frac{\|A_n s_{n,k} - b_n^\delta\|^{\frac{rp}{p^*}}}{\|A_n^* J_r(A_n s_{n,k} - b_n^\delta)\|^p}.$$

Using (41), (42), and (43) we obtain (16) with $v_{n,k} = A_n s_{n,k} - b_n^\delta$ for all $\lambda_{n,k} \leq \lambda_{n,k}^{\max}$. Now, from Assumption 1 (A.5), $\lambda_{n,k}^{\max} \geq \lambda_{n,k}^{\min}$ with

$$\lambda_{n,k}^{\min} := \left(\frac{1-c_0}{\beta M^{p^*}}\right)^{\frac{p}{p^*}} \|A_n s_{n,k} - b_n^\delta\|^{p-r}$$

and by restricting $\lambda_{n,k}$ to the interval $\left[\lambda_{n,k}^{\min}, \lambda_{n,k}^{\max}\right]$ we also obtain Assumption 3 with

$$h(t) = \left(\frac{1-c_0}{\beta M^{p^*}}\right)^{\frac{p}{p^*}} t^{p-1}.$$

In variations of the Landweber, Steepest Descent, and Minimal Error methods, we have an explicit step size $\lambda_{n,k}$, see e.g., [20, eq. (42)]. If c_0 is properly chosen, we can guarantee that $\lambda_{n,k} \in \left[\lambda_{n,k}^{\min}, \lambda_{n,k}^{\max}\right]$. Hence, all the assumptions hold for these methods.

5.2. **Iterated-Tikhonov methods.** For $\lambda_{n,k} > 0$, we define

(45) $$x_{n,k+1} := \operatorname*{argmin}_{x\in X} T_{n,k}(x), \text{ where } T_{n,k}(x) := \frac{\lambda_{n,k}}{r}\|A_n(x-x_n) - b_n^\delta\|^r + \Delta_{\xi_{n,k}}(x, x_{n,k})\,.$$

Hence, $0 \in \partial T_{n,k}(x_{n,k+1})$, and therefore, the vector $\xi_{n,k} - \lambda_{n,k}A_n^* J_r(A_n(x_{n,k+1} - x_n) - b_n^\delta)$ belongs to $\partial f(x_{n,k+1})$. We thus define the inner iteration

(46) $$\xi_{n,k+1} := \xi_{n,k} - \lambda_{n,k}A_n^* J_r(A_n(x_{n,k+1} - x_n) - b_n^\delta),$$

with $\lambda_{n,k} > 0$ to be determined below. Hence, $\xi_{n,k+1} \in \partial f(x_{n,k+1})$, meaning that $x_{n,k+1} = \nabla f^*(\xi_{n,k+1})$. Obviously, here we set $\omega_{n,k} = A_n^* J_r(A_n(x_{n,k+1} - x_n) - b_n^\delta)$ in [4.1] of Algorithm 1. Moreover, from $T_{n,k}(x_{n,k+1}) \leq T_{n,k}(x_{n,k})$, we find that $\|A_n s_{n,k+1} - b_n^\delta\| \leq \|A_n s_{n,k} - b_n^\delta\|$.

Remark 5.3. *If $k_{\max,n} \equiv 1$, then $k_n \equiv 1$ and* `REGINN` *becomes*

$$x_{n+1} = \operatorname*{argmin}_{x\in X}\left\{\frac{\lambda_{n,0}}{r}\|F'(x_n)(x-x_n)+F(x_n)-y^\delta\|^r + \Delta_{\xi_n}(x,x_n)\right\},$$

$$\xi_{n+1} = \xi_n - \lambda_{n,0}F'(x_n)^* J_r(F'(x_n)(x_{n+1}-x_n)+F(x_n)-y^\delta)\,,$$

which is a variant of the Levenberg-Marquardt method. In case the forward operator F is linear, the above iteration is the iterated-Tikhonov method.

Let $x^\star$ be as in (A.3) of Assumption 1. As usual, $e_n := x^\star - x_n$ and $s_{n,k} := x_{n,k} - x_n$. We start with the verification of Assumption 2. By the three points identity (6),

$$\begin{aligned}\Delta(x^\star, x_{n,k+1}) - \Delta(x^\star, x_{n,k}) &= -\Delta(x_{n,k+1}, x_{n,k}) + \langle \xi_{n,k+1} - \xi_{n,k}, x_{n,k+1} - x^\star\rangle \\ &\le \langle \xi_{n,k+1} - \xi_{n,k}, x_{n,k+1} - x^\star\rangle.\end{aligned} \tag{47}$$

From (46) and properties of the duality mapping,

$$\begin{aligned}\langle \xi_{n,k+1} - \xi_{n,k}, x_{n,k+1} - x^\star\rangle &= -\lambda_{n,k}\langle A_n^* J_r(A_n s_{n,k+1} - b_n^\delta), s_{n,k+1} - e_n\rangle \\ &\le \lambda_{n,k}\|A_n s_{n,k+1} - b_n^\delta\|^{r-1}\left[\|A_n e_n - b_n^\delta\| - \|A_n s_{n,k+1} - b_n^\delta\|\right].\end{aligned} \tag{48}$$

By (11),

$$\begin{aligned}2\varphi(\|x_{n,k+1} - x_{n,k}\|) &\le \langle \xi_{n,k+1} - \xi_{n,k}, x_{n,k+1} - x_{n,k}\rangle \\ &= -\lambda_{n,k}\langle J_r(A_n s_{n,k+1} - b_n^\delta), A_n(s_{n,k+1} - s_{n,k})\rangle \\ &\le \lambda_{n,k}\|A_n s_{n,k+1} - b_n^\delta\|^{r-1}(\|A_n s_{n,k} - b_n^\delta\| - \|A_n s_{n,k+1} - b_n^\delta\|) \\ &\le \lambda_{n,k}\|A_n s_{n,k+1} - b_n^\delta\|^{r-1}\|A_n s_{n,k} - b_n^\delta\| \le \lambda_{n,k}\|A_n s_{n,k} - b_n^\delta\|^r\,.\end{aligned}$$

Therefore, if $0 < \lambda_{n,k} \le \lambda_{n,k}^{\max}$, with

$$\lambda_{n,k}^{\max} := \frac{2\varphi(\frac{1-c_0}{M}\|A_n s_{n,k} - b_n^\delta\|)}{\|A_n s_{n,k} - b_n^\delta\|^r}\,,$$

where $c_0 \in (0,1)$, then

$$\|A_n s_{n,k} - b_n^\delta\| - \|A_n s_{n,k+1} - b_n^\delta\| \le M\|x_{n,k+1} - x_{n,k}\| \le (1-c_0)\|A_n s_{n,k} - b_n^\delta\|\,. \tag{49}$$

Combining the above inequality with (47) and (48) we end up with (16) when $v_{n,k} := A_n s_{n,k+1} - b_n^\delta$. Thus, Assumption 2 holds for the iterated-Tikhonov method.

To verify Assumption 3, define $\lambda_{n,k}^{\min} := c_1\lambda_{n,k}^{\max}$, where $c_1 \in (0,1)$. If $\lambda_{n,k} \in [\lambda_{n,k}^{\min}, \lambda_{n,k}^{\max}]$ then (16) is satisfied and, by taking (49) into account,

$$\lambda_{n,k}\|v_{n,k}\|^{r-1} \ge \lambda_{n,k}^{\min} c_0^{r-1}\|A_n s_{n,k} - b_n^\delta\|^{r-1} = 2c_1c_0^{r-1}\frac{\varphi(\frac{1-c_0}{M}\|A_n s_{n,k} - b_n^\delta\|)}{\|A_n s_{n,k} - b_n^\delta\|}\,.$$

Thus, Assumption 3 applies with

$$h(t) := 2c_1c_0^{r-1}\frac{\varphi(\frac{1-c_0}{M}t)}{t}\,.$$

From (46), Assumption 4 trivially holds with $\gamma_{n,k} \equiv 0$, $\alpha_{n,k} \equiv 0$ and $c_1 = 0$. The proof for Assumption 5 is given in [21, Thm. 8].

Remark 5.4. *There are many variants of the iterated-Tikhonov Method (iTM) depending on the choice of the regularization parameters $\lambda_{n,k}$. If $\lambda_{n,k}$ is constant in k, then the iTM is called stationary. Otherwise, it is called nonstationary. Further, the regularization parameter in the nonstationary iTM may be computed either a priori or a posteriori, that* *is, $\lambda_{n,k}$ is chosen before or during the iteration, respectively. Finally, an a posteriori choice of the regularization parameters may be computed either explicitly or implicitly.*

The version of the iTM presented above employs an explicit a posteriori rule for choosing the regularization parameters. In contrast, the range-relaxed strategy presented in the next subsection, uses an implicit a posteriori rule for choosing $\lambda_{n,k}$, see (52) below.

5.3. **Range-relaxed strategy.** This strategy is a particular type of the nonstationary iterated-Tikhonov method with *a posteriori* choice of the sequence of regularization parameters, see, e.g., [18].

In this method, the current iterate $x_{n,k}$ is "separated" from the set

$$S := \{x \in B_\rho(x_0) \mid F(x) = y\} \subseteq X$$

of solutions by the closed and convex set $\Omega_\gamma := \{x \in X \mid \|A_n(x - x_n) - b_n^\delta\| \le \gamma\}$, $\gamma > 0$. Then, $x_{n,k}$ is projected onto Ω_γ to generate the next iterate $x_{n,k+1}$:

$$x_{n,k+1} = \operatorname{argmin}\big\{\Delta_{\xi_{n,k}}(x, x_{n,k}) \;\big|\; x \in \Omega_\gamma\big\}\,. \tag{50}$$

Note that, if $\gamma < \|A_n(x_{n,k} - x_n) - b_n^\delta\|$ then $x_{n,k} \notin \Omega_\gamma$.

Now, let $x^+ \in S$. From the TCC (Assumption 1 (A.4)) follows

$$\begin{aligned}\|A_n(x^+ - x_n) - b_n^\delta\| &\le \|F(x^+) - F(x_n) - F'(x_n)(x^+ - x_n)\| + \|b_n^\delta - b_n\| \\ &\le \eta\|F(x^+) - F(x_n)\| + \delta \le \eta\left(\|b_n^\delta\| + \|b_n^\delta - b_n\|\right) + \delta \\ &\le \overline{\mu}_n\|b_n^\delta\|\end{aligned} \tag{51}$$

with $\overline{\mu}_n$ as in (17). This means that $S \subseteq \Omega_\gamma$ whenever $\overline{\mu}_n\|b_n^\delta\| \le \gamma$.

Assuming $\overline{\mu}_n\|b_n^\delta\| < \|A_n(x_{n,k} - x_n) - b_n^\delta\|$, the minimization problem (50) can be translated into the following task (see, e.g., [18, Lem.2]): find a pair $(x_{n,k+1}, \lambda_{n,k}) \in X \times (0, \infty)$ such that (45) is satisfied and

$$\overline{\mu}_n\|b_n^\delta\| < \|A_n(x_{n,k+1} - x_n) - b_n^\delta\| < \|A_n(x_{n,k} - x_n) - b_n^\delta\|\,.$$

So, the range-relaxed strategy can be interpreted as the iterated-Tikhonov method with an implicit choice of the regularization parameters. To guarantee the convergence and stability properties, one needs to use convex combinations of the left and right-hand terms in the above inequalities. Summarizing, the range-relaxed strategy is defined as in (45) with the parameter $\lambda_{n,k} > 0$ being chosen such that the computed $x_{n,k+1}$ satisfies

$$(1-t_0)\overline{\mu}_n\|b_n^\delta\| + t_0\|A_n s_{n,k} - b_n^\delta\| \le \|A_n s_{n,k+1} - b_n^\delta\| \le (1-t_1)\overline{\mu}_n\|b_n^\delta\| + t_1\|A_n s_{n,k} - b_n^\delta\|\,, \tag{52}$$

where $0 < t_0 \le t_1 < 1$ are prespecified constants.

Now, Assumption 2 is an immediate consequence of (52) with (47) and (48). In fact, (16) holds for $c_0 = t_0$ and $v_{n,k} = A_n s_{n,k+1} - b_n^\delta$.

Next we consider Assumption 3. From (11) and (46),

$$\begin{aligned}2\varphi(\|x_{n,k+1} - x_{n,k}\|) &\le \langle \xi_{n,k+1} - \xi_{n,k}, x_{n,k+1} - x_{n,k}\rangle \\ &= -\lambda_{n,k}\langle J_r(A_n s_{n,k+1} - b_n^\delta), A_n(x_{n,k+1} - x_{n,k})\rangle \\ &\le M\lambda_{n,k}\|v_{n,k}\|^{r-1}\|x_{n,k+1} - x_{n,k}\|.\end{aligned}$$

Thus,

$$\frac{\varphi(\|x_{n,k+1} - x_{n,k}\|)}{\|x_{n,k+1} - x_{n,k}\|} \le \frac{M}{2}\lambda_{n,k}\|v_{n,k}\|^{r-1}\,. \tag{53}$$

Further, for $k < k_n$ (see (13) and Theorem 4.1),

$$\|A_n s_{n,k} - b_n^\delta\| \geq \mu_n \|b_n^\delta\| > \frac{\overline{\mu}_n}{c_0}\|b_n^\delta\| \,. \tag{54}$$

Hence, by (54), where $c_0 = t_0$, and the second inequality of (52),

$$\begin{aligned} \|x_{n,k+1} - x_{n,k}\| \geq \frac{\|A_n(x_{n,k+1} - x_{n,k})\|}{M} &\geq \frac{\|A_n s_{n,k} - b_n^\delta\| - \|A_n s_{n,k+1} - b_n^\delta\|}{M} \\ &\geq \frac{(1-t_1)(1-t_0)}{M}\|A_n s_{n,k} - b_n^\delta\| \,. \end{aligned} \tag{55}$$

Since $t \mapsto \varphi(t)/t$ is non-decreasing, from (53) and (55) we get that

$$\frac{\varphi\left(\frac{(1-t_1)(1-t_0)}{M}\|A_n s_{n,k} - b_n^\delta\|\right)}{\frac{(1-t_1)(1-t_0)}{M}\|A_n s_{n,k} - b_n^\delta\|} \leq \frac{\varphi(\|x_{n,k+1} - x_{n,k}\|)}{\|x_{n,k+1} - x_{n,k}\|} \leq \frac{M}{2}\lambda_{n,k}\|v_{n,k}\|^{r-1} \,.$$

Now, Assumption 3 holds for

$$h(t) := \frac{2\varphi\left(\frac{(1-t_1)(1-t_0)}{M}t\right)}{(1-t_1)(1-t_0)t}.$$

From (46), Assumption 4 trivially holds with $\gamma_{n,k} \equiv 0$, $\alpha_{n,k} \equiv 0$ and $c_1 = 0$. The validation of Assumption 5 is given in [21, Thm. 8].

5.4. **Mixed gradient-Tikhonov methods.** The main issue in applying a Tikhonov-like method in the inner iteration of REGINN is that a minimizer of the Tikhonov functional must be determined in each step. This may be a costly task. In practice, a gradient-like method can be applied to the Tikhonov functional in order to approximate its minimizer. The mixed method consists in applying a gradient-like method to a Tikhonov functional but performing only a few steps and then changing the Tikhonov functional before the minimizer is found, see, e.g., [19].

A gradient method usually works with the iteration

$$\xi_{n,k+1} = \xi_{n,k} - \lambda_{n,k}\omega_{n,k} \,, \tag{56}$$

where $\lambda_{n,k} > 0$ and $\omega_{n,k} \in \partial H(x_{n,k})$. Here, $H\colon X \longrightarrow \overline{\mathbb{R}}$ is an objective functional. Typically, $H(x) = r^{-1}\|A_n(x - x_n) - b_n^\delta\|^r$, in which case iteration (56) becomes (40). For mixed methods, we use the Tikhonov functional $H(x) = r^{-1}\|A_n(x - x_n) - b_n^\delta\|^r + \alpha_{n,k}P_{n,k}(x)$ as objective functional. Here, $P_{n,k}\colon X \longrightarrow [0,\infty]$ is a proper, l.s.c., and convex penalization term and $(\alpha_{n,k})_k \subseteq (0,\infty)$ is a sequence of regularization parameters. Thus, the iteration of mixed methods is given by

$$\begin{aligned} \xi_{n,k+1} &= \xi_{n,k} - \lambda_{n,k}\left[A_n^* J_r(A_n(x_{n,k} - x_n) - b_n^\delta) + \alpha_{n,k}\gamma_{n,k}\right] , \\ x_{n,k+1} &= \nabla f^*(\xi_{n,k+1}) \,, \end{aligned} \tag{57}$$

with $\gamma_{n,k} \in \partial P_{n,k}(x_{n,k})$.

Remark 5.5. *If we use $P_{n,k}(x) = \Delta_{\xi_{n,k-1}}(x, x_{n,k-1})$, where, by definition, $\xi_{n,-1} = \xi_{n,0} = \xi_n$ (consequently, $x_{n,-1} = x_{n,0} = x_n$), then, the iteration of the mixed method becomes*

$$\begin{aligned} \xi_{n,k+1} &= \xi_{n,k} - \lambda_{n,k}\left[A_n^* J_r(A_n(x_{n,k} - x_n) - b_n^\delta) + \alpha_{n,k}\left(\xi_{n,k} - \xi_{n,k-1}\right)\right] , \\ x_{n,k+1} &= \nabla f^*(\xi_{n,k+1}) \,. \end{aligned}$$

The term in the brackets of the first equation belongs to the subdifferential of the Tikhonov functional $H(x) = T_{n,k}(x) = r^{-1}\|A_n(x - x_n) - b_n^\delta\|^r + \alpha_{n,k}\Delta_{\xi_{n,k-1}}(x, x_{n,k-1})$ at $x_{n,k}$. In this situation, the mixed method can be interpreted as the iterated-Tikhonov method with *inexact minimization. Indeed, $x_{n,k+1}$ is not the exact minimizer of $T_{n,k}$, but only an approximation obtained after performing one single step of a gradient method applied to $T_{n,k}$ with initial guess $x_{n,k}$. A similar reasoning applies to the functional $P_{n,k}(x) = p^{-1}\|x - x_{n,k-1}\|^p$, $p > 1$.*

If $P_{n,k}(x) = P_n(x) = \Delta(x, x_n)$ or $P_n(x) = p^{-1}\|x - x_n\|^p$, then the mixed method becomes a variation of the Tikhonov-Phillips method with inexact minimization.

Remark 5.6. *If $k_{\max,n} \equiv 1$, then* `REGINN` *becomes a variation of the mixed method presented in* [19].

Let $x^\star$ be as in Assumption 1 (A.3). As usual, we set $e_n := x^\star - x_n$ and $s_{n,k} := x_{n,k} - x_n$. We start with the verification of Assumption 2. Assume that $P_{n,k}(x^\star) < \infty$ and choose $d_{n,k} > 0$ such that $P_{n,k}(x^\star) \le d_{n,k}$. Then, as $\gamma_{n,k} \in \partial P_{n,k}(x_{n,k})$, we obtain from the definition of a subdifferential that

$$\langle \gamma_{n,k}, x^\star - x_{n,k} \rangle \le P_{n,k}(x^\star) - P_{n,k}(x_{n,k}) \le d_{n,k}\,.$$

We now choose $(\alpha_{n,k})$ such that

$$\alpha_{n,k} \le (1 - c_2)\, d_{n,k}^{-1}\, \|A_n s_{n,k} - b_n^\delta\|^r \tag{58}$$

and

$$\alpha_{n,k}\|\gamma_{n,k}\| \le c_1 \|A_n s_{n,k} - b_n^\delta\|^{r-1} \min\{\|A_n s_{n,k} - b_n^\delta\|, 1\}\,, \tag{59}$$

where $c_2 \in (0,1)$ and $c_1 > 0$. Similarly to (42), we obtain that

$$\begin{aligned} \langle \xi_{n,k+1} - \xi_{n,k}, x_{n,k} - x^\star \rangle &\le \lambda_{n,k}\|A_n s_{n,k} - b_n^\delta\|^{r-1} \left[\|A_n e_n - b_n^\delta\| - \|A_n s_{n,k} - b_n^\delta\|\right] \\ &\qquad + \lambda_{n,k}\alpha_{n,k} d_{n,k} \\ &\le \lambda_{n,k}\|A_n s_{n,k} - b_n^\delta\|^{r-1} \left[\|A_n e_n - b_n^\delta\| - c_2\|A_n s_{n,k} - b_n^\delta\|\right]. \end{aligned} \tag{60}$$

Set $\nabla_{n,k} := A_n^* J_r(A_n s_{n,k} - b_n^\delta) + \alpha_{n,k}\gamma_{n,k}$ and fix $c_0 \in (0, c_2)$. Following (43) and (44), we choose $0 < \lambda_{n,k}^{\min} < \lambda_{n,k}^{\min}$ such that $\lambda_{n,k} \in \left[\lambda_{n,k}^{\min}, \lambda_{n,k}^{\max}\right]$ implies

$$\frac{c_3\|A_n s_{n,k} - b_n^\delta\|^r}{M\|A_n s_{n,k} - b_n^\delta\|^{r-1} + \alpha_{n,k}\|\gamma_{n,k}\|} \le g(\lambda_{n,k}\|\nabla_{n,k}\|) \le \frac{(c_2 - c_0)\|A_n s_{n,k} - b_n^\delta\|^r}{\|\nabla_{n,k}\|}\,, \tag{61}$$

where $0 < c_3 \le c_2 - c_0$. Thus,

$$\Delta(x_{n,k}, x_{n,k+1}) \le \varphi^\#(\lambda_{n,k}\|\nabla_{n,k}\|) \le (c_2 - c_0)\lambda_{n,k}\|A_n s_{n,k} - b_n^\delta\|^r$$

and from (41) we obtain (16) with $v_{n,k} = A_n s_{n,k} - b_n^\delta$. The verification of Assumption 2 is complete.

To verify Assumption 3, let $w\colon (0,\infty) \longrightarrow (0,\infty)$ be any bijective and increasing function such that $w \ge g$. From (59) there is a constant $c_4 > 0$ such that

$$\|\nabla_{n,k}\| \le M\|A_n s_{n,k} - b_n^\delta\|^{r-1} + \alpha_{n,k}\|\gamma_{n,k}\| \le c_4\|A_n s_{n,k} - b_n^\delta\|^{r-1}.$$

Thus, from the first inequality in (61), we get

$$w^{-1}\left(\frac{c_3}{c_4}\|A_n s_{n,k} - b_n^\delta\|\right) \le \lambda_{n,k}\|\nabla_{n,k}\| \le c_4\lambda_{n,k}\|v_{n,k}\|^{r-1}.$$

Hence, Assumption 3 holds with

$$h(t) := \frac{1}{c_4}\, w^{-1}\left(\frac{c_3}{c_4}\, t\right).$$

From (57) and (59), Assumption 4 holds. For Assumption 5, see Appendix A.

Remark 5.7. *To implement the mixed method, one must choose the sequence* $(\alpha_{n,k})$, *which in turn depends on the upper bounds* $d_{n,k}$ *of the sequence* $(P_{n,k}(x^\star))$, *see* (58). *In some cases, it is easy to determine these upper bounds. For instance, if* $P_{n,k}(x) = p^{-1}\|x-x_0\|^p$, $p>1$, *then using Assumption* 1 (A.3) *and* (9), *one can prove that* $P_{n,k}(x^\star) \le p^{-1}(\rho/2)^p =: d_{n,k}$. *A similar reasoning applies in case* $P_{n,k}(x) = p^{-1}\|x\|^p$. *If, as in Remark* 5.5, $P_{n,k}(x) = \Delta(x, x_{n,k-1})$, *then one may use induction, along with Assumption* 1 (A.3), *to validate Assumptions* 2, 3, *and* $P_{n,k}(x^\star) < \varphi(\rho/2) =: d_{n,k}$ *at the same time.*

6. Tikhonov-Phillips as inner iteration

In this section, we provide a convergence analysis of `REGINN` with the Tikhonov-Phillips method as inner iteration. As it does not fit into our general framework of the previous sections, see Remark 6.2 below, it needs a separate treatment. We emphasize that the convergence result presented is nevertheless novel.

We define

$$\begin{aligned} T_{n,k}(x) &:= \frac{\lambda_{n,k}}{r}\|A_n(x-x_n) - b_n^\delta\|^r + \Delta_{\xi_n}(x, x_n)\,, \\ x_{n,k+1} &:= \operatorname*{argmin}_{x\in X} T_{n,k}(x)\,, \\ \xi_{n,k+1} &:= \xi_n - \lambda_{n,k}A_n^* J_r(A_n(x_{n,k+1}-x_n) - b_n^\delta)\,, \end{aligned} \tag{62}$$

with $\lambda_{n,k} > 0$ satisfying, for each $n \in \mathbb{N}$,

$$1 < \frac{\lambda_{n,k+1}}{\lambda_{n,k}} \le \sigma \quad \text{and} \quad \lim_{k\to\infty} \lambda_{n,k} = \infty\,, \tag{63}$$

where $1 < \sigma < (r+1)/r$. Moreover, we assume that

$$\sum_{n=0}^{\infty} \lambda_{n,0} = \infty\,. \tag{64}$$

Observe that the difference between the Tikhonov functionals in (62) and (45) occurs in the second entry of the Bregman distance. Observe too, that $0 \in \partial T_{n,k}(x_{n,k+1})$, which implies $\xi_{n,k+1} \in \partial f(x_{n,k+1})$, meaning that $x_{n,k+1} = \nabla f^*(\xi_{n,k+1})$.

Because the sequence $(\lambda_{n,k})_k$ is increasing and grows to infinity, one can prove that (see, e.g., [24, Lem. 30])

$$\|A_n s_{n,k+1} - b_n^\delta\| \le \|A_n s_{n,k} - b_n^\delta\| \quad \text{and} \quad \lim_{k\to\infty}\|A_n s_{n,k} - b_n^\delta\| = \inf_{s\in X}\|A_n s - b_n^\delta\|\,. \tag{65}$$

Remark 6.1. *In contrast to the iterated-Tikhonov method, the use of the Tikhonov-Phillips method as inner iteration of* `REGINN` *always results in a variation of the Levenberg-Marquardt method, independently of the choice of* $k_{\max,n}$.

Similarly to (48), we obtain, for every solution $x^\star \in B_\rho(x_0)$, that

$$\begin{aligned} \langle \xi_{n,k+1} - \xi_n, x_{n,k+1} - x^\star\rangle \le& -\lambda_{n,k}\|A_n s_{n,k+1} - b_n^\delta\|^r \\ &+ \lambda_{n,k}\|A_n s_{n,k+1} - b_n^\delta\|^{r-1}\|A_n e_n - b_n^\delta\|\,. \end{aligned} \tag{66}$$

Now, the inequality $T_{n,k-1}(x_{n,k}) \le T_{n,k-1}(x_{n,k+1})$ implies

$$-\lambda_{n,k-1}\|A_n s_{n,k+1} - b_n^\delta\|^r \le -\lambda_{n,k-1}\|A_n s_{n,k} - b_n^\delta\|^r + r\Delta(x_{n,k+1}, x_n)\,,$$

which, in view of (63), yields

$$-\lambda_{n,k}\|A_n s_{n,k+1} - b_n^\delta\|^r \le -\frac{\lambda_{n,k}}{\sigma}\|A_n s_{n,k} - b_n^\delta\|^r + r\Delta(x_{n,k+1}, x_n)\,.$$

Then, by using the three points identity (6), we obtain, considering (66), that

(67)
$$\begin{aligned}\Delta(x^\star, x_{n,k+1}) - \Delta(x^\star, x_n) &= -\Delta(x_{n,k+1}, x_n) + \langle \xi_{n,k+1} - \xi_n, x_{n,k+1} - x^\star\rangle \\ &\le -\frac{\lambda_{n,k}}{\sigma}\|A_n s_{n,k} - b_n^\delta\|^r + \lambda_{n,k}\|A_n s_{n,k+1} - b_n^\delta\|^{r-1}\|A_n e_n - b_n^\delta\| \\ &\qquad + (r-1)\Delta(x_{n,k+1}, x_n)\,.\end{aligned}$$

But, the inequality $T_{n,k}(x_{n,k+1}) \le T_{n,k}(x_n)$ implies $\Delta(x_{n,k+1}, x_n) \le r^{-1}\lambda_{n,k}\|b_n^\delta\|^r$. Then, from (67) and the inequality in (65) we get

$$\begin{aligned}\Delta(x^\star, x_{n,k+1}) - \Delta(x^\star, x_n) \le \lambda_{n,k}\big\{&\|A_n s_{n,k} - b_n^\delta\|^{r-1} \\ &\big[\|A_n e_n - b_n^\delta\| - c_1\|A_n s_{n,k} - b_n^\delta\|\big] + c_2\|b_n^\delta\|^r\big\}.\end{aligned} \tag{68}$$

with $c_1 = 1/\sigma \in (0,1)$ and $c_2 = (1 - 1/r) \in (0, c_1)$.

Remark 6.2. *The Tikhonov-Phillips method behaves similarly to the previous ones, and many of its properties are, to some extent, similar. However, the fact that the k-th inner iterate does not depend on the previous one makes its iteration, defined in* (62)*, and the inequality* (68) *above to be considerably different from the previous ones. This hampers the inclusion of the method in a unified convergence analysis.*

We assume that $\eta < c_1 - c_2$ and $\tau > (1+\eta)/(c_1 - c_2 - \eta)$. Then, for any $n < n_\delta$,

$$\overline{\mu}_n + c_2 < \frac{1+\eta}{\tau} + \eta + c_2 < c_1\,, \tag{69}$$

where $\overline{\mu}_n$ is defined in (17). Now, since

$$\lim_{t\to 1^-}\left(\frac{\frac{1+\eta}{\tau} + \eta}{t} + \frac{c_2}{t^r}\right) < c_1\,,$$

there exists $\mu_{\min} \in \big((1+\eta)/\tau + \eta, 1\big)$ such that $\mu_n \in [\mu_{\min}, 1)$ implies

$$\frac{\frac{1+\eta}{\tau} + \eta}{\mu_n} + \frac{c_2}{\mu_n^r} < c_1\,. \tag{70}$$

By (65) and (51),

$$\lim_{k\to\infty}\|A_n s_{n,k} - b_n^\delta\| \le \|A_n e_n - b_n^\delta\| \le \overline{\mu}_n\|b_n^\delta\| \le \left(\frac{1+\eta}{\tau} + \eta\right)\|b_n^\delta\|\,,$$

and, since $(1+\eta)/\tau + \eta < \mu_{\min} \le \mu_n$, we get $\widehat{k}_n < \infty$, i.e., the inner iteration terminates. Using again (51), we obtain

$$\|A_n e_n - b_n^\delta\| \le \frac{\overline{\mu}_n}{\mu_n}\mu_n\|b_n^\delta\| \le \frac{\overline{\mu}_n}{\mu_n}\|A_n s_{n,k_n-1} - b_n^\delta\|.$$

Now, from (68) follows that

$$\begin{aligned}\Delta(x^\star, x_{n+1}) - \Delta(x^\star, x_n) &= \Delta(x^\star, x_{n,k_n}) - \Delta(x^\star, x_n) \\ &\le \lambda_{n,k_n-1}\big\{\|A_n s_{n,k_n-1} - b_n^\delta\|^{r-1}\big[\|A_n e_n - b_n^\delta\| \\ &\qquad - c_1\|A_n s_{n,k_n-1} - b_n^\delta\|\big] + c_2\|b_n^\delta\|^r\big\} \\ &\le -C_n\lambda_{n,k_n-1}\|A_n s_{n,k_n-1} - b_n^\delta\|^r\,,\end{aligned}$$

with

$$C_n = c_1 - \left(\frac{\overline{\mu}_n}{\mu_n} + \frac{c_2}{\mu_n^r}\right).$$

Observe that $C_n > 0$ by (69) and (70). Hence,

$$\Delta(x^\star, x_{n+1}) < \Delta(x^\star, x_n), \quad \text{for } n = 0, \dots, n_\delta - 1. \tag{71}$$

Setting $\mu_n = \mu_{\min}$ in (70) we obtain

$$0 < \overline{C} := c_1 - \left(\frac{\frac{1+\eta}{\tau} + \eta}{\mu_{\min}} + \frac{c_2}{\mu_{\min}^r}\right) < C_n \quad \text{for } n < n_\delta.$$

Consequently, any $j \leq n_\delta$ must satisfy

$$\begin{aligned}
\overline{C}(\tau\delta)^r \sum_{n=0}^{j-1} \lambda_{n,0} \leq \overline{C} \sum_{n=0}^{j-1} \lambda_{n,0} \|b_n^\delta\|^r &\leq \sum_{n=0}^{j-1} C_n \frac{\lambda_{n,k_n-1}}{\mu_{\min}^r} \|A_n s_{n,k_n-1} - b_n^\delta\|^r \\
&\leq \frac{1}{\mu_{\min}^r} \sum_{n=0}^{j-1} [\Delta(x^\star, x_n) - \Delta(x^\star, x_{n+1})] \\
&= \frac{1}{\mu_{\min}^r} (\Delta(x^\star, x_0) - \Delta(x^\star, x_j)) \leq \frac{1}{\mu_{\min}^r} \Delta(x^\star, x_0) < \infty.
\end{aligned}$$

In view of (64), this proves $n_\delta < \infty$ whenever $\delta > 0$.

Now, by (71), the statement of Theorem 4.8 carries over to the Tikhonov-Phillips method. Moreover, Corollary 4.4 holds as well with $c_0 = 1$.

As before, if there is no premature termination in the noiseless situation ($\delta = 0$), we have

$$\sum_{n=0}^{\infty} \lambda_{n,0} \|b_n\|^r \leq \sum_{n=0}^{\infty} \frac{\lambda_{n,k_n-1}}{\mu_{\min}^r} \|A_n s_{n,k_n-1} - b_n\|^r < \infty, \tag{72}$$

which, recalling (64), implies $F(x_n) \longrightarrow y$ as $n \longrightarrow \infty$.

Convergence in the noiseless case is obtained similarly to the proof of Theorem 4.6. Indeed, if $n_\delta = \infty$ (for $\delta = 0$), fix a subsequence (x_{n_j}) satisfying (30) and proceed with

$$\begin{aligned}
|\langle \xi_{n_\ell} - \xi_{n_m}, x_{n_\ell} - x^\star\rangle| &= \left|\sum_{n=n_m}^{n_\ell-1} \langle \xi_{n+1} - \xi_n, e_{n_\ell}\rangle\right| = \left|\sum_{n=n_m}^{n_\ell-1} \langle \xi_{n,k_n} - \xi_n, e_{n_\ell}\rangle\right| \\
&= \left|\sum_{n=n_m}^{n_\ell-1} -\lambda_{n,k_n-1} \langle J_r(A_n s_{n,k_n} - b_n), A_n e_{n_\ell}\rangle\right| \\
&\leq \sum_{n=n_m}^{n_\ell-1} \lambda_{n,k_n-1} \|A_n s_{n,k_n} - b_n\|^{r-1} \|A_n e_{n_\ell}\| \\
&\leq \frac{3\eta+1}{\mu_{\min}} \sum_{n=n_m}^{n_\ell-1} \lambda_{n,k_n-1} \|A_n s_{n,k_n-1} - b_n\|^r,
\end{aligned}$$

where the last estimate is due to (32) and (65). Finally, (31) with (71) and (72) reveal (x_{n_j}) to be a Cauchy sequence. As in the proof of Theorem 4.6, we conclude that the whole sequence (x_n) converges to a solution of (1).

Still assuming that $\delta = 0$ and $n_\delta = \infty$, we obtain from the definition (62) together with (35), that for all $n \in \mathbb{N}$,

$$\xi_{n,k_n} - \xi_n \in \mathcal{R}(F'(x_n)^*) \subseteq \mathcal{N}(F'(x_n))^\perp \subseteq \mathcal{N}(F'(x^\dagger))^\perp .$$

Thus, for all $m \in \mathbb{N}$ we have

$$\xi_m - \xi_0 = \sum_{n=0}^{m-1}(\xi_{n+1} - \xi_n) = \sum_{n=0}^{m-1}(\xi_{n,k_n} - \xi_n) \in \mathcal{N}(F'(x^\dagger))^\perp .$$

Further, by replacing the inequality (37) by

$$\sum_{n=n_{M_1}}^{\infty} \lambda_{n,k_n-1}\|A_n s_{n,k_n-1} - b_n\|^r < \frac{\mu_{\min}}{3\eta + 1}\epsilon ,$$

we obtain, in view of the convergence in the noiseless case, that Corollary 4.7 holds true for the Tikhonov-Phillips method.

Theorems 4.9 and 4.10 are verified only based on the structure of `REGINN`, Assumption 5 (which can be proved as a particular case of [21, Thm. 8]), and convergence in the noiseless case. Corollary 4.11 follows from Theorem 4.10 and Corollary 4.7. Therefore, these results also apply to the Tikhonov-Phillips method.

Appendix A. Stability of gradient and mixed methods

In what follows, we prove Assumption 5 for the gradient and mixed methods.

