IMA Journal of Numerical Analysis (2021) **41**, 2962–2989
https://doi.org/10.1093/imanum/draa050
Advance Access publication on 16 September 2020

Range-relaxed criteria for choosing the Lagrange multipliers in the Levenberg–Marquardt method

A. Leitão* and F. Margotti
Department of Mathematics, Federal University of St. Catarina, P.O. Box 476, 88040-900 Florianópolis, Brazil
*Corresponding author: acgleitao@gmail.com

and

B. F. Svaiter
IMPA, Estr.Dona Castorina 110, 22460-320 Rio de Janeiro, Brazil

[Received on 12 December 2019; revised on 28 May 2020]

In this article we propose a novel strategy for choosing the Lagrange multipliers in the Levenberg–Marquardt method for solving ill-posed problems modeled by nonlinear operators acting between Hilbert spaces. Convergence analysis results are established for the proposed method, including monotonicity of iteration error, geometrical decay of the residual, convergence for exact data, stability and semi-convergence for noisy data. Numerical experiments are presented for an elliptic parameter identification two-dimensional electrical impedance tomography problem. The performance of our strategy is compared with standard implementations of the Levenberg–Marquardt method (using *a priori* choice of the multipliers).

Keywords: nonlinear ill-posed problems; Levenberg–Marquardt method; Lagrange multipliers.

1. Introduction

In this article we address the Levenberg–Marquardt (LM) method (Levenberg, 1944; Marquardt, 1963), which is a well-established iterative method for obtaining stable approximate solutions of nonlinear ill-posed operator equations (Hanke, 1997; Deuflhard *et al.*, 1998) (see also the textbooks Engl *et al.*, 1996; Kaltenbacher *et al.*, 2008 and the references therein).

The novelty of our approach consists in adopting a range-relaxed criteria for the choice of the Lagrange multipliers in the LM method. Our approach is inspired in the recent paper (Boiger *et al.*, 2020), where a range-relaxed criteria was proposed for choosing the Lagrange multipliers in the iterated Tikhonov method for linear ill-posed problems.

With our strategy the new iterate is obtained as the projection of the current one onto a level set of the linearized residual function. This level belongs to an interval (or *range*), which is defined by the current (nonlinear) residual and by the noise level. As a consequence, the admissible Lagrange multipliers (in each iteration) shall belong to a nondegenerate interval instead of being a single value (see (1.4)). This fact reduces the computational burden of evaluating the multipliers. Moreover, under appropriate assumptions, the choice of the above-mentioned range enforces geometrical decay of the residual (see (3.5)).

The resulting method (see Section 2) proves, in the preliminary numerical experiments (see Section 4), to be more efficient than the classical geometrical choice of the Lagrange multipliers, typically used in implementations of LM-type methods.

1.1 *The model problem*

The *exact case* of the *inverse problem* we are interested in consists of determining an unknown quantity $x \in X$ from the set of data $y \in Y$, where X, Y are Hilbert spaces, and y is obtained by indirect measurements of the parameter x, this process being described by the model

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

with $F : D(F) \subseteq X \to Y$ being a nonlinear ill-posed operator. In practical situations one does not know the data exactly. Instead, an approximate measured data $y^\delta \in Y$, satisfying

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

is available, where $\delta > 0$ is the (known) noise level.

Standard methods for finding a solution of (1.1) are based in the use of *iterative type* regularization methods (Landweber, 1951; Hanke *et al.*, 1995; Engl *et al.*, 1996; Bakushinsky & Kokurin, 2004; Kaltenbacher *et al.*, 2008), which include the LM method or *Tikhonov type* regularization methods (Tikhonov, 1963; Tikhonov & Arsenin, 1977; Seidman & Vogel, 1989; Morozov, 1993; Scherzer, 1993; Engl *et al.*, 1996).

1.2 *The LM method*

In what follows we briefly revise the LM method, which was proposed separately by Levenberg (1944) and Marquardt (1963) for solving nonlinear optimization problems. The LM method for solving the nonlinear ill-posed operator equation (1.1) was originally considered in Deuflhard *et al.* (1998); Hanke (1997), and is defined by

$$x_{k+1}^\delta \;=\; \arg\min \left\{ \|y^\delta - F(x_k^\delta) - F'(x_k^\delta)(x - x_k^\delta)\|^2 + \alpha_k \|x - x_k^\delta\|^2 \right\},\; k = 0, 1, \ldots$$

Here $F'(z) : X \to Y$ is the Fréchet-derivative of F in $z \in D(F)$, $F'(z)^* : Y \to X$ is the corresponding adjoint operator, and $x_0^\delta \in X$ is some initial guess (possibly incorporating *a priori* knowledge about the exact solution(s) of $F(x) = y$). Moreover, $\{\alpha_k\}$ is a sequence of positive relaxation parameters (or Lagrange multipliers), aiming to guarantee convergence and stability of the iteration. This method can be summarized as follows:

$$x_{k+1}^\delta = x_k^\delta + h_k, \;\text{ with }\; h_k := \left(F'(x_k^\delta)^* F'(x_k^\delta) + