

A KACZMARZ VERSION OF THE REGINN-LANDWEBER ITERATION FOR ILL-POSED PROBLEMS IN BANACH SPACES*

FÁBIO MARGOTTI[†], ANDREAS RIEDER[†], AND ANTONIO LEITÃO[‡]

Abstract. In this work we present and analyze a Kaczmarz version of the iterative regularization scheme REGINN-Landweber for nonlinear ill-posed problems in Banach spaces [Q. Jin, *Inverse Problems* 28 (2012), 065002]. Kaczmarz methods are designed for problems which split into smaller subproblems which are then processed cyclically during each iteration step. Under standard assumptions on the Banach space and on the nonlinearity we prove stability and (norm) convergence as the noise level tends to zero. Further, we test our scheme on the inverse problem of two dimensional electric impedance tomography not only to illustrate our theoretical findings but also to study the influence of different Banach spaces on the reconstructed conductivities.

Key words. nonlinear ill-posed problems, Banach spaces, iterative regularization

AMS subject classifications. 65J20, 65J15

DOI. 10.1137/130923956

1. Introduction. Our goal is to solve the nonlinear ill-posed problem

$$(1) \quad F(x) = y,$$

where $F: D(F) \subset X \rightarrow Y$ operates between the *Banach spaces* X and Y with domain of definition $D(F)$. This kind of inverse problem gained a lot of interest over the last years because several applications and constraints are formulated quite naturally in a Banach space framework: sparsity, uniform and impulsive noise, preservation of discontinuities (edges), etc. In parameter identification tasks, for instance, the searched for parameter often appears in the governing partial differential equations as an L^∞ -coefficient, e.g., electrical impedance tomography [2, 4, 5] (see our numerical experiment section below).

A variety of techniques for solving (1) in Banach spaces is on the market and the field has meanwhile reached a considerable level of maturity. For an overview, examples, and references we point to the monograph [26], to the special section *Tackling inverse problems in a Banach space environment* (*Inverse Problems*, 28 (10), 2012) and to the topical review article [13].

In the present work we contribute to the analysis of iterative solvers where we combine the Kaczmarz approach for ill-posed problems as introduced by [22, 23, 18, 3, 8] with an inexact Newton iteration due to [9, 24]. Both concepts, which we now put together, have already been investigated separately in a Banach space setting; see [25, 16, 12] for Landweber regularization methods, [21, 15] for Landweber–Kaczmarz methods, [14] for inexact Newton-type methods.

Algorithm REGINN (REGularization based on INexact Newton iteration) [24] for solving (1) in Hilbert spaces is a Newton-type algorithm which updates the actual

*Received by the editors June 6, 2013; accepted for publication (in revised form) March 25, 2014; published electronically June 17, 2014.

<http://www.siam.org/journals/sinum/52-3/92395.html>

[†]Department of Mathematics, Karlsruhe Institute of Technology (KIT), D-76128 Karlsruhe, Germany (fabio.margotti@partner.kit.edu, andreas.rieder@kit.edu). The first author was supported by the German Academic Exchange Service (DAAD).

[‡]Department of Mathematics, Federal University of St. Catarina, 88040-900 Florianópolis, Brazil (acgleitao@gmail.com). This author was partially supported by the Brazilian National Research Council CNPq, grant 303098/2009–0.

iterate x_n by adding a correction step

$$(2) \quad x_{n+1} = x_n + s_n.$$

To find the Newton update s_n , we assume F to be continuously Fréchet differentiable with derivative $F': D(F) \rightarrow \mathcal{L}(X, Y)$ and solve approximately the linearized equation

$$(3) \quad A_n s = b_n,$$

where $A_n := F'(x_n)$ and $b_n := y - F(x_n)$. In fact, for a fixed $\mu \in]0, 1[$, s_n is picked such that

$$(4) \quad \|A_n s_n - b_n\| < \mu \|b_n\|.$$

Typically we apply an iterative regularization method to (3) to find an s_n satisfying (4). This iteration is called *inner* iteration and iteration (2) is called *outer* iteration.

To formulate a Kaczmarz version of REGINN we further assume that problem (1) splits into $d \in \mathbb{N}$ “smaller” subproblems, that is, Y factorizes into Banach spaces Y_0, \dots, Y_{d-1} : $Y = Y_0 \times Y_1 \times \dots \times Y_{d-1}$. Accordingly, $F = (F_0, F_1, \dots, F_{d-1})^\top$, $F_j: D(F) \subset X \rightarrow Y_j$, and $y = (y_0, y_1, \dots, y_{d-1})^\top$. Thus, (1) can be written as follows: find $x \in D(F)$ such that

$$(5) \quad F_j(x) = y_j, \quad j = 0, \dots, d-1.$$

The idea is to solve the large-scale system (1) by a cyclic iteration where at each step REGINN is applied to only one of the equations (5). This approach breaks the large system down into d smaller subproblems and thus permits use of all the information contained in the data while avoiding a large system.

Here is a short outline of the paper: in section 2 we recall from [6, 26] needed notation and results concerning the geometry of Banach spaces and Bregman distances. The experienced Banach space user may skip this section. Then, in sections 3 and 4 we introduce the K-REGINN-Landweber method and prove convergence as well as stability properties. Our results generalize and complement the investigations of [14] and [21]. Notably we verify strong convergence of the method from [14] where only weak convergence was established. In the final section we present a variety of numerical experiments for the inverse problem of two dimensional electric impedance tomography. Here we study the performance of K-REGINN-Landweber, compare Hilbert and Banach space settings, as well as different noise models.

2. Basic facts about the geometry of Banach spaces. If the context is clear, we use always a generic constant $C > 0$ even it takes different values, avoiding the unnecessary index enumeration. Sometimes we use also the notation \lesssim . This symbol means $a(x) \lesssim b(x)$ if and only if there exists a positive constant C independent of x such that $a(x) \leq C b(x)$ for all x .

Next we collect needed facts about the geometry of Banach spaces. For proofs and more details we refer to [6].

We define the *modulus of smoothness* of the Banach space X as

$$\rho_X(\tau) := \frac{1}{2} \sup \{ \|x + \tau y\| + \|x - \tau y\| - 2 : \|x\| = \|y\| = 1 \}, \quad \tau \geq 0,$$

and the *modulus of convexity* as

$$\delta_X(\epsilon) := \inf \left\{ 1 - \frac{\|x + y\|}{2} : \|x\| = \|y\| = 1, \|x - y\| \geq \epsilon \right\}, \quad 0 \leq \epsilon \leq 2.$$

The space X is called *uniformly smooth* if $\lim_{\tau \rightarrow 0^+} \frac{\rho_X(\tau)}{\tau} = 0$ and *uniformly convex* if $\delta_X(\epsilon) > 0$ for all $0 < \epsilon \leq 2$. Let $1 < s < \infty$ be fixed. We call X *s-smooth* if $\rho_X(\tau) \leq C_1\tau^s$ for all $\tau \geq 0$ and we call it *s-convex* if $\delta_X(\epsilon) \geq C_2\epsilon^s$ for all $0 \leq \epsilon \leq 2$, where $C_1, C_2 > 0$ are constants. Of course X *s-convex* implies X uniformly convex and X *s-smooth* implies X uniformly smooth.

A continuous and strictly increasing function $\varphi: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that $\varphi(0) = 0$ and $\lim_{t \rightarrow \infty} \varphi(t) = \infty$ is called a *gauge function*. The *duality mapping* associated with the gauge function φ is the mapping $J_\varphi: X \rightarrow 2^{X^*}$ defined by

$$J_\varphi(x) := \{x^* \in X^* : \langle x^*, x \rangle = \|x^*\| \|x\| \text{ and } \|x^*\| = \varphi(\|x\|)\},$$

where $\langle \cdot, \cdot \rangle: X^* \times X \rightarrow \mathbb{R}$ is the duality pairing.

The duality mapping associated with the gauge function $t \mapsto t^{p-1}$ for some $p > 1$ has the special notation J_p

$$J_p(x) := \left\{ x^* \in X^* : \langle x^*, x \rangle = \|x^*\| \|x\| \text{ and } \|x^*\| = \|x\|^{p-1} \right\}.$$

The duality mapping J_2 is called the *normalized* duality mapping. For any two gauge functions φ_1 and φ_2 and $x \in X$ we have the relation $\varphi_1(\|x\|)J_{\varphi_2}(x) = \varphi_2(\|x\|)J_{\varphi_1}(x)$. In particular

$$(6) \quad J_r(x) = \|x\|^{r-t} J_t(x), \quad t, r > 1.$$

A *selection* $j_\varphi: X \rightarrow X^*$ of the duality mapping J_φ is a mapping which satisfies $j_\varphi(x) \in J_\varphi(x)$ for all $x \in X$. If X is uniformly smooth, then it is reflexive. In this case, each duality mapping is single valued and continuous. If X is also uniformly convex, then each duality mapping is bijective with continuous inverse $J_\varphi^{-1} = J_{\varphi^{-1}}^*$, where $J_{\varphi^{-1}}^*: X^* \rightarrow X^{**} \cong X$ is the duality mapping associated with the gauge function φ^{-1} . In the particular case $\varphi(t) = t^{p-1}$ we have $J_p^{-1} = J_{p^*}^*$, where $J_{p^*}^*: X^* \rightarrow X$ is the duality mapping associated with the gauge function $\varphi(t) = t^{p^*-1}$ and p^* is the conjugate of p , i.e., $\frac{1}{p} + \frac{1}{p^*} = 1$.

Suppose that X is uniformly smooth. Then J_p is single valued and we can define the *Bregman distance* $\Delta_p: X \times X \rightarrow [0, \infty[$ as

$$\Delta_p(x, y) := \frac{1}{p} \|y\|^p - \frac{1}{p} \|x\|^p - \langle J_p(x), y - x \rangle.$$

A straightforward calculation shows that

$$(7) \quad \Delta_p(x, y) = \Delta_p(x, z) - \Delta_p(y, z) - \langle J_p(y) - J_p(x), z - y \rangle$$

for all $x, y, z \in X$. Moreover, $\Delta_p(x, y) = 0$ iff $x = y$ and

$$(\Delta_p(x_n, x))_{n \in \mathbb{N}} \text{ uniformly bounded} \implies (x_n)_{n \in \mathbb{N}} \text{ uniformly bounded.}$$

If X is additionally uniformly convex then

$$(8) \quad \lim_{n \rightarrow \infty} \|x_n - x\| = 0 \iff \lim_{n \rightarrow \infty} \Delta_p(x_n, x) = 0$$

for sequences $(x_n)_{n \in \mathbb{N}} \subset X$ and $x \in X$.

For an *s-smooth* Banach space we have that (see [14, inequality (2.2)])

$$(9) \quad \Delta_p(x, y) \leq C_{p,s} \|x - y\|^s \left(\|y\|^{p-s} + \|x - y\|^{p-s} \right)$$

for all $p > 1$, where $C_{p,s} > 0$ is a constant dependent only on p and s .

If X is an s -convex Banach space and $p \leq s$ then

$$\Delta_p(x, y) \geq C \|x - y\|^s (\|x\| + \|y\|)^{p-s}$$

for all $x, y \in X$; $C > 0$ is again a constant dependent only on p and s ; see, e.g., [26, Corollary 2.61(a)]. In this case,

$$(10) \quad \Delta_p(x, y) \geq C \|x - y\|^s$$

for all $x, y \in B_R(0, \|\cdot\|) := \{z \in X : \|z\| \leq R\}$, where $C > 0$ depends only on p, s , and R .

3. The K-REGINN-Landweber method: Definition and first results. Jin [14] considered REGINN using the Landweber regularization as inner iteration in a Banach space framework. The main goal of this section is to define and analyze a Kaczmarz version of his algorithm. As a byproduct of our analysis we improve Jin’s convergence result; see Remark 14 below.

We first introduce some notation and main assumptions. By $[n] := n \bmod d$ we denote the remainder of integer division. Furthermore, we slightly change the notation from the previous sections:

$$b_n := y_{[n]} - F_{[n]}(x_n) \quad \text{and} \quad A_n := F'_{[n]}(x_n).$$

Assumption 1. The right-hand side of (1) and (5) is achievable: there is an $x^+ \in D(F)$ with $y = F(x^+)$.

(a) There exists some $\rho > 0$ such that

$$B_\rho(x^+, \Delta_p) := \{v \in X : \Delta_p(v, x^+) \leq \rho\} \subset D(F).$$

(b) There exists a constant $M \geq 0$ such that

$$\|F'_j(v)\| \leq M \text{ for all } v \in B_\rho(x^+, \Delta_p) \text{ and } j = 0, \dots, d-1.$$

(c) (Tangential cone condition) There exists a constant $0 \leq \eta < 1$ such that

$$\|F_j(v) - F_j(w) - F'_j(w)(v - w)\| \leq \eta \|F_j(v) - F_j(w)\|$$

for all $v, w \in B_\rho(x^+, \Delta_p)$ and $j = 0, \dots, d-1$.

(d) The Banach space X is uniformly smooth and uniformly convex.

As the spaces Y_j are arbitrary for now, the duality mapping $J_r, r > 1$, might not be single valued. Then, $j_r : Y_j \rightarrow Y_j^*$ denotes one selection of J_r . Under Assumption 1 we define the K-REGINN-Landweber method which consists of two iterative schemes. Let $x_0 \in D(F)$ be our starting guess and suppose that the n th iterate $x_n \in D(F)$ is already defined. Then, x_n is updated to x_{n+1} in the *outer iteration* as in (2) by adding the Newton step s_{n,k_n} determined in the *inner iteration*, i.e., $x_{n+1} := x_n + s_{n,k_n}$. The increment s_{n,k_n} is obtained by the Landweber method applied to the linearized system (3): set $u_{n,0} := 0 \in X^*$ and suppose that $u_{n,k}$ is already defined. Then,

$$(11) \quad u_{n,k+1} := u_{n,k} + \omega_{n,k} A_n^* j_r(b_n - A_n s_{n,k}),$$

where

$$s_{n,k} := z_{n,k} - x_n \quad \text{and} \quad z_{n,k} := J_{p^*}^*(J_p(x_n) + u_{n,k}).$$

Further, $\omega_{n,k}$ is a scale factor which might depend on n as well as k , and $A_n^* : Y_{[n]}^* \rightarrow$

X^* is the adjoint operator of A_n . Thus,

$$(12) \quad x_{n+1} = x_n + s_{n,k_n} = z_{n,k_n} = J_{p^*}^*(J_p(x_n) + u_{n,k_n}),$$

where the final (inner) index k_n is determined as follows: choose $k_{\max} \in \mathbb{N}$ and $\mu \in]0, 1[$, set

$$(13) \quad k_{REG} := \min \{k \in \{1, \dots, k_{\max}\} : \|b_n - A_n s_{n,k}\| < \mu \|b_n\|\},$$

using $\min \emptyset = \infty$. Then,

$$(14) \quad k_n = \begin{cases} 0 & : b_n = 0, \\ k_{REG} & : k_{REG} \leq k_{\max}, \\ k_{\max} & : k_{REG} > k_{\max}. \end{cases}$$

Observe that $k_n = 0$ if and only if $b_n = 0$ in which case $x_{n+1} = x_n$. In the unlikely event that $b_{n+i} = 0$ for $i = 0, \dots, d - 1$ we can conclude that x_n is a solution of (5) as $y_j = F_j(x_n)$ for $j = 0, \dots, d - 1$. In any case we will prove that $(x_n)_{n \in \mathbb{N}}$ converges to a solution (see Theorem 5 below).

Remark 2. If $k_{\max} = 1$, then $k_n \in \{0, 1\}$ for all $n \in \mathbb{N}$ and the method assumes the form

$$x_{n+1} = J_{p^*}^*(J_p(x_n) + \omega_n A_n^* j_r(y_{[n]} - F_{[n]}(x_n))),$$

where $\omega_n := \omega_{n,0}$, which is a kind of Kaczmarz–Landweber iteration applied to (1) (in [21] the authors suggested a special way to define the step size ω_n).

If $k_n = k_{REG}$ and $\mu < 1 - 2\eta$ we have a qualified decrease in the nonlinear residual by at least the factor $\Lambda = (\mu + \eta)/(1 - \eta) < 1$. Indeed, standard arguments (see, e.g., [9] and [24, Proof of Lemma 4.1]) lead to

$$(15) \quad \|y_{[n]} - F_{[n]}(x_{n+1})\| < \Lambda \|y_{[n]} - F_{[n]}(x_n)\|.$$

It is this property which accounts for the efficiency of inexact Newton methods and which motivates the definition of k_n . Generically, $k_n = k_{REG}$ (provided k_{\max} sufficiently large), however, we had to introduce k_{\max} because we need an upper bound for k_n in our following analysis.¹ Nevertheless, the important relation

$$(16) \quad \mu \|b_n\| \leq \|b_n - A_n s_{n,k}\| \quad k = 0, \dots, k_n - 1,$$

remains true which allows us to rely on some results of Jin [14].

In the next two theorems we prove that K-REGINN-Landweber (12) is well defined and converges.

THEOREM 3. *Let X and Y be Banach spaces where X is s -convex for some $s > 1$. Choose $1 < p \leq s$ and $r > 1$. Let Assumption 1 hold true and start with $x_0 \in B_\rho(x^+, \Delta_\rho)$. If $\eta < \mu < 1$ and*

$$\omega_{n,k} := \min \left\{ \omega_{n,k}^{(1)}, \omega_{n,k}^{(2)} \right\}$$

with

$$(17) \quad \begin{aligned} \omega_{n,k}^{(1)} &:= \theta_1 \|A_n\|^{-p} \|b_n - A_n s_{n,k}\|^{p-r}, \\ \omega_{n,k}^{(2)} &:= \theta_2 \|A_n\|^{-s} \|z_{n,k}\|^{p-s} \|b_n - A_n s_{n,k}\|^{s-r}, \end{aligned}$$

¹In case $d = 1$ we can actually cope with $k_{\max} = \infty$; cf. Remark 14 below.

for some positive constants θ_1 and θ_2 such that

$$C_0 := 1 - \frac{\eta}{\mu} - C_{p^*,s^*} \left(\theta_1^{p^*-1} + \theta_2^{s^*-1} \right) > 0,$$

where C_{p^*,s^*} is the constant from (9), then the method is well defined and all iterations remain in $B_\rho(x^+, \Delta_p)$. Moreover,

$$(18) \quad \Delta_p(x_{n+1}, x^+) \leq \Delta_p(x_n, x^+)$$

for all $n \in \mathbb{N}$.

Proof. We use an inductive argument. Assume that $x_n \in B_\rho(x^+, \Delta_p)$ is well defined. As k_n is then well defined we have $x_{n+1} \in X$. In view of (16) we can argue exactly as in [14, Theorem 3.1] to show that

$$\Delta_p(z_{n,k+1}, x^+) - \Delta_p(z_{n,k}, x^+) \leq -C_0 \omega_{n,k} \|b_n - A_n s_{n,k}\|^r$$

for all $k = 0, \dots, k_n - 1$. Summing up yields

$$(19) \quad C_0 \sum_{k=0}^{k_n-1} \omega_{n,k} \|b_n - A_n s_{n,k}\|^r \leq \Delta_p(x_n, x^+) - \Delta_p(x_{n+1}, x^+)$$

which first implies (18) and then $x_{n+1} \in B_\rho(x^+, \Delta_p)$. \square

By (18) we find that

$$(20) \quad \Delta_p(x_n, x^+) \leq \Delta_p(x_0, x^+).$$

Hence, $(x_n)_{n \in \mathbb{N}}$ is uniformly bounded.

LEMMA 4. *Adopt all the hypotheses from Theorem 3. Then,*

$$(21) \quad \Delta_p(x_{n+1}, x_n) \leq C (\Delta_p(x_n, x^+) - \Delta_p(x_{n+1}, x^+))$$

for all $n \in \mathbb{N}$.

Proof. By (7),

$$(22) \quad \Delta_p(x_{n+1}, x_n) \leq \Delta_p(x_{n+1}, x^+) - \Delta_p(x_n, x^+) + |\langle J_p(x_n) - J_p(x_{n+1}), x^+ - x_n \rangle|.$$

Further,

$$\begin{aligned} & |\langle J_p(x_n) - J_p(x_{n+1}), x^+ - x_n \rangle| \\ &= |\langle J_p(x_n) - (J_p(x_n) + u_{n,k_n}), x^+ - x_n \rangle| \\ &= |\langle u_{n,k_n}, x_n - x^+ \rangle| = \left| \sum_{k=0}^{k_n-1} \langle u_{n,k+1} - u_{n,k}, x_n - x^+ \rangle \right| \\ &= \left| \sum_{k=0}^{k_n-1} \omega_{n,k} \langle j_r(b_n - A_n s_{n,k}), A_n(x_n - x^+) \rangle \right| \\ &\leq \sum_{k=0}^{k_n-1} \omega_{n,k} \|b_n - A_n s_{n,k}\|^{r-1} \|A_n(x_n - x^+)\| \\ &\leq (\eta + 1) \sum_{k=0}^{k_n-1} \omega_{n,k} \|b_n - A_n s_{n,k}\|^{r-1} \|b_n\|, \end{aligned}$$

where we have used the tangential cone condition from Assumption 1 in the last inequality. By (16) and (19),

$$\begin{aligned} |\langle J_p(x_n) - J_p(x_{n+1}), x^+ - x_n \rangle| &\leq \frac{\eta + 1}{\mu} \sum_{k=0}^{k_n-1} \omega_{n,k} \|b_n - A_n s_{n,k}\|^r \\ &\leq \frac{\eta + 1}{\mu C_0} (\Delta_p(x_n, x^+) - \Delta_p(x_{n+1}, x^+)). \end{aligned}$$

Inserting this bound into (22), we arrive at (21) with $C = \frac{\eta+1}{\mu C_0} - 1 > 0$. \square

THEOREM 5. *Again, let the assumptions of Theorem 3 hold true where $1 < p \leq s \leq r$ and*

$$(23) \quad \omega_{n,k} := \min \left\{ \omega_{n,k}^{(1)}, \omega_{n,k}^{(2)}, \bar{\omega} \right\}$$

with a constant $\bar{\omega} > 0$. Then, the sequence $(x_n)_{n \in \mathbb{N}} \subset X$ converges strongly to a solution of (1). If x^+ is the unique solution in $B_\rho(x^+, \Delta_p)$, then $x_n \rightarrow x^+$ as $n \rightarrow \infty$.

Proof. Theorem 3 remains true under (23). We will prove that the sequence $(x_n)_{n \in \mathbb{N}}$ is a Cauchy sequence. Let $m, l \in \mathbb{N}$ with $m \leq l$ and write $m = m_0 d + m_1$ as well as $l = l_0 d + l_1$, where $m_0, l_0 \in \mathbb{N}$ and $m_1, l_1 \in \{0, \dots, d - 1\}$. Of course $m_0 \leq l_0$. Choose $z_0 \in \{m_0, \dots, l_0\}$ such that

$$(24) \quad \begin{aligned} \sum_{j=0}^{d-1} (\|y_j - F_j(x_{z_0 d+j})\| + \|x_{z_0 d+j+1} - x_{z_0 d+j}\|) \\ \leq \sum_{j=0}^{d-1} (\|y_j - F_j(x_{n_0 d+j})\| + \|x_{n_0 d+j+1} - x_{n_0 d+j}\|) \end{aligned}$$

for all $n_0 \in \{m_0, \dots, l_0\}$, and define $z := z_0 d + z_1$, where $z_1 = l_1$ for $z_0 = l_0$ and $z_1 = d - 1$ otherwise. This setting guarantees that $m \leq z \leq l$. By (20), the sequence $(x_n)_{n \in \mathbb{N}}$ is uniformly bounded. It follows, see (10), that

$$\|x_m - x_l\|^s \leq 2^s (\|x_m - x_z\|^s + \|x_z - x_l\|^s) \lesssim (\Delta_p(x_m, x_z) + \Delta_p(x_l, x_z)).$$

In view of (7) we obtain

$$(25) \quad \begin{aligned} \|x_m - x_l\|^s &\lesssim (\Delta_p(x_m, x^+) - \Delta_p(x_z, x^+)) + (\Delta_p(x_l, x^+) - \Delta_p(x_z, x^+)) \\ &\quad + |\langle J_p(x_z) - J_p(x_m), x^+ - x_z \rangle| + |\langle J_p(x_z) - J_p(x_l), x^+ - x_z \rangle|. \end{aligned}$$

Due to the monotonicity (18) the first two terms in the right-hand side converge to zero as $m, l \rightarrow \infty$. We estimate the last two:

$$\begin{aligned} |\langle J_p(x_z) - J_p(x_m), x^+ - x_z \rangle| &= \left| \sum_{n=m}^{z-1} \langle J_p(x_{n+1}) - J_p(x_n), x^+ - x_z \rangle \right| \\ &\leq \sum_{n=m}^{z-1} |\langle u_{n,k_n}, x_z - x^+ \rangle|. \end{aligned}$$

Analogously,

$$|\langle J_p(x_z) - J_p(x_l), x^+ - x_z \rangle| \leq \sum_{n=z}^{l-1} |\langle u_{n,k_n}, x_z - x^+ \rangle|.$$

Then, the last two terms in (25)

$$f(z, m, l) := |\langle J_p(x_z) - J_p(x_m), x_z - x^+ \rangle| + |\langle J_p(x_z) - J_p(x_l), x_z - x^+ \rangle|$$

are estimated by

$$(26) \quad f(z, m, l) \leq \sum_{n=m}^{l-1} |\langle u_{n,k_n}, x_z - x^+ \rangle|.$$

We proceed as in the proof of Lemma 4 and get

$$(27) \quad \begin{aligned} |\langle u_{n,k_n}, x_z - x^+ \rangle| &\leq \sum_{k=0}^{k_n-1} |\langle u_{n,k+1} - u_{n,k}, x_z - x^+ \rangle| \\ &= \sum_{k=0}^{k_n-1} |\omega_{n,k} \langle j_r(b_n - A_n s_{n,k}), A_n(x_z - x^+) \rangle| \\ &\leq \sum_{k=0}^{k_n-1} \omega_{n,k} \|b_n - A_n s_{n,k}\|^{r-1} \|A_n(x_z - x^+)\|. \end{aligned}$$

Employing Assumption 1(c) yields

$$(28) \quad \begin{aligned} \|A_n(x_z - x^+)\| &\leq \|A_n(x_n - x^+)\| + \|A_n(x_z - x_n)\| \\ &\leq \|b_n\| + \|b_n - A_n(x^+ - x_n)\| + \|F_{[n]}(x_z) - F_{[n]}(x_n)\| \\ &\quad + \left\| F_{[n]}(x_z) - F_{[n]}(x_n) - F'_{[n]}(x_n)(x_z - x_n) \right\| \\ &\leq (\eta + 1) (\|b_n\| + \|F_{[n]}(x_z) - F_{[n]}(x_n)\|) \\ &\leq (\eta + 1) (2\|b_n\| + \|y_{[n]} - F_{[n]}(x_z)\|). \end{aligned}$$

Further,

$$\begin{aligned} \|y_{[n]} - F_{[n]}(x_z)\| &= \|y_{[n]} - F_{[n]}(x_{z_0 d + d - 1})\| \\ &\leq \|y_{[n]} - F_{[n]}(x_{z_0 d + [n]})\| + \sum_{j=[n]}^{d-2} \|F_{[n]}(x_{z_0 d + j + 1}) - F_{[n]}(x_{z_0 d + j})\| \\ &\leq \|y_{[n]} - F_{[n]}(x_{z_0 d + [n]})\| \\ &\quad + \frac{1}{1 - \eta} \sum_{j=[n]}^{d-2} \left\| F'_{[n]}(x_{z_0 d + j})(x_{z_0 d + j + 1} - x_{z_0 d + j}) \right\| \\ &\leq \left(1 + \frac{M}{1 - \eta} \right) \sum_{j=0}^{d-1} (\|y_j - F_j(x_{z_0 d + j})\| + \|x_{z_0 d + j + 1} - x_{z_0 d + j}\|). \end{aligned}$$

We write $n = n_0 d + n_1$ for some $n_0 \in \{m_0, \dots, l_0\}$ and $n_1 \in \{0, \dots, d - 1\}$. Then, from the definition of z_0 (24),

$$(29) \quad \|y_{[n]} - F_{[n]}(x_z)\| \leq \left(1 + \frac{M}{1 - \eta} \right) \sum_{j=0}^{d-1} (\|y_j - F_j(x_{n_0 d + j})\| + \|x_{n_0 d + j + 1} - x_{n_0 d + j}\|).$$

Recalling (16), inserting (29) in (28), (28) in (27), and (27) in (26), we arrive at

$$(30) \quad f(z, m, l) \lesssim \left(\sum_{n=m}^{l-1} \sum_{k=0}^{k_n-1} \omega_{n,k} \|b_n - A_n s_{n,k}\|^r \right) + g(z, m, l) + h(z, m, l)$$

with the abbreviations

$$g(z, m, l) := \sum_{n=m}^{l-1} \sum_{k=0}^{k_n-1} \omega_{n,k} \|b_n - A_n s_{n,k}\|^{r-1} \sum_{j=0}^{d-1} \|y_j - F_j(x_{n_0 d+j})\|,$$

$$h(z, m, l) := \sum_{n=m}^{l-1} \sum_{k=0}^{k_n-1} \omega_{n,k} \|b_n - A_n s_{n,k}\|^{r-1} \sum_{j=0}^{d-1} \|x_{n_0 d+j+1} - x_{n_0 d+j}\|.$$

The first term in the right-hand side of (30) can be estimated by (19).

We estimate the last two terms. By

$$\|b_n\| \leq \frac{M}{1-\eta} \|x^+ - x_n\|$$

we see that $(b_n)_{n \in \mathbb{N}}$ is uniformly bounded because $(x_n)_{n \in \mathbb{N}}$ is uniformly bounded. From the definition of $\omega_{n,k}$ ((23) and (17)), Assumption 1(b), $1 < p \leq s \leq r$, $s_{n,0} = 0$, and $z_{n,0} = x_n$, we deduce that

$$\omega_{n,0}^{\frac{1}{r-1}} = \min \left\{ \theta_1 \|A_n\|^{-p} \|b_n\|^{p-r}, \theta_2 \|A_n\|^{-s} \|x_n\|^{p-s} \|b_n\|^{s-r}, \bar{\omega} \right\}^{\frac{1}{r-1}} \geq C_1 > 0$$

for all $n \in \mathbb{N}$. Moreover,

$$(31) \quad \|b_n\| \leq \frac{1}{C_1} \omega_{n,0}^{\frac{1}{r-1}} \|b_n\| = \frac{1}{C_1} \omega_{n,0}^{\frac{1}{r-1}} \|b_n - A_n s_{n,0}\| \leq \frac{1}{C_1} v_n,$$

where $v_n := \sum_{k=0}^{k_n-1} \omega_{n,k}^{\frac{1}{r-1}} \|b_n - A_n s_{n,k}\|$. As $k_n \leq k_{\max}$,

$$\sum_{k=0}^{k_n-1} \omega_{n,k} \|b_n - A_n s_{n,k}\|^{r-1} = \sum_{k=0}^{k_n-1} \left(\omega_{n,k}^{\frac{1}{r-1}} \|b_n - A_n s_{n,k}\| \right)^{r-1} \leq k_{\max} v_n^{r-1}.$$

Then, we estimate the second term in the right-hand side of (30) using (31) according to

$$(32) \quad g(z, m, l) \leq \frac{k_{\max}}{C_1} \sum_{n_0=m_0}^{l_0} \left(\sum_{n_1=0}^{d-1} v_{n_0 d+n_1}^{r-1} \sum_{j=0}^{d-1} v_{n_0 d+j} \right)$$

$$\lesssim \sum_{n_0=m_0 d}^{l_0} \left(\sum_{n_1=0}^{d-1} v_{n_0 d+n_1}^r \right) = \sum_{n=m_0 d}^{l_0 d+d-1} v_n^r$$

$$= \sum_{n=m_0 d}^{l_0 d+d-1} \left(\sum_{k=0}^{k_n-1} \omega_{n,k}^{\frac{1}{r-1}} \|b_n - A_n s_{n,k}\| \right)^r$$

$$\lesssim \sum_{n=m_0 d}^{l_0 d+d-1} \sum_{k=0}^{k_n-1} \omega_{n,k} \|b_n - A_n s_{n,k}\|^r,$$

where we have used $\omega_{n,k} \leq \bar{\omega}$ in the last step to establish $\omega_{n,k}^{\frac{r}{r-1}} \lesssim \omega_{n,k}$.

Similarly, we bound the rightmost term in (30):

$$\begin{aligned}
 (33) \quad h(z, m, l) &\leq k_{\max} \sum_{n_0=m_0}^{l_0} \left(\sum_{n_1=0}^{d-1} v_{n_0 d+n_1}^{r-1} \sum_{j=0}^{d-1} \|x_{n_0 d+j+1} - x_{n_0 d+j}\| \right) \\
 &\lesssim \sum_{n_0=m_0}^{l_0} \sum_{n_1=0}^{d-1} (v_{n_0 d+n_1}^r + \|x_{n_0 d+n_1+1} - x_{n_0 d+n_1}\|^r) \\
 &= \sum_{n=m_0 d}^{l_0 d+d-1} (v_n^r + \|x_{n+1} - x_n\|^r) \\
 &\lesssim \sum_{n=m_0 d}^{l_0 d+d-1} \sum_{k=0}^{k_n-1} \omega_{n,k} \|b_n - A_n s_{n,k}\|^r + \sum_{n=m_0 d}^{l_0 d+d-1} \|x_{n+1} - x_n\|^r.
 \end{aligned}$$

Now concentrate on the sum on the right. Using (10) and Lemma 4 we obtain

$$\|x_{n+1} - x_n\|^s \lesssim \Delta_p(x_n, x_{n+1}) \lesssim (\Delta_p(x_n, x^+) - \Delta_p(x_{n+1}, x^+)).$$

As $r \geq s > 1$, we have for m, l large enough that

$$\sum_{n=m_0 d}^{l_0 d+d-1} \|x_{n+1} - x_n\|^r \leq \sum_{n=m_0 d}^{l_0 d+d-1} \|x_{n+1} - x_n\|^s \lesssim (\Delta_p(x_{m_0 d}, x^+) - \Delta_p(x_{l_0 d+d}, x^+)).$$

Using this fact, inserting (33) and (32) in (30), (30) in (25), taking (19) into account, we finally end up with

$$\begin{aligned}
 \|x_m - x_l\|^s &\lesssim (\Delta_p(x_m, x^+) - \Delta_p(x_z, x^+)) + (\Delta_p(x_l, x^+) - \Delta_p(x_z, x^+)) \\
 &\quad + (\Delta_p(x_{m_0 d}, x^+) - \Delta_p(x_{l_0 d+d}, x^+)).
 \end{aligned}$$

By monotonicity (18), we conclude that $\Delta_p(x_q, x^+) \rightarrow \beta \geq 0$ as $q \rightarrow \infty$. Then, the right side of the above inequality converges to zero as $m, l \rightarrow \infty$, which proves $(x_n)_{n \in \mathbb{N}}$ to be a Cauchy sequence. As X is complete, it converges to some $a \in X$. From (31) and (19) we obtain

$$(34) \quad \sum_{n=0}^{\infty} \|b_n\|^r \lesssim \sum_{n=0}^{\infty} v_n^r \lesssim \sum_{n=0}^{\infty} \sum_{k=0}^{k_n-1} \omega_{n,k} \|b_n - A_n s_{n,k}\|^r < \infty,$$

that is, $\|y_{[n]} - F_{[n]}(x_n)\| = \|b_n\| \rightarrow 0$ as $n \rightarrow \infty$. Since the F_j 's are continuous, $y_j = F_j(a)$ and a is a solution of (5) and (1). If (1) has only one solution in $B_\rho(x^+, \Delta_p)$, then $a = x^+$. \square

Remark 6. The hypothesis $s \leq r$ is a technicality needed in our proof of Theorem 5. It is not necessary for $d = 1$ (this is Jin's method [14]). However, $s \leq r$ is not a strong requirement anyway. Indeed, we can realize the duality mapping J_r by (6); see section 5 below. Moreover, as $r \geq s$ we can bound the scale factors $\omega_{n,k}$ (11) from below by a positive constant which facilitates the implementation of the K-REGINN-Landweber method; see Remark 15 below.

4. The K-REGINN-Landweber method: Regularization property.

4.1. Noise in the data. We suppose now that the right-hand sides $y_j = F_j(x^+)$ in (5) are not available but noisy versions satisfying

$$(35) \quad \|y_j - y_j^{\delta_j}\| \leq \delta_j$$

will be. The nonnegative *noise levels* $\delta_j, j = 0, \dots, d - 1$, are assumed to be known. Moreover, following [14] we even allow the operators F_j and their derivatives F'_j to be perturbed (thus modeling, for instance, discretization errors). Their perturbed versions are F_j^h and G_j^h , respectively, where

$$(36) \quad \|F_j^h(x) - F_j(x)\| \leq \beta_{h,j} \text{ and } \|G_j^h(x) - F'_j(x)\| \leq \pi_{h,j}$$

for all $x \in B_\rho(x^+, \Delta_\rho)$ and $j = 0, \dots, d - 1$. Further, all $\beta_{h,j}$'s and $\pi_{h,j}$'s converge to zero as $h \rightarrow 0$. In principle we could perform a convergence analysis with respect to δ and h independently of each other. For the ease of presentation, however, we couple h to δ such that $h \rightarrow 0$ whenever $\delta \rightarrow 0$. Here,

$$(37) \quad \delta := \max \{\delta_j : j = 0, \dots, d - 1\} > 0.$$

To formulate the method for noisy data, we introduce new notation

$$A_n^\delta := G_{[n]}^h(x_n^\delta) \quad \text{and} \quad b_n^\delta := y_{[n]}^{\delta_{[n]}} - F_{[n]}^h(x_n^\delta).$$

With this notation the method is defined exactly as in (12) but all quantities are replaced by their respective noisy counterparts (indicated by a superscript δ_i or δ):

$$x_{n+1}^\delta = x_n^\delta + s_{n,k_n^\delta}^\delta = z_{n,k_n^\delta}^\delta = J_{p^*}^*(J_p(x_n^\delta) + u_{n,k_n^\delta}^\delta), \quad x_0^\delta = x_0 \in D(F),$$

where

$$(38) \quad u_{n,k+1}^\delta := u_{n,k}^\delta + \omega_{n,k}^\delta A_n^{\delta,*} j_r (b_n^\delta - A_n^\delta s_{n,k}^\delta), \quad u_{n,0}^\delta = 0,$$

with $A_n^{\delta,*} = (A_n^\delta)^*$. For the definition of k_n^δ see (14) where we modify the case $b_n = 0$ to a discrepancy principle to avoid noise amplification in the outer iteration: choose $R > 0$ such that

$$(39) \quad B_\rho(x^+, \Delta_\rho) \subset B_R(x^+, \|\cdot\|)$$

and fix $\tau > 1$. If for some $n \in \mathbb{N}$ we have

$$(40) \quad \|b_n^\delta\| \leq \tau (\delta_{[n]} + \beta_{h,[n]} + R\pi_{h,[n]})$$

then we define $k_n^\delta := 0$ resulting in $x_{n+1}^\delta = x_n^\delta$. We stop the outer iteration as soon as (40) is satisfied d consecutive times. Our approximate solution of (1) is then x_N^δ , where $N = N(\delta)$ is the smallest number² which satisfies

$$(41) \quad \left\| y_j^{\delta_j} - F_j^h(x_N^\delta) \right\| \leq \tau (\delta_j + \beta_{h,j} + R\pi_{h,j}), \quad j = 0, \dots, d - 1;$$

see Algorithm 1 for an implementation in pseudocode. The monotone decrease of the residuals (15) applies accordingly in the noisy setting of this section.

This method is well defined and terminates: A result similar to Theorem 3 holds true.

²The number N is chosen by a posteriori strategy; it thus depends actually on δ and y^δ : $N = N(\delta, y^\delta)$. But we stick to the simpler notation $N = N(\delta)$.

ALGORITHM 1. K-REGINN-LANDWEBER.

Input: x_N^δ ; $(y_i^{\delta_i}, \delta_i)$; F_i^h ; $\beta_{h,i}$; G_i^h ; $\pi_{h,i}$, $i = 0, \dots, d-1$; μ ; k_{\max} ; R ; τ ;

Output: x_N^δ with $\|y_i^{\delta_i} - F_i(x_N^\delta)\| \leq \tau(\delta_i + \beta_{h,i} + R\pi_{h,i})$, $i = 0, \dots, d-1$;

$\ell := 0$; $x_0 := x_N^\delta$; $c := 0$;

while $c < d$ **do**

for $i = 0 : d-1$ **do**

$n := \ell d + i$;

$b_n^\delta := y_i^{\delta_i} - F_i^h(x_n^\delta)$; $A_n^\delta := G_i^h(x_n^\delta)$;

if $\|b_n^\delta\| \leq \tau(\delta_i + \beta_{h,i} + R\pi_{h,i})$ **then**

$x_{n+1}^\delta := x_n^\delta$; $c := c + 1$;

else

$k := 0$; $u_{n,0}^\delta := 0$; $s_{n,0}^\delta := 0$;

repeat

 choose $\omega_{n,k}^\delta$; $u_{n,k+1}^\delta := u_{n,k}^\delta + \omega_{n,k}^\delta A_n^{\delta,*} j_r(b_n^\delta - A_n^\delta s_{n,k}^\delta)$;

$k := k + 1$; $s_{n,k}^\delta := J_{p^*}^*(J_p(x_n^\delta) + u_{n,k}^\delta) - x_n^\delta$;

until $\|b_n^\delta - A_n^\delta s_{n,k}^\delta\| < \mu \|b_n^\delta\|$ or $k = k_{\max}$

$x_{n+1}^\delta := x_n^\delta + s_{n,k}^\delta$; $c := 0$;

end if

end for

$\ell := \ell + 1$;

end while

$x_N^\delta := x_{\ell d - c}^\delta$;

THEOREM 7. *Let X and Y be Banach spaces where X is s -convex for some $s > 1$. Choose $1 < p \leq s$ and $r > 1$. Let Assumption 1 hold true and start with $x_0 \in B_\rho(x^+, \Delta_p)$.*

If $\eta < \mu < 1$, $\tau > \frac{1+\eta}{\mu-\eta}$, and

$$\omega_{n,k}^\delta := \min \left\{ \omega_{n,k}^{\delta,(1)}, \omega_{n,k}^{\delta,(2)} \right\},$$

where

$$(42) \quad \begin{aligned} \omega_{n,k}^{\delta,(1)} &:= \theta_1 \|A_n^\delta\|^{-p} \|b_n^\delta - A_n^\delta s_{n,k}^\delta\|^{p-r}, \\ \omega_{n,k}^{\delta,(2)} &:= \theta_2 \|A_n^\delta\|^{-s} \|z_{n,k}^\delta\|^{p-s} \|b_n^\delta - A_n^\delta s_{n,k}^\delta\|^{s-r}, \end{aligned}$$

for some positive constants θ_1 and θ_2 satisfying

$$c_0 := 1 - \frac{\eta}{\mu} - \frac{1+\eta}{\mu\tau} - C_{p^*,s^*} \left(\theta_1^{p^*-1} + \theta_2^{s^*-1} \right) > 0$$

with C_{p^*,s^*} from (9), then the method is well defined and all iterations remain in $B_\rho(x^+, \Delta_p)$. Moreover, the algorithm terminates with some number $N = N(\delta) \in \mathbb{N}$ and

$$(43) \quad \Delta_p(x_{n+1}^\delta, x^+) \leq \Delta_p(x_n^\delta, x^+)$$

for all $n \leq N - 1$. Furthermore,

$$(44) \quad \|x_n^\delta\| \leq C \text{ for all } n \leq N$$

with $C > 0$ independent of n, N , and δ .

Proof. We use an inductive argument again. Suppose that x_n^δ is well defined in $B_\rho(x^+, \Delta_p)$ while the stopping criterion (41) is not met yet, that is, there is a $k \in \{1, \dots, d\}$ such that $x_{n-k}^\delta \neq x_n^\delta$. We will show that x_{n+1}^δ is well defined in $B_\rho(x^+, \Delta_p)$.

If x_n^δ satisfies (40), then $x_{n+1}^\delta = x_n^\delta$ is well defined and belongs to $B_\rho(x^+, \Delta_p)$. Of course in this case the inequality (43) trivially holds. If x_n^δ does not satisfy (40), then $k_n^\delta \geq 1$ and x_{n+1}^δ is well defined in X . We follow again [14, Theorem 3.2] to validate

$$\Delta_p(z_{n,k+1}^\delta, x^+) - \Delta_p(z_{n,k}^\delta, x^+) \leq -c_0 \omega_{n,k}^\delta \|b_n^\delta - A_n^\delta s_{n,k}^\delta\|^r$$

for all $k = 0, \dots, k_n^\delta - 1$. Then

$$c_0 \sum_{k=0}^{k_n^\delta-1} \omega_{n,k}^\delta \|b_n^\delta - A_n^\delta s_{n,k}^\delta\|^r \leq \Delta_p(x_n^\delta, x^+) - \Delta_p(x_{n+1}^\delta, x^+),$$

which proves (43). Hence, $x_{n+1}^\delta \in B_\rho(x^+, \Delta_p)$. As in the proof of Theorem 3.2 in [14] we obtain that $\omega_{n,k}^\delta \|b_n^\delta - A_n^\delta s_{n,k}^\delta\|^r \geq C$ for those x_n^δ which do not satisfy (40). The constant $C > 0$ is independent of n and k . In this case,

$$c_0 C k_n^\delta \leq c_0 \sum_{k=0}^{k_n^\delta-1} \omega_{n,k}^\delta \|b_n^\delta - A_n^\delta s_{n,k}^\delta\|^r \leq \Delta_p(x_n^\delta, x^+) - \Delta_p(x_{n+1}^\delta, x^+).$$

If x_n^δ satisfies (40), then $k_n^\delta = 0$ and the above inequality holds trivially true. Assume that Algorithm 1 does not terminate. Then, we sum up the above inequality from $n = 0$ to $n = l - 1$ for arbitrary $l \in \mathbb{N}$ to yield

$$c_0 C \sum_{n=0}^{l-1} k_n^\delta \leq \Delta_p(x_0^\delta, x^+) - \Delta_p(x_l^\delta, x^+) < \Delta_p(x_0^\delta, x^+).$$

Thus, $\sum_{n=0}^\infty k_n^\delta < \infty$ which is true if and only if the sequence $(k_n^\delta)_n$ is finite. Therefore the termination index $N = N(\delta)$ is well defined.

From (43),

$$\Delta_p(x_n^\delta, x^+) \leq \Delta_p(x_0, x^+) < \infty, \quad n \leq N,$$

which implies that $\|x_n^\delta\| \leq C$ for all $n \leq N$, with some $C > 0$ independent of n, N , and δ . \square

Weak convergence is now a matter of standard arguments; see, e.g., [20, Corollary 3.5].

THEOREM 8. *Let all the assumptions of Theorem 7 hold true. If the operators $F_j, j = 0, \dots, d - 1$, are weakly sequentially closed then for any sequences $(y_j^{(\delta_j)_i})_{i \in \mathbb{N}}, (F_j^{h_i})_{i \in \mathbb{N}}$, and $(G_j^{h_i})_{i \in \mathbb{N}}$ with $\delta^{(i)} = \max\{(\delta_j) : j = 0, \dots, d - 1\} \rightarrow 0$ (consequently $h_i \rightarrow 0$) as $i \rightarrow \infty$, the sequence $(x_{N(\delta^{(i)})}^{\delta^{(i)}})_{i \in \mathbb{N}}$ contains a subsequence that converges weakly to a solution of (1) in $B_\rho(x^+, \Delta_p)$. If x^+ is the unique solution of (1) in $B_\rho(x^+, \Delta_p)$, then $(x_{N(\delta)}^\delta)_{\delta > 0}$ converges weakly to x^+ as $\delta = \max\{\delta_j : j = 0, \dots, d - 1\} \rightarrow 0$.*

4.2. Strong convergence. Here we investigate norm convergence of the family $(x_{N(\delta)}^\delta)_{\delta>0}$ when the noise level δ (37) tends to zero. To reduce the notational burden we will not use the δ_j 's in their original meaning (35) anymore. From now on we denote by (δ_i) a sequence of noise levels as defined in (37), i.e.,

$$\delta_i := \max\{(\delta_j)_i : j = 0, \dots, d - 1\}.$$

We prove strong convergence along the line of arguments from [10]; see also [20, section 3.2]. To this end we suppose subsequently that x^+ is the unique solution of (1) in $B_\rho(x^+, \Delta_p)$ and we define the sets \tilde{X}_n .

DEFINITION 9. Set $\tilde{X}_0 := \{x_0\}$. Suppose that the set \tilde{X}_n is already defined and that it is finite. We define \tilde{X}_{n+1} by the following procedure: for each $\xi_n \in \tilde{X}_n$, define $\sigma_{n,0} := 0 \in X^*$ and $\sigma_{n,k+1}$ like $u_{n,k+1}$ in (11) but with ξ_n in place of x_n and $\sigma_{n,k}$ in place of $u_{n,k}$, i.e.,

$$\sigma_{n,k+1} := \sigma_{n,k} + \omega_{n,k}^{\xi_n} F'_{[n]}(\xi_n)^* j_r \left[\tilde{b}_n - F'_{[n]}(\xi_n) s_{n,k}^{\xi_n} \right],$$

where $\tilde{b}_n := y_{[n]} - F_{[n]}(\xi_n)$, $s_{n,k}^{\xi_n} := J_{p^*}(J_p(\xi_n) + \sigma_{n,k}) - \xi_n$, and $\omega_{n,k}^{\xi_n}$ as defined in (23) and (17) (with ξ_n and $\sigma_{n,k}$ in place of x_n and $u_{n,k}$, respectively).

Define

$$(45) \quad k_{REG}(\xi_n) := \min \left\{ k \in \{1, \dots, k_{\max}\} : \left\| \tilde{b}_n - F'_{[n]}(\xi_n) s_{n,k}^{\xi_n} \right\| < \mu \left\| \tilde{b}_n \right\| \right\}$$

and

$$(46) \quad k_n(\xi_n) := \begin{cases} 0 & : \tilde{b}_n = 0, \\ k_{REG}(\xi_n) & : k_{REG}(\xi_n) \leq k_{\max}, \\ k_{\max} & : k_{REG}(\xi_n) > k_{\max}. \end{cases}$$

Then, $\xi_n + s_{n,k_n(\xi_n)}^{\xi_n} \in \tilde{X}_{n+1}$. Furthermore, if

$$(47) \quad \left\| \tilde{b}_n - F'_{[n]}(\xi_n) s_{n,k_n(\xi_n)-\ell}^{\xi_n} \right\| = \mu \left\| \tilde{b}_n \right\| \quad \text{for } \ell = 1, \dots, \ell_n(\xi_n) < k_n(\xi_n),$$

then the elements $\xi_n + s_{n,k_n(\xi_n)-\ell}^{\xi_n}$, $\ell = 1, \dots, \ell_n(\xi_n)$, belong to the set \tilde{X}_{n+1} as well.

We call $\xi_n \in \tilde{X}_n$ the predecessor of $\xi_n + s_{n,k_n(\xi_n)-\ell}^{\xi_n} \in \tilde{X}_{n+1}$ for $\ell = 0, 1, \dots, \ell_n(\xi_n)$, and these ones successors of ξ_n .

Of course $x_n \in \tilde{X}_n$ and \tilde{X}_n is finite for all $n \in \mathbb{N}$. Moreover, from (18),

$$(48) \quad \Delta_p(\xi_{n+1}, x^+) \leq \Delta_p(\xi_n, x^+)$$

whenever $\xi_{n+1} \in \tilde{X}_{n+1}$ is a successor of $\xi_n \in \tilde{X}_n$.

In a certain sense the sequence $(x_n^{\delta_i})_{i \in \mathbb{N}}$ converges to the set \tilde{X}_n provided the image spaces Y_j share further properties.

LEMMA 10. Let all the assumptions of Theorem 7 hold true where

$$\omega_{n,k}^\delta := \min \left\{ \omega_{n,k}^{\delta,(1)}, \omega_{n,k}^{\delta,(2)}, \bar{\omega} \right\}$$

with a constant $\bar{\omega} > 0$. Additionally, let all Y_j 's be uniformly smooth.

If $\delta_i \rightarrow 0$ as $i \rightarrow \infty$, then for $n \leq N(\delta_i)$ with $\delta_i > 0$ sufficiently small, the sequence $(x_n^{\delta_i})_{i \in \mathbb{N}}$ splits into convergent subsequences, all of which converge to an element of \tilde{X}_n .

Proof. We prove the statement by induction. For $n = 0$, $x_0^{\delta_i} = x_0 \rightarrow x_0 \in \widetilde{X}_0$ as $i \rightarrow \infty$. Now, suppose that for some $n \in \mathbb{N}$ with $n + 1 \leq N(\delta_i)$ for i large enough, $(x_n^{\delta_i})_{i \in \mathbb{N}}$ splits into convergent subsequences, all of which converge to an element of \widetilde{X}_n . To simplify the notation, let $(x_n^{\delta_i})_{i \in \mathbb{N}}$ itself be a subsequence which converges to an element of \widetilde{X}_n :

$$(49) \quad \lim_{i \rightarrow \infty} x_n^{\delta_i} = \xi_n, \text{ where } \xi_n \in \widetilde{X}_n.$$

We have to prove that the sequence $(x_{n+1}^{\delta_i})_{i \in \mathbb{N}}$ splits in convergent subsequences, each one converging to an element of \widetilde{X}_{n+1} . To this end, we verify by induction with respect to k that

$$(50) \quad u_{n,k}^{\delta_i} \rightarrow \sigma_{n,k} \text{ as } i \rightarrow \infty \text{ for all } k \leq k_n(\xi_n).$$

In fact, for $k = 0$, $u_{n,0}^{\delta_i} = 0 = \sigma_{n,0} \rightarrow \sigma_{n,0}$ as $i \rightarrow \infty$. If we suppose for some $k < k_n(\xi_n) - 1$ that $u_{n,k}^{\delta_i} \rightarrow \sigma_{n,k}$ as $i \rightarrow \infty$, then, as Y_j is uniformly smooth, the selection $j_r: Y_j \rightarrow Y_j^*$ is unique and continuous. Further, the mappings $J_p, J_p^*, F_j,$ and F'_j are also continuous which implies together with (36) and $\lim_{i \rightarrow \infty} x_n^{\delta_i} = \xi_n$ that $s_{n,k}^{\delta_i} \rightarrow s_{n,k}^{\xi_n}$ and $\omega_{n,k}^{\delta_i} \rightarrow \omega_{n,k}^{\xi_n}$ as $i \rightarrow \infty$. Thus,

$$u_{n,k+1}^{\delta_i} \xrightarrow{i \rightarrow \infty} \sigma_{n,k} + \omega_{n,k}^{\xi_n} F'_{[n]}(\xi_n)^* j_r \left[\widetilde{b}_n - F'_{[n]}(\xi_n) s_{n,k}^{\xi_n} \right] = \sigma_{n,k+1}$$

yielding (50). With a similar argument,

$$(51) \quad s_{n,k}^{\delta_i} \xrightarrow{i \rightarrow \infty} s_{n,k}^{\xi_n} \text{ for all } k \leq k_n(\xi_n)$$

resulting in

$$(52) \quad \lim_{i \rightarrow \infty} \left\| b_n^{\delta_i} - F'_{[n]}(x_n^{\delta_i}) s_{n,k}^{\delta_i} \right\| = \left\| \widetilde{b}_n - F'_{[n]}(\xi_n) s_{n,k}^{\xi_n} \right\| \text{ for all } k \leq k_n(\xi_n).$$

Now we have to differentiate three cases.

Case 1. $1 \leq k_n(\xi_n) = k_{REG}(\xi_n)$. From the definition (45) of $k_{REG}(\xi_n)$ we have

$$\left\| \widetilde{b}_n - F'_{[n]}(\xi_n) s_{n,k_n(\xi_n)}^{\xi_n} \right\| < \mu \left\| \widetilde{b}_n \right\|.$$

It follows from (52) that, for i large enough,

$$\left\| b_n^{\delta_i} - F'_{[n]}(x_n^{\delta_i}) s_{n,k_n(\xi_n)}^{\delta_i} \right\| < \mu \left\| b_n^{\delta_i} \right\|,$$

which implies in view of (14) that $k_n^{\delta_i} \leq k_n(\xi_n)$. Now, if

$$(53) \quad \left\| \widetilde{b}_n - F'_{[n]}(\xi_n) s_{n,k_n(\xi_n)-1}^{\xi_n} \right\| > \mu \left\| \widetilde{b}_n \right\|,$$

then, for i large enough,

$$\left\| b_n^{\delta_i} - F'_{[n]}(x_n^{\delta_i}) s_{n,k_n(\xi_n)-1}^{\delta_i} \right\| > \mu \left\| b_n^{\delta_i} \right\|,$$

which implies that $k_n(\xi_n) - 1 < k_n^{\delta_i}$ and then $k_n^{\delta_i} = k_n(\xi_n)$. It follows from (51) that

$$(54) \quad \lim_{i \rightarrow \infty} x_{n+1}^{\delta_i} = \lim_{i \rightarrow \infty} \left(x_n^{\delta_i} + s_{n,k_n^{\delta_i}}^{\delta_i} \right) = \xi_n + \lim_{i \rightarrow \infty} s_{n,k_n(\xi_n)}^{\delta_i} = \xi_n + s_{n,k_n(\xi_n)}^{\xi_n} \in \widetilde{X}_{n+1}.$$

But, if (53) is not true, let $\ell_n(\xi_n)$ be the largest number such that (47) holds. Then,

$$\left\| \tilde{b}_n - F'_{[n]}(\xi_n) s_{n, k_n(\xi_n) - (\ell_n(\xi_n) + 1)}^{\xi_n} \right\| > \mu \left\| \tilde{b}_n \right\|$$

and so we have $k_n(\xi_n) - \ell_n(\xi_n) \leq k_n^{\delta_i}$ which implies $k_n^{\delta_i} \in \{k_n(\xi_n) - \ell_n(\xi_n), \dots, k_n(\xi_n)\}$. This means that the sequence $(k_n^{\delta_i})_{i \in \mathbb{N}}$ has the limit points $k_n(\xi_n) - \ell_n(\xi_n), \dots, k_n(\xi_n)$ and accordingly the sequence $(x_{n+1}^{\delta_i})_{i \in \mathbb{N}}$ splits into $\ell_n(\xi_n) + 1$ convergent subsequences, each one converging to an element of \tilde{X}_{n+1} by definition of this set.

Case 2. $k_n(\xi_n) = k_{\max}$. In this case,

$$(55) \quad \left\| \tilde{b}_n - F'_{[n]}(\xi_n) s_{n, k_n(\xi_n)}^{\xi_n} \right\| \geq \mu \left\| \tilde{b}_n \right\|.$$

If this inequality is strict, due to (52), for i large enough, we have

$$\left\| b_n^{\delta_i} - F'_{[n]}(x_n^{\delta_i}) s_{n, k_n(\xi_n)}^{\delta_i} \right\| > \mu \left\| b_n^{\delta_i} \right\|$$

which implies that $k_n^{\delta_i} \geq k_n(\xi_n) = k_{\max}$ and then $k_n^{\delta_i} = k_{\max}$. As in (54), $\lim_{i \rightarrow \infty} x_{n+1}^{\delta_i} \in \tilde{X}_{n+1}$.

If equality holds in (55) we consider again the number $\ell_n(\xi_n)$ of (47) to validate $k_n(\xi_n) - \ell_n(\xi_n) \leq k_n^{\delta_i}$ and then that $k_n^{\delta_i} \in \{k_{\max} - \ell_n(\xi_n), \dots, k_{\max}\}$. As before, we conclude that $(x_{n+1}^{\delta_i})_{i \in \mathbb{N}}$ splits into $\ell_n(\xi_n) + 1$ convergent subsequences, each one converging to a point of \tilde{X}_{n+1} .

Case 3. $k_n(\xi_n) = 0$. This case comes with $\tilde{b}_n = 0$, that is, $F'_{[n]}(\xi_n) = y_{[n]}$. Then, $\xi_n \in \tilde{X}_{n+1}$. By an inductive argument using (49) we obtain from the definition of $s_{n, k}^{\delta_i}$ (see the **repeat** loop of Algorithm 1) that $\lim_{i \rightarrow \infty} s_{n, k}^{\delta_i} = 0$ for each k . As $k_n^{\delta_i} \leq k_{\max}$, the sequence $(x_n^{\delta_i} + s_{n, k_n^{\delta_i}}^{\delta_i})$ splits into $k_{\max} + 1$ subsequences at most as $i \rightarrow \infty$. All these subsequences converge to $\xi_n \in \tilde{X}_{n+1}$. \square

The next lemma and its corollary establish uniform convergence of $(\tilde{X}_n)_n$ to x^+ .

LEMMA 11. *Under the assumptions of Theorem 5 there is, for any $\epsilon > 0$, a constant $M = M(\epsilon) \in \mathbb{N}$ such that*

$$\Delta_p(\xi_n, x^+) < \epsilon \quad \text{for all } n \geq M(\epsilon) \text{ and all } \xi_n \in \tilde{X}_n.$$

Proof. See Appendix A for the proof. \square

COROLLARY 12. *Let all assumptions of Theorem 5 hold true. Then, Lemma 11 is true with the norm in X : for any $\epsilon > 0$ there is a constant $M = M(\epsilon) \in \mathbb{N}$ such that*

$$\|x^+ - \xi_n\| < \epsilon \quad \text{for all } n \geq M(\epsilon) \text{ and all } \xi_n \in \tilde{X}_n.$$

Proof. From Lemma 11 we conclude that the sequence $(\xi_n)_{n \in \mathbb{N}}$ is uniformly bounded. Then the result is a consequence of (10). \square

We are able to prove strong convergence.

THEOREM 13. *Let all the assumptions of Theorem 7 hold true with $1 < p \leq s \leq r$ and*

$$\omega_{n, k}^\delta := \min \left\{ \omega_{n, k}^{\delta, (1)}, \omega_{n, k}^{\delta, (2)}, \bar{\omega} \right\},$$

where $\bar{\omega} > 0$ is a constant. Additionally, let all Y_j 's be uniformly smooth. Then,

$$\lim_{\delta \rightarrow 0} \|x_{N(\delta)}^\delta - x^+\| = 0,$$

where x^+ is the solution of (1) which we assumed to be unique in $B_\rho(x^+, \Delta_p)$.

Proof. Let $\delta_i \rightarrow 0$ as $i \rightarrow \infty$. The cases $N(\delta_i) \rightarrow n \in \mathbb{N}$ as $i \rightarrow \infty$ and $N(\delta_i)$ bounded can be dealt with as in the proof of Theorem 3.7 in [20]. Now, let $N(\delta_i) \rightarrow \infty$ as $i \rightarrow \infty$ and let $\epsilon > 0$ be given. From (44) and (10) we conclude that there exists some positive constant $C_1 > 0$ such that

$$\|x_n^{\delta_i} - x^+\|^s \leq C_1 \Delta_p(x_n^{\delta_i}, x^+) \quad \text{for all } n \leq N(\delta_i).$$

But for all $n \in \mathbb{N}$ fixed, we have $N(\delta_i) \geq n$ provided i is large enough. It follows from (43) that

$$\|x_{N(\delta_i)}^{\delta_i} - x^+\|^s \leq C_1 \Delta_p(x_{N(\delta_i)}^{\delta_i}, x^+) \leq C_1 \Delta_p(x_n^{\delta_i}, x^+)$$

for each n fixed and i large enough. Therefore, it is sufficient to prove that

$$(56) \quad \Delta_p(x_n^{\delta_i}, x^+) \leq \frac{\epsilon^s}{C_1}$$

for some $n \in \mathbb{N}$ fixed and i sufficiently large. From the definition of Δ_p ,

$$\Delta_p(x_n^{\delta_i}, x^+) \leq \left| \frac{1}{p} \|x_n^{\delta_i}\|^p - \frac{1}{p} \|x^+\|^p \right| + |\langle J_p(x_n^{\delta_i}), x^+ - x_n^{\delta_i} \rangle|.$$

As the norm is a continuous function, there is a $\gamma_1 = \gamma_1(\epsilon) > 0$ such that

$$\left| \frac{1}{p} \|x_n^{\delta_i}\|^p - \frac{1}{p} \|x^+\|^p \right| \leq \frac{\epsilon^s}{2C_1}$$

whenever $\|x^+ - x_n^{\delta_i}\| \leq \gamma_1$. Moreover, from (44), if $\|x^+ - x_n^{\delta_i}\| \leq \gamma_2 = \gamma_2(\epsilon) := \frac{\epsilon^s}{2C_1^{p-1}}$, we have from the properties of J_p ,

$$|\langle J_p(x_n^{\delta_i}), x^+ - x_n^{\delta_i} \rangle| \leq \|x_n^{\delta_i}\|^{p-1} \|x^+ - x_n^{\delta_i}\| \leq \frac{\epsilon^s}{2C_1},$$

which implies that (56) is satisfied whenever

$$(57) \quad \|x^+ - x_n^{\delta_i}\| \leq \gamma = \gamma(\epsilon) := \min\{\gamma_1, \gamma_2\}$$

for some $n \in \mathbb{N}$ fixed and i large enough. To validate (57), we again follow ideas of [20]. According to Corollary 12 we find an $n \in \mathbb{N}$ such that

$$\|x^+ - \xi_n\| \leq \frac{\gamma}{2} \text{ for all } \xi_n \in \tilde{X}_n.$$

According to Lemma 10 there is an $I(\gamma) \in \mathbb{N}$ such that for all integers $i \geq I(\gamma)$ there exists a $\xi_n^{(i)} \in \tilde{X}_n$ with

$$\|x_n^{\delta_i} - \xi_n^{(i)}\| \leq \frac{\gamma}{2}.$$

Then, for all $i \geq I(\gamma)$ we can find an appropriate $\xi_n^{(i)} \in \tilde{X}_n$ with

$$\|x^+ - x_n^{\delta_i}\| \leq \|x^+ - \xi_n^{(i)}\| + \|\xi_n^{(i)} - x_n^{\delta_i}\| \leq \gamma.$$

This is (57) and completes the proof. \square

Remark 14. If the number d of equations is equal to 1 and if we set $k_{\max} = \infty$ then Algorithm 1 coincides with the REGINN-Landweber method presented and analyzed by Jin in [14] where, however, only weak convergence was established; cf. [14, Remark 3.4]. In view of the proof of Lemma 10 we are able to prove strong convergence with little additional effort. In fact, from [14] we know that k_n^δ (or k_n in the noiseless situation) and $N(\delta)$ are well defined (in other words: the **repeat** and **while** loops of Algorithm 1 terminate). Further, strong convergence in the noiseless situation is also provided. In this situation, only Cases 1 and 3 in the proof of Lemma 10 need to be considered. The proof of Case 1 remains the same; however, we need to modify the arguments for $k_n(\xi_n) = 0$ (Case 3) because we are not allowed to assume $k_n^\delta \leq k_{\max} < \infty$ uniformly in δ . To this end let $(x_n^{\delta_i})$ converge to ξ_n with $k_n(\xi_n) = 0$ which yields $F(\xi_n) = y$ (F consists of only 1 component) and $\xi_n = x^+$ by the uniqueness assumption on x^+ in $B_\rho(x^+, \Delta_p)$. Hence, $x^+ \in \tilde{X}_{n+1}$. Assume that $(x_{n+1}^{\delta_i})$ does not converge to x^+ as $i \rightarrow \infty$. Then, there are an $\epsilon > 0$ and a subsequence $(\delta_{i_m})_m$ such that

$$\epsilon < \|x_{n+1}^{\delta_{i_m}} - x^+\|^s \stackrel{(10)}{\lesssim} \Delta_p(x_{n+1}^{\delta_{i_m}}, x^+) \stackrel{(43)}{\lesssim} \Delta_p(x_{n+1}^{\delta_{i_m}}, x^+) \xrightarrow{m \rightarrow \infty} \Delta_p(x^+, x^+) = 0,$$

which contradicts the assumption. As a consequence, strong convergence of Jin’s method is a byproduct of our analysis.

Relying on a variational source condition Kaltenbacher and Tomba [17] could recently validate strong convergence with rates of an iteratively regularized Gauss–Newton version of Jin’s method.

Remark 15. Observe that the sequences (b_n^δ) and $(z_{n,k}^\delta)$ are uniformly bounded and because $\|A_n^\delta\| \leq M$, we have that $\|b_n^\delta - A_n^\delta s_{n,k}^\delta\|$ is also uniformly bounded. As $1 < p \leq s \leq r$, we conclude that (see (42)) $\min\{\omega_{n,k}^{\delta,(1)}, \omega_{n,k}^{\delta,(2)}\} \geq C$, where the constant $C > 0$ is independent of n, k , and δ . If $\bar{\omega} \leq C$, then

$$\omega_{n,k}^\delta = \min\{\omega_{n,k}^{\delta,(1)}, \omega_{n,k}^{\delta,(2)}, \bar{\omega}\} = \bar{\omega}$$

which allows us to use a small constant step size for our numerical experiments in the following section.

5. Numerical experiments. To demonstrate its strengths and weaknesses we apply K-REGINN-Landweber to the inverse problem of electric impedance tomography (EIT) introduced by Calderón [4]. We give a rough explanation of EIT; for more details concerning modeling and mathematical results see, e.g., both overview articles [2, 5].

Let $\Omega \subset \mathbb{R}^2$ be a bounded and simply connected Lipschitz domain. We apply some electric currents $g: \partial\Omega \rightarrow \mathbb{R}$ on its boundary and record the resulting potentials $f: \partial\Omega \rightarrow \mathbb{R}$ on its boundary as well. The goal is to reconstruct the electric conductivity $\gamma: \Omega \rightarrow \mathbb{R}$ in the whole of Ω .

The governing equation of the continuum model in weak formulation is the elliptic variational problem: find $u \in H_{\diamond}^1(\Omega) := \{v \in H^1(\Omega) : \int_{\partial\Omega} v = 0\}$ such that

$$(58) \quad \int_{\Omega} \gamma \nabla u \nabla \varphi = \int_{\partial\Omega} g \varphi \quad \text{for all } \varphi \in H_{\diamond}^1(\Omega).$$

If $\gamma \in L_+^\infty(\Omega) := \{v \in L^\infty(\Omega) : v \geq C \text{ a.e.}\}$ for a positive constant C and $g \in H_{\diamond}^{-1/2}(\partial\Omega) = H_{\diamond}^{1/2}(\partial\Omega)^*$, where $H_{\diamond}^{1/2}(\partial\Omega) := \{v \in H^{1/2}(\partial\Omega) : \int_{\partial\Omega} v = 0\}$, then due

to standard elliptic theory there exists a unique solution of (58). Furthermore, its trace $f = u|_{\partial\Omega}$ is in $H_{\diamond}^{1/2}(\partial\Omega)$ and the mapping $\Lambda_{\gamma}: H_{\diamond}^{-1/2}(\partial\Omega) \ni g \mapsto f \in H_{\diamond}^{1/2}(\partial\Omega)$ is a linear homeomorphism called *Neumann-to-Dirichlet* map (in short NtD).

The *forward operator* $F: D(F) \subset L^{\infty}(\Omega) \rightarrow \mathcal{L}(H_{\diamond}^{-1/2}(\partial\Omega), H_{\diamond}^{1/2}(\partial\Omega))$ associated with EIT is defined by

$$(59) \quad F(\gamma) := \Lambda_{\gamma},$$

where $D(F) := L^{\infty}_+(\Omega)$. Solving (59) for γ given Λ_{γ} is the EIT inverse problem which is uniquely solvable [1].

In practice the full NtD map is not completely available, only $d \in \mathbb{N}$ potentials $\Lambda_{\gamma}g_j$, $j = 0, \dots, d - 1$, can be observed which are induced by the currents $g_j \in H_{\diamond}^{-1/2}(\partial\Omega)$. This fact leads us to introduce the operators $F_j: D(F) \subset L^{\infty}(\Omega) \rightarrow H_{\diamond}^{1/2}(\partial\Omega)$, $F_j(\gamma) := \Lambda_{\gamma}g_j$, $j = 0, \dots, d - 1$. We are now in the situation (5). Moreover, F_j is Fréchet differentiable; see, e.g., [19]. Unfortunately, $L^{\infty}(\Omega)$ is not a Banach space covered by our analysis of K-REGINN-Landweber (it is not s -convex). Since $L^{\infty}(\Omega) \subset L^p(\Omega)$ (Ω is bounded) we could redefine the mappings F_j as $F_j: D(F) \subset L^p(\Omega) \rightarrow H_{\diamond}^{1/2}(\partial\Omega)$. Indeed, $L^p(\Omega)$, $1 < p < \infty$, is q -smooth and s -convex for $q := \min\{p, 2\}$ and $s := \max\{p, 2\}$; see, e.g., [6]. The duality mapping $J_p: L^p(\Omega) \rightarrow L^{p^*}(\Omega)$ is given by

$$(60) \quad J_p(f) = |f|^{p-1} \text{sign}(f).$$

We are, however, in trouble again because F_j is certainly not Fréchet differentiable with respect to the L^p -topology ($D(F)$ contains no interior points). We suggest a pragmatic approach as remedy: restrict the searched for conductivity to a finite dimensional space $V \subset L^{\infty}(\Omega)$, that is, consider

$$(61) \quad F_j: V_+ \subset V_p \rightarrow L^2(\partial\Omega_j), \quad \gamma \mapsto \Lambda_{\gamma}g_j, \quad V_+ = D(F) \cap V,$$

where $\partial\Omega_j$ is this part of the boundary where the measurements are actually taken. We write $V_p = (V, \|\cdot\|_{L^p(\Omega)})$ to emphasize that $V = \text{span}\{v_1, \dots, v_M\}$ is equipped with the L^p -norm.³ This model is reasonable for two reasons: 1. only finitely many degrees of freedom of the conductivity can be determined from finitely many measurements; 2. from a computational point of view we are bound to a finite dimensional setting anyway.

The operators (61) have Fréchet derivatives $F'_j: \text{int}(V_+) \rightarrow \mathcal{L}(V_p, L^2(\partial\Omega_j))$, $F'_j(\gamma)h = w_j|_{\partial\Omega_j}$, where $w_j \in H_{\diamond}^1(\Omega)$ is the unique solution of

$$(62) \quad \int_{\Omega} \gamma \nabla w_j \nabla \varphi = - \int_{\Omega} h \nabla u_j \nabla \varphi \quad \text{for all } \varphi \in H_{\diamond}^1(\Omega)$$

with u_j solving (58) for $g = g_j$. The associated adjoint is $F'_j(\gamma)^*: L^2(\partial\Omega_j) \rightarrow V_{p^*}$, $F'_j(\gamma)^*z = \sum_i \lambda_i(z)v_i$, where $\lambda = (\lambda_1(z), \dots, \lambda_M(z))^{\top} \in \mathbb{R}^M$ uniquely solves

$$A\lambda = b, \quad \text{where } A \in \mathbb{R}^{M \times M} \text{ and } b \in \mathbb{R}^M \text{ with } A_{i,j} = \int_{\Omega} v_i v_j \text{ and } b_i = \int_{\partial\Omega_j} z F'_j(\gamma)v_i.$$

³The v_i 's are naturally assumed to be linearly independent and chosen such that V_+ has a nonempty interior: $\text{int}(V_+) \neq \emptyset$.

Let ψ_z and $w_{i,j} \in H^1_{\diamond}(\Omega)$ be the unique solutions of (58) for $g = z$ and (62) for $h = v_i$, respectively. Then,

$$b_i = \int_{\partial\Omega_j} z F'_j(\gamma)v_i = \int_{\partial\Omega_j} z w_{i,j} = \int_{\Omega} \gamma \nabla \psi_z \nabla w_{i,j} = - \int_{\Omega} v_i \nabla u_j \nabla \psi_z.$$

Now we introduce our test environment: Ω is the unit square $(0, 1)^2$ and we feed in the $d = 4m$ ($m \in \mathbb{N}$) independent currents

$$g_j(x, y) := \begin{cases} \cos(2j_0\pi x) \cos(2j_0\pi y) & : (x, y) \in \Gamma_{j_1}, \\ 0 & : \text{otherwise,} \end{cases}$$

where $j = 4(j_0 - 1) + (j_1 - 1)$, $j_0 = 1, \dots, m$, $j_1 = 1, \dots, 4$. Further, $\Gamma_1 := [0, 1] \times \{1\}$, $\Gamma_2 := \{1\} \times [0, 1]$, $\Gamma_3 := [0, 1] \times \{0\}$, and $\Gamma_4 := \{0\} \times [0, 1]$ are the faces of Ω . We operate with $\partial\Omega_j = \partial\Omega \setminus \Gamma_{j_1}$ (we do not take measurements where we apply currents).

The basis functions v_i of the conductivity space V are constructed by a Delaunay triangulation $\mathcal{T} = \{T_i : i = 1, \dots, M\}$ provided by the MATLAB pde toolbox,⁴ where $M = 3066$. Set $v_i = \chi_{T_i}$, $i = 1, \dots, M$ (χ_B denotes the indicator function of the set B). For this basis, A reduces to diagonal matrix $A = \text{diag}(|T_i|)$, where $|T_i|$ represents the area of triangle T_i and then $\lambda_i(z) = b_i/|T_i|$.

Remark 16. The above choice of V guarantees injectivity of $F'_j(\gamma)$, $\gamma \in \text{int}(V_+)$. Moreover, F_j satisfies the tangential cone condition (Assumption 1); see [19, section 3].

The exact solution $\gamma^+ = \sum_{i=1}^M \alpha_i v_i$ models a constant background conductivity 0.1 and two ball-like inclusions with conductivity 1. Denoting the centroid of T_i by ξ_i we set

$$(63) \quad \gamma^+ = \sum_{i=1}^M \alpha_i v_i \quad \text{with } \alpha_i = \begin{cases} 1 & : \xi_i \in B_1 \cup B_2, \\ 0.1 & : \text{otherwise,} \end{cases}$$

where B_1 and B_2 are (open) balls with radii 0.15 about the centers $(0.35, 0.35)$ and $(0.65, 0.65)$, respectively. Figure 1 displays γ^+ . The corresponding data $\Lambda_{\gamma^+} g_j$ have been computed by the finite element method (FEM). We assume to know the background conductivity and, thus, start with initial iterate $\gamma_0 \equiv 0.1$. The other input parameters of Algorithm 1 are $\tau = 1.5$, $k_{\max} = 10^6$, $\mu = 0.8$, and $\omega_{n,k} = 0.001$; see Remark 15. Further, $\beta_{h,j} = \pi_{h,j} = 0$. During the iteration the elliptic problems (58) and (62) have been solved by FEM also, however, using a different and coarser discretization mesh as for generating the data to avoid the most obvious inverse crime.

The *relative L^2 -error* of the n th iterate γ_n is denoted by

$$e_n := 100 \frac{\|\gamma_n - \gamma^+\|_{L^2(\Omega)}}{\|\gamma^+\|_{L^2(\Omega)}}.$$

Our initial error is $e_0 \approx 87.4\%$. As $Y_j = L^2(\partial\Omega_j)$ is a Hilbert space, the normalized duality mapping is the identity operator. For this reason, we have chosen $r = 2$ and then $j_r(f) = f$ for all $f \in Y_j$. Due to the restriction $1 < p \leq s \leq r$ (see, e.g., Theorem 13), we have that $1 < p \leq 2$.

⁴MATLAB is a trademark of The MathWorks, Inc.

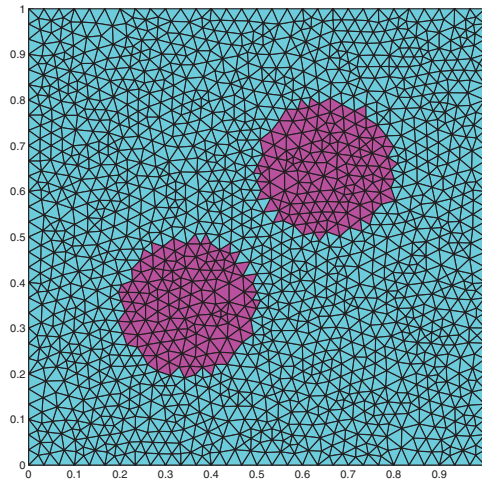


FIG. 1. The searched for conductivity γ^+ (63) superimposed by the corresponding triangulation \mathcal{T} of Ω .

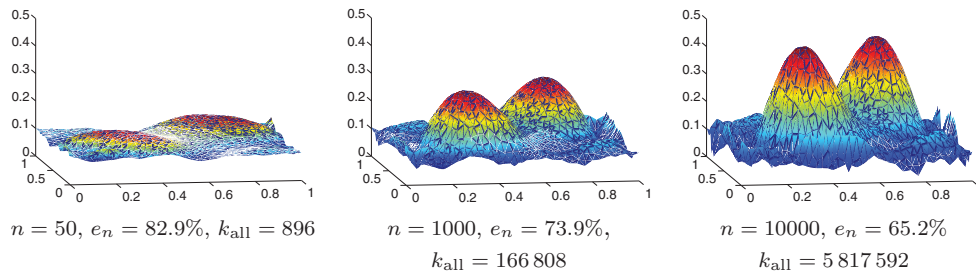


FIG. 2. Some iterates of *K-REGINN-Landweber* for $d = 4$. Below each plot the iteration index, the relative error, and the overall number of inner iterations are given (The plots actually do not show $\gamma_n = \sum \alpha_i^n v_i$ but smoothed versions which are obtained by piecewise linearly interpolating the coefficients (α_i^n)).

In our first experiments we work with the Hilbert space $X = V_2$, i.e., $p = 2$ to illustrate the convergence results of Theorems 5 and 13. Figure 2 shows different iterations where $d = 4$ and no noise (other than noise from discretization) is present ($\delta = 0$). The slow convergence of the inner (Landweber) iteration can be observed when looking at the number k_{all} of overall inner iterations (sweeps through the **repeat** loop of Algorithm 1). The locations of the inclusions appear clearly after only a few (outer) iterations. This observation is quite remarkable as only one current per face was applied. For a possible analytic explanation we refer to [11].

The convergence $\gamma_{N(\delta)}^\delta \rightarrow \gamma^+$ as $\delta \rightarrow 0$ (Theorem 13) is somewhat illustrated in Table 1. To this end we corrupted the simulated exact data by artificial random noise of relative noise level δ , that is, the perturbed data are

$$(64) \quad \Lambda_{\gamma+g_j}^\delta = \Lambda_{\gamma+g_j} + \delta \|\Lambda_{\gamma+g_j}\|_{L^2(\partial\Omega_j)} \text{per}_j, \quad j = 0, \dots, d - 1,$$

where per_j is a uniformly distributed random variable such that $\|\text{per}_j\|_{L^2(\partial\Omega_j)} = 1$. In contrast to the previous sections δ denotes here a relative noise level.

Remark 17. We emphasize that $k_{\text{max}} = 10^6$ is never reached in all our experiments: The **repeat** loop terminates as $\|b_n^\delta - A_n^\delta s_{n,k}^\delta\| < \mu \|b_n^\delta\|$ happens for a $k < k_{\text{max}}$.

TABLE 1
Relative L^2 error of $\gamma_{N(\delta)}^\delta$ for decreasing noise level δ . Here, $d = 4$.

δ (%)	8.0	4.0	2.0	1.0	0.5
$N(\delta)$	8	22	58	104	248
k_{all}	20	267	1 491	3 871	18 381
$e_{N(\delta)}$ (%)	87.1	85.6	82.9	81.7	80.4

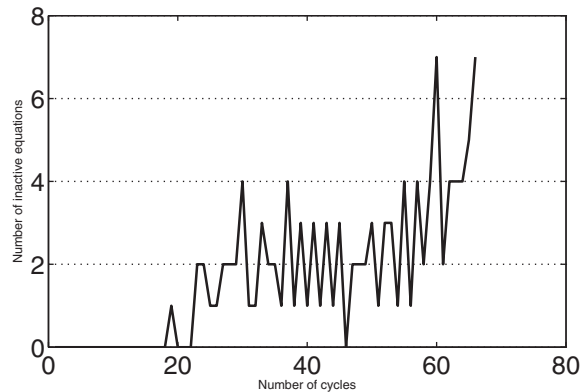


FIG. 3. Number of inactive equations per cycle for $d = 8$.

If, however, we reduce k_{max} to a much smaller value, say $k_{\text{max}} = 50$, then the **repeat** loop is stopped predominantly because $k = k_{\text{max}}$ is reached. In the experimental setting underlying Table 1 K-REGINN-Landweber terminates then with even smaller k_{all} 's. The price to pay is worse accuracy (larger reconstruction error).

With our next experiments we highlight the mode of operation of Algorithm 1. One advantage of Kaczmarz over classical methods is that not all equations are active in one cycle.⁵ An equation is inactive in a cycle if the **then** branch of the **if** block is executed. Figure 3 displays the number of inactive equations as a function of the cycle number (which is the variable ℓ in Algorithm 1). Here, $d = 8$, $\delta = 1\%$, and all other input parameters including the spaces remain the same as before. Notice that whereas at the beginning the algorithm works with all equations in each cycle, a considerable number of them are not used toward the end.

As a complement to Figure 3 we present Table 2 where more internal information is available: the number of inner iterations per equation and per cycle is listed. Also the number of inactive equations is noted. Here, $d = 4$, $\tau = 2$, and $\delta = 2\%$.

Regularization in appropriate Banach spaces is known to foster sparsity and steep gradients. This feature is demonstrated in Figure 4 which collects reconstructions with respect to different spaces $X = V_2$ ($p = 2$) and $X = V_{1.1}$ ($p = 1.1$), different noise levels δ , and different numbers of equations d . All other parameters remain as before and τ is reset to 1.5. The Banach space norm $p = 1.1$ separates the two inclusion sharper than the Hilbert space norm independently of the noise level. If the available information increases this separation becomes even better (last row).

Finally we illustrate L^1 fitting which is well suited to cope with impulsive noise; see, e.g., [7]. To this end we set $Y_j = L^t(\partial\Omega_j)$ for some $t > 1$ but close to 1 and

⁵A cycle is one sweep through all equations in (5). More precisely, one run of the **for** loop of Algorithm 1.

TABLE 2
 Number of inner iterations per equation and per cycle.

no. of eqn.	no. of cycle									
	1	2	3	4	5	6	7	8	9	10
1	2	8	14	19	20	22	24	25	49	0
2	2	7	18	25	29	32	35	0	31	0
3	2	7	15	22	28	36	43	0	0	–
4	4	11	19	24	29	34	41	0	0	–
# iter./cycle	10	33	66	90	106	124	143	25	80	0
# inactive eqns.	0	0	0	0	0	0	0	3	2	2

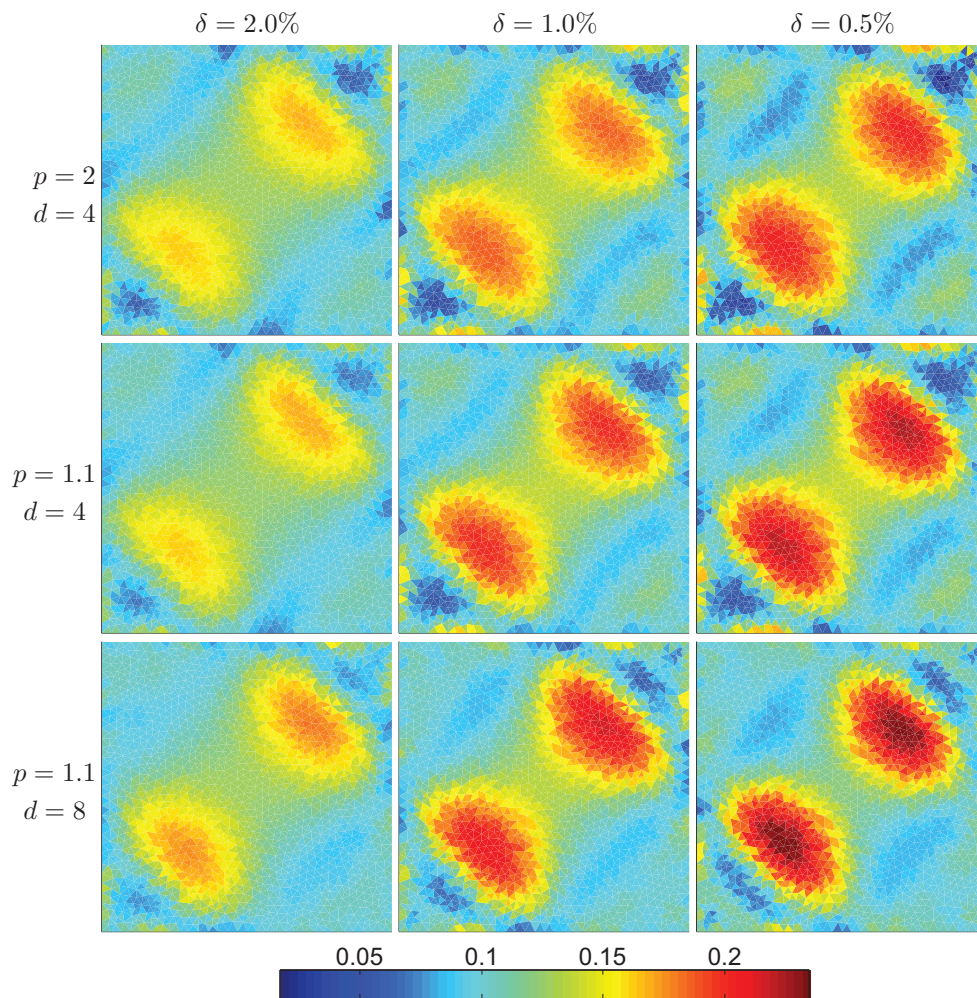


FIG. 4. Reconstructed conductivities with respect to different spaces, number of equations, and noise levels. The color scale is the same for all images and ranges from 0.018 (darkest blue) to 0.23 (darkest red).

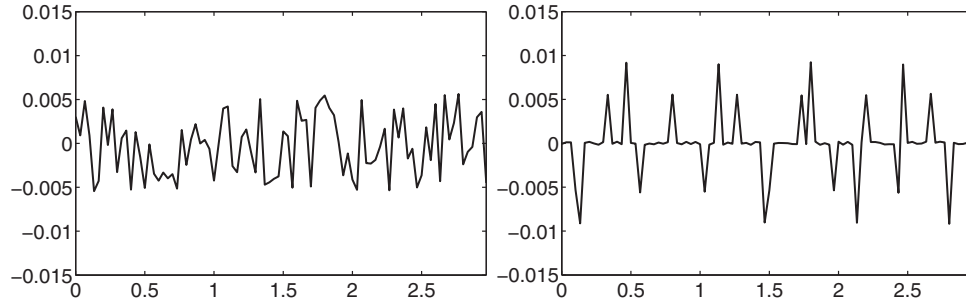


FIG. 5. *Left: uniform noise. Right: impulsive noise. Both noise signals are scaled to have the same L^2 -norm.*

consider

$$F_j: V_+ \subset V_p \rightarrow L^t(\partial\Omega_j), \quad 1 < t \leq 2.$$

As $\|g\|_{L^t(\partial\Omega_j)} \lesssim \|g\|_{L^2(\partial\Omega_j)}$, the new F_j is Fréchet differentiable with the same derivative. If we choose $p = 2$, i.e., $X = V_2 (\subset L^2(\Omega))$, then $s = 2$ and due to the restriction $s \leq r$ (see Theorem 13), we cannot use directly the duality mapping J_r for $r = t$ in the space $L^t(\partial\Omega_j)$ for $t < 2$. However, in view of (6) and (60), we obtain the normalized duality mapping J_2 in $L^t(\partial\Omega_j)$ as

$$J_2(f) = \|f\|_{L^t(\partial\Omega_j)}^{2-t} J_t(f) = \|f\|_{L^t(\partial\Omega_j)}^{2-t} |f|^{t-1} \text{sign}(f)$$

which allows us to use $r = 2$.

To test the performance of K-REGINN-Landweber under impulsive noise we superimposed standard uniform noise with some highly inconsistent data points. These *outliers* may arise from procedural measurement errors in practical applications. Figure 5 shows a plot of such a kind of noise. In our experiments we scaled both kinds of noise such that the relative L^2 noise is 1% in each case; cf. (64). See Figure 6 for a visual impression of the reconstructed conductivities. Below each image we note the relative error $e_{N(\delta)}$, the number of outer iterations $N(\delta)$, and the overall number of inner iterations k_{all} . Under uniform noise the L^2 setting yields visually the best reconstruction. As expected, if the noise is impulsive, the $L^{1.01}$ approach is best from a qualitative as well as a quantitative point of view.