Let $n \in \mathbb{N}$ and $k \in \mathbb{N}$ be fixed. Assume that $n < \overline{n}$ and $k < \limsup_{j\to\infty} k_n^{\delta_j}$. Also, assume that

$$\xi_{n,m}^{\delta_j} \longrightarrow \xi_{n,m} \quad \text{as} \quad j \longrightarrow \infty , \quad \text{for} \quad m = 0, \dots, k . \tag{73}$$

We shall prove that, there is a subsequence of (δ_j) (depending on k) such that $\xi_{n,k+1}^{\delta_j} \longrightarrow \xi_{n,k+1}$ as $j \longrightarrow \infty$.

We start with the proof of gradient methods.

From (44) and the continuity of the function g, the sequences $(\lambda_{n,k}^{\min,\delta_j})_j$ and $(\lambda_{n,k}^{\max,\delta_j})_j$ are bounded, with $(\lambda_{n,k}^{\min,\delta_j})_j$ being bounded away from zero. Therefore, taking a subsequence if necessary, we may assume that $\lambda_{n,k}^{\min,\delta_j} \longrightarrow \lambda^{\min} > 0$ and $\lambda_{n,k}^{\max,\delta_j} \longrightarrow \lambda^{\max} \geq \lambda^{\min} > 0$. Since $\lambda_{n,k}^{\delta_j} \in \left[\lambda_{n,k}^{\min,\delta_j}, \lambda_{n,k}^{\max,\delta_j}\right]$ for all j, we conclude that, for a subsequence, $\lambda_{n,k}^{\delta_j} \longrightarrow \lambda$ and $\lambda \in [\lambda^{\min}, \lambda^{\max}]$. Now, observe that $x_n^{\delta_j} = \nabla f^*(\xi_{n,0}^{\delta_j}) \longrightarrow \nabla f^*(\xi_{n,0}) = x_n$ as $j \longrightarrow \infty$. By (44),

$$c_1\|A_n s_{n,k} - b_n\|^r \leq g(\lambda\|A_n^* J_r(A_n s_{n,k} - b_n)\|) \leq \frac{(1 - c_0)\|A_n s_{n,k} - b_n\|^r}{\|A_n^* J_r(A_n s_{n,k} - b_n)\|} . \tag{74}$$

Thus, by defining $\lambda_{n,k} := \lambda$, we get $\lambda_{n,k}^{\delta_j} \longrightarrow \lambda_{n,k} \in [\lambda^{\min}, \lambda^{\max}]$ as $j \longrightarrow \infty$.

Finally, from the continuity of F' and of J_r on Y (Assumption 1 (A6)), together with (73) and (2) we obtain

$$\xi_{n,k+1}^{\delta_j} = \xi_{n,k}^{\delta_j} - \lambda_{n,k}^{\delta_j} A_n^{\delta_j,*} J_r(A_n^{\delta_j} s_{n,k}^{\delta_j} - b_n^{\delta_j}) \longrightarrow \xi_{n,k} - \lambda_{n,k} A_n^* J_r(A_n s_{n,k} - b_n) = \xi_{n,k+1} , \tag{75}$$

as we wanted.

Concerning the mixed methods, we additionally assume the sequences $(d_{n,k})$ to be independent of the noise level[3] δ and the convex functionals $P_{n,k}\colon X \longrightarrow [0,\infty]$ in (57) to be Fréchet differentiable. Then, its derivative is continuous, see, e.g., [1, Cor. 4.2.12]. Hence, $\gamma_{n,k}^{\delta_j} = \nabla P_{n,k}(x_{n,k}^{\delta_j}) \longrightarrow \nabla P_{n,k}(x_{n,k}) = \gamma_{n,k}$ as $j \longrightarrow \infty$. Further, by (58),

$$\alpha_{n,k}^{\delta_j} \leq \frac{1-c_2}{d_{n,k}} \| A_n^{\delta_j} s_{n,k}^{\delta_j} - b_n^{\delta_j} \|^r ,$$

from which we conclude, taking a subsequence if necessary, that $\alpha_{n,k}^{\delta_j} \longrightarrow \alpha_{n,k}$ as $j \longrightarrow \infty$, where $\alpha_{n,k} \leq (1-c_2) d_{n,k}^{-1} \| A_n s_{n,k} - b_n \|^r$.

Taking (61) into account, we deduce, similarly to (74) and (75), that $\xi_{n,k+1}^{\delta_j} \longrightarrow \xi_{n,k+1}$ as $j \longrightarrow \infty$, and the proof is complete.

References

[1] J. M. Borwein and J. D. Vanderwerff, *Convex functions: constructions, characterizations and counterexamples*, vol. 109 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 2010, https://doi.org/10.1017/CBO9781139087322.

[2] I. Cioranescu, *Geometry of Banach spaces, duality mappings and nonlinear problems*, vol. 62 of Mathematics and its Applications, Kluwer Academic Publishers Group, Dordrecht, 1990, https://doi.org/10.1007/978-94-009-2121-4.

[3] C. Clason and B. Jin, *A semismooth Newton method for nonlinear parameter identification problems with impulsive noise*, SIAM J. Imaging Sci., 5 (2012), pp. 505–538, http://dx.doi.org/10.1137/110826187.

[4] I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., 57 (2004), pp. 1413–1457, http://dx.doi.org/10.1002/cpa.20042.

[5] R. S. Dembo, S. C. Eisenstat, and T. Steihaug, *Inexact Newton methods*, SIAM J. Numer. Anal., 19 (1982), pp. 400–408, http://dx.doi.org/10.1137/0719025.

[6] I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*, vol. 28 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, english ed., 1999, https://doi.org/10.1137/1.9781611971088. Translated from the French.

[7] M. Eller, R. Griesmaier, and A. Rieder, *Tangential cone condition for the full waveform forward operator in the viscoelastic regime: the non-local case*, CRC 1173 Preprint 2023/8, Karlsruhe Institute of Technology, 2023, https://doi.org/10.5445/IR/1000155827.

[8] R. Gu and B. Han, *Inexact Newton regularization in Banach spaces based on two-point gradient method with uniformly convex penalty terms*, Appl. Numer. Math., 160 (2021), pp. 122–145, https://doi.org/10.1016/j.apnum.2020.09.018.

[9] M. Hanke, *A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems*, Inverse Problems, 13 (1997), pp. 79–95, https://doi.org/10.1088/0266-5611/13/1/007.

[10] M. Hanke, *Regularizing properties of a truncated Newton-CG algorithm for nonlinear inverse problems*, Numer. Funct. Anal. Optim., 18 (1997), pp. 971–993, http://dx.doi.org/10.1080/01630569708816804.

[11] Q. Jin, *Inexact Newton-Landweber iteration for solving nonlinear inverse problems in Banach spaces*, Inverse Problems, 28 (2012), pp. 065002, 15, http://dx.doi.org/10.1088/0266-5611/28/6/065002.

[12] Q. Jin, *Inexact Newton-Landweber iteration in Banach spaces with nonsmooth convex penalty terms*, SIAM J. Numer. Anal., 53 (2015), pp. 2389–2413, https://doi.org/10.1137/130940505.

[3]The number $d_{n,k}$ is an upper bound for $P_{n,k}(x^\star)$. In principle, the functional $P_{n,k}$ may depend on δ, as in $P_{n,k}^\delta(x) = \|x - x_{n,k}^\delta\|^2$. However, for the cases of interest, we can assume that $(d_{n,k}^\delta)_\delta$ is bounded whenever $\delta \in [0,\delta_0)$ with $\delta_0 > 0$ constant. Thus, one may replace $(d_{n,k}^\delta)_\delta$ by its upper bound. So, w.l.o.g., we assume that $(d_{n,k})$ is independent of δ.

[13] Q. Jin and H. Yang, *Levenberg-Marquardt method in Banach spaces with general convex regularization terms*, Numer. Math., 133 (2016), pp. 655–684, https://doi.org/10.1007/s00211-015-0764-z.

[14] B. Kaltenbacher, T. T. N. Nguyen, and O. Scherzer, *The tangential cone condition for some coefficient identification model problems in parabolic pdes*, in Time-dependent Problems in Imaging and Parameter Identification, Springer, Cham, 2021, pp. 121–163, https://doi.org/10.1007/978-3-030-57784-1_5.

[15] A. Lechleiter and A. Rieder, *Newton regularizations for impedance tomography: convergence by local injectivity*, Inverse Problems, 24 (2008), pp. 065009, 18, https://doi.org/10.1088/0266-5611/24/6/065009.

[16] A. Lechleiter and A. Rieder, *Towards a general convergence theory for inexact Newton regularizations*, Numer. Math., 114 (2010), pp. 521–548, http://dx.doi.org/10.1007/s00211-009-0256-0.

[17] A. Leitão, F. Margotti, and B. F. Svaiter, *Range-relaxed criteria for choosing the Lagrange multipliers in the Levenberg–Marquardt method*, IMA J. Numer. Anal., 41 (2021), pp. 2962–2989, https://doi.org/10.1093/imanum/draa050.

[18] M. P. Machado, F. Margotti, and A. Leitão, *On nonstationary iterated Tikhonov methods for ill-posed equations in Banach spaces*, in New trends in parameter identification for mathematical models, Trends Math., Birkhäuser/Springer, Cham, 2018, pp. 175–193, https://doi.org/10.1007/978-3-319-70824-9_10.

[19] F. Margotti, *Mixed gradient-Tikhonov methods for solving nonlinear ill-posed problems in Banach spaces*, Inverse Problems, 32 (2016), pp. 125012, 25, http://dx.doi.org/10.1088/0266-5611/32/12/125012.

[20] F. Margotti, *Inexact Newton regularization combined with gradient methods in Banach spaces*, Inverse Problems, 34 (2018), p. 075007, http://stacks.iop.org/0266-5611/34/i=7/a=075007.

[21] F. Margotti and E. Hafemann, *Range-relaxed strategy applied to the Levenberg-Marquardt method with uniformly convex penalization term in Banach spaces*, Inverse Problems, 38 (2022), p. 095001 (35pp), https://doi.org/10.1088/1361-6420/ac7e68.

[22] F. Margotti and A. Rieder, *An inexact Newton regularization in Banach spaces based on the nonstationary iterated Tikhonov method*, J. Inverse Ill-Posed Probl., 23 (2015), pp. 373–392, http://dx.doi.org/10.1515/jiip-2014-0035.

[23] F. Margotti, A. Rieder, and A. Leitão, *A Kaczmarz version of the REGINN-Landweber iteration for ill-posed problems in Banach spaces*, SIAM J. Numer. Anal., 52 (2014), pp. 1439–1465, http://dx.doi.org/10.1137/130923956.

[24] F. J. Margotti, *On Inexact Newton Methods for Inverse Problems in Banach Spaces*, PhD thesis, Karlsruher Institut für Technologie (KIT), 2015, https://dx.doi.org/10.5445/IR/1000048606.

[25] L. Pieronek and A. Rieder, *On the iterative regularization of non-linear illposed problems in L^∞*, CRC 1173 Preprint 2021/46, Karlsruhe Institute of Technology, dec 2021, https://doi.org/10.5445/IR/1000140578.

[26] A. Rieder, *On the regularization of nonlinear ill-posed problems via inexact Newton iterations*, Inverse Problems, 15 (1999), pp. 309–327, http://dx.doi.org/10.1088/0266-5611/15/1/028.

[27] A. Rieder, *On convergence rates of inexact Newton regularizations*, Numer. Math., 88 (2001), pp. 347–365, http://dx.doi.org/10.1007/PL00005448.

[28] A. Rieder, *Inexact Newton regularization using conjugate gradients as inner iteration*, SIAM J. Numer. Anal., 43 (2005), pp. 604–622 (electronic), http://dx.doi.org/10.1137/040604029.

[29] C. Zălinescu, *On uniformly convex functions*, J. Math. Anal. Appl., 95 (1983), pp. 344–374, https://doi.org/10.1016/0022-247X(83)90112-9.

[30] C. Zălinescu, *Convex analysis in general vector spaces*, World Scientific Publishing Co., Inc., River Edge, NJ, 2002, https://doi.org/10.1142/9789812777096.

Department of Mathematics, Federal Univ. of St. Catarina, P.O. Box 476, 88040-900 Florianópolis, SC, Brazil
Email address: fabio.margotti@ufsc.br
Email address: marco.pauleti@gmail.com

Department of Mathematics, Karlsruhe Institute of Technology, D-76128 Karlsruhe, Germany
Email address: andreas.rieder@kit.edu