Remark 18. For the last experiment we have equipped K-REGINN-Landweber with a strategy to choose the tolerance μ (13) adaptively. The scheme suggested by [24, sect. 6] works as expected provided each equation is treated separately; the overall number of inner iterations is greatly reduced.

Appendix A. Proof of Lemma 11. Here we provide a proof of Lemma 11 by a more general result. The technique of this proof is adapted from Hanke [10, Proposition 4.3].

PROPOSITION 19. *Let all assumptions of Theorem 5 hold true. Let $(x_n^{(l)})_n$ be a sequence generated by a run of K-REGINN-Landweber using the arbitrary sequence $(k_n^{(l)})_n$ of stopping indices in the inner iteration, that is, $x_{n+1}^{(l)} = x_n^{(l)} + s_{n, k_n^{(l)}}$, $x_0^{(l)} = x_0$. Each stopping index must satisfy $k_n^{(l)} \in \{1, \dots, k_n\}$ and $k_n^{(l)} = 0$ only if $k_n = 0$, where k_n is the stopping index from (14). Then, for each $\epsilon > 0$, there exists a number $M(\epsilon) \in \mathbb{N}$ independent of l such that*

$$\Delta_p(x_n^{(l)}, x^+) \leq \epsilon \text{ for all } n \geq M(\epsilon) \text{ and all sequences } (k_n^{(l)})_n.$$

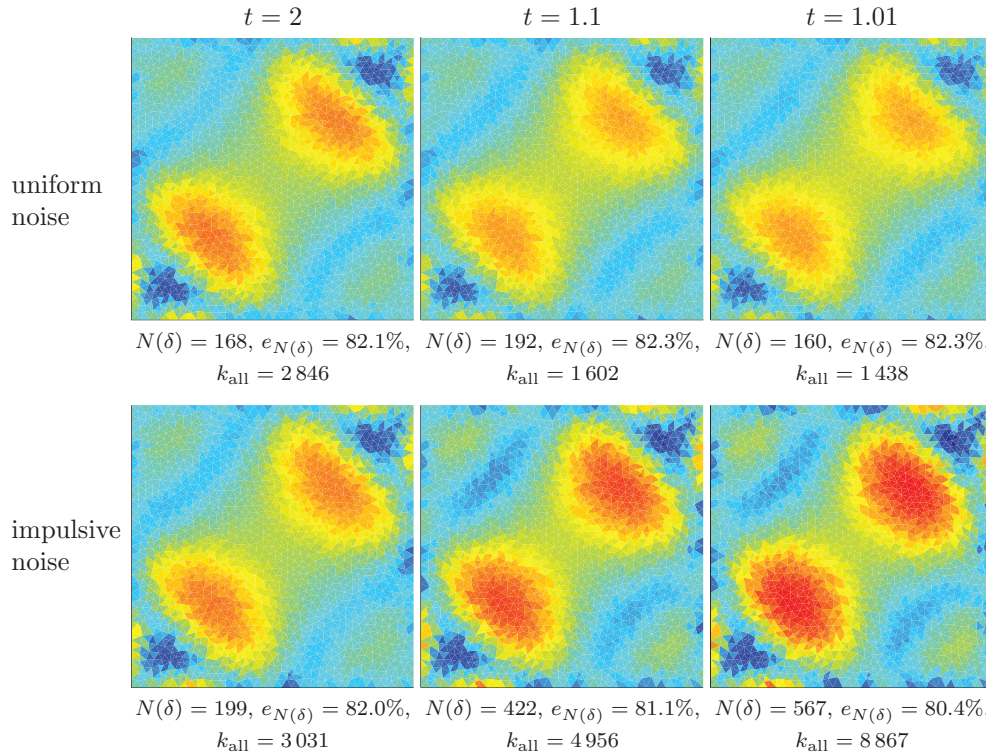


FIG. 6. Reconstructed conductivities for $d = 4$ with respect to different noises and different data fitting norms. The relative L^2 noise level is 1% for each type of noise. The used color scale is as in Figure 4.

Proof. As $\|A_n s_{n,k} - b_n\| \geq \mu \|b_n\|$ holds for any $k = 0, \dots, k_n^{(l)} - 1$, Theorem 5 holds and all sequences $(x_n^{(l)})_n$ converge to x^+ . Assume that the proposition is not true. Then, there exists some $\epsilon > 0$ and strictly increasing sequences $(n_j)_j, (l_j)_j \subset \mathbb{N}$ such that

$$\Delta_p(x_{n_j}^{(l_j)}, x^+) > \epsilon \text{ for all } j \in \mathbb{N},$$

where $(x_n^{(l_j)})_n$ represents the sequence generated by the sequence of stopping indices $(k_n^{(l_j)})_n$. The iterates $x_{n_j}^{(l_j)}$ must be generated by infinitely many different sequences of stopping indices (otherwise, as $x_{n_j}^{(l)} \rightarrow x^+$ as $j \rightarrow \infty$ for each l fixed and as the l_j 's have only a finite number of values, we would have $\Delta_p(x_{n_j}^{(l_j)}, x^+) < \epsilon$ for n_j large enough). Next we reorder the numbers l_j (excluding some iterates if necessary) such that

$$(65) \quad \Delta_p(x_{n_l}^{(l)}, x^+) > \epsilon \text{ for all } l \in \mathbb{N}.$$

Now, as $k_0 < \infty$, there exists $\widehat{k}_0 \in \{0, \dots, k_0\}$ such that $\widehat{k}_0 = k_0^{(l)}$ for infinitely many $l \in \mathbb{N}$. Let $\mathcal{L}_0 \subset \mathbb{N}$ be the set of those indices l . Next, fix \widehat{k}_0 as the stopping index for the first inner iteration and consider the second outer iteration; again, $k_1 < \infty$, and there is a $\widehat{k}_1 \in \{0, \dots, k_1\}$ such that $\widehat{k}_1 = k_1^{(l)}$ for infinitely many $l \in \mathcal{L}_0 \setminus \{1\}$. Those l 's

are collected in \mathcal{L}_1 and so on. It follows that there exists a sequence of stopping indices $(\widehat{k}_n)_n$ and unbounded sets $\mathcal{L}_n \subset \mathbb{N} \setminus \{1, \dots, n\}$, $n \in \mathbb{N}_0$, with $\mathcal{L}_0 \supset \mathcal{L}_1 \supset \mathcal{L}_2 \supset \dots$, such that

$$(66) \quad \widehat{k}_n = k_n^{(l)} \text{ for all } l \in \mathcal{L}_n, n \in \mathbb{N}_0.$$

Denote by $(\widehat{x}_n)_n$ the sequence corresponding to the sequence $(\widehat{k}_n)_n$ of stopping indices.

Observe that, if $l \in \mathcal{L}_0$, then from (66) $x_1^{(l)} = x_0^{(l)} + s_{0, k_0^{(l)}} = \widehat{x}_0 + s_{0, \widehat{k}_0} = \widehat{x}_1$. Similarly, if $l \in \mathcal{L}_1$ ($\subset \mathcal{L}_0$), then $x_2^{(l)} = x_1^{(l)} + s_{1, k_1^{(l)}} = \widehat{x}_1 + s_{1, \widehat{k}_1} = \widehat{x}_2$. By induction,

$$(67) \quad l \in \mathcal{L}_n \implies x_{n+1}^{(l)} = \widehat{x}_{n+1} \text{ for all } n \in \mathbb{N}_0.$$

Furthermore, $\widehat{x}_n \rightarrow x^+$ as $n \rightarrow \infty$ which implies that there exists $M = M(\epsilon) \in \mathbb{N}$ such that

$$(68) \quad \Delta_p(\widehat{x}_n, x^+) \leq \epsilon \text{ for all } n \geq M.$$

Finally, for $l \in \mathcal{L}_M$ fixed, the errors $\Delta_p(x_n^{(l)}, x^+)$ are monotonically decreasing with n ; see (18). In particular, $n_l \geq l \geq M + 1$ (because $l \in \mathcal{L}_M \subset \mathbb{N} \setminus \{1, \dots, M\}$) which implies in view of (67) and (68)

$$\Delta_p(x_{n_l}^{(l)}, x^+) \leq \Delta_p(x_{M+1}^{(l)}, x^+) = \Delta_p(\widehat{x}_{M+1}, x^+) \leq \epsilon$$

contradicting (65). \square

As each element of \widetilde{X}_n has the form $x_n^{(l)}$ for some l , Lemma 11 is an immediate consequence of the above proposition.

REFERENCES

- [1] K. ASTALA AND L. PÄIVÄRINTA, *Calderón's inverse conductivity problem in the plane*, Ann. of Math. (2), 163 (2006), pp. 265–299.
- [2] L. BORCEA, *Electrical impedance tomography*, Inverse Problems, 18 (2002), pp. R99–R136.
- [3] M. BURGER AND B. KALTENBACHER, *Regularizing Newton–Kaczmarz methods for nonlinear ill-posed problems*, SIAM J. Numer. Anal., 44 (2006), pp. 153–182.
- [4] A.P. CALDERÓN, *On an inverse boundary value problem*, Comput. Appl. Math., 25 (2006), pp. 133–138.
- [5] M. CHENEY, D. ISAACSON, AND J.C. NEWELL, *Electrical impedance tomography*, SIAM Rev., 41 (1999), pp. 85–101.
- [6] I. CIORANESCU, *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*, Math. Appl. 62, Kluwer Academic, Dordrecht, 1990.
- [7] C. CLASON AND B. JIN, *A semismooth Newton method for nonlinear parameter identification problems with impulsive noise*, SIAM J. Imaging Sci., 5 (2012), pp. 505–536.
- [8] M. HALTMEIER, A. LEITÃO, AND O. SCHERZER, *Kaczmarz methods for regularizing nonlinear ill-posed equations. I. Convergence analysis*, Inverse Problems Imaging, 1 (2007), pp. 289–298.
- [9] M. HANKE, *A regularizing Levenberg–Marquardt scheme, with applications to inverse ground-water filtration problems*, Inverse Problems, 13 (1997), pp. 79–95.
- [10] M. HANKE, *Regularizing properties of a truncated Newton–CG algorithm for nonlinear inverse problems*, Numer. Funct. Anal. Optim., 18 (1997), pp. 971–993.
- [11] B. HARRACH AND J.K. SEO, *Exact shape-reconstruction by one-step linearization in electrical impedance tomography*, SIAM J. Math. Anal., 42 (2010), pp. 1505–1518.
- [12] T. HEIN AND K.S. KAZIMIERSKI, *Accelerated Landweber iteration in Banach spaces*, Inverse Problems, 26 (2010), 055002.
- [13] B. JIN AND P. MAASS, *Sparsity regularization for parameter identification problems*, Inverse Problems, 28 (2012), 123001.

- [14] Q. JIN, *Inexact Newton-Landweber iteration for solving nonlinear inverse problems in Banach spaces*, Inverse Problems, 28 (2012), 065002.
- [15] Q. JIN AND W. WANG, *Landweber iteration of Kaczmarz type with general non-smooth convex penalty functionals*, Inverse Problems, 29 (2013), 085011.
- [16] B. KALTENBACHER, F. SCHÖPFER, AND T. SCHUSTER, *Iterative methods for nonlinear ill-posed problems in Banach spaces: Convergence and applications to parameter identification problems*, Inverse Problems, 25 (2009), 065003.
- [17] B. KALTENBACHER AND I. TOMBA, *Convergence rates for an iteratively regularized Newton-Landweber iteration in Banach space*, Inverse Problems, 29 (2013), 25010.
- [18] R. KOWAR AND O. SCHERZER, *Convergence analysis of a Landweber-Kaczmarz method for solving nonlinear ill-posed problems*, in Ill-Posed and Inverse Problems VSP, Utrecht, The Netherlands, 2002, pp. 69–90.
- [19] A. LECHLEITER AND A. RIEDER, *Newton regularizations for impedance tomography: Convergence by local injectivity*, Inverse Problems, 24 (2008), 065009.
- [20] A. LECHLEITER AND A. RIEDER, *Towards a general convergence theory for inexact Newton regularizations*, Numer. Math., 114 (2010), pp. 521–548.
- [21] A. LEITÃO AND M. MARQUES ALVES, *On Landweber-Kaczmarz methods for regularizing systems of ill-posed equations in Banach spaces*, Inverse Problems, 28 (2012), 104008.
- [22] F. NATTERER, *Algorithms in ultrasound tomography*, in Radon Transforms and Tomography, South Hadley, MA, 2000, Contemp. Math. 278, AMS, Providence, RI, 2001, pp. 49–56.
- [23] F. NATTERER AND F. WÜBBELING, *Mathematical Methods in Image Reconstruction*, SIAM Monogr. Math. Model. Comput., SIAM, Philadelphia, 2001.
- [24] A. RIEDER, *On the regularization of nonlinear ill-posed problems via inexact Newton iterations*, Inverse Problems, 15 (1999), pp. 309–327.
- [25] F. SCHÖPFER, A.K. LOUIS, AND T. SCHUSTER, *Nonlinear iterative methods for linear ill-posed problems in Banach spaces*, Inverse Problems, 22 (2006), pp. 311–329.
- [26] T. SCHUSTER, B. KALTENBACHER, B. HOFMANN, AND K.S. KAZIMIERSKI, *Regularization Methods in Banach Spaces*, Radon Ser. Comput. Appl. Math. 10, Walter de Gruyter, Berlin, 2012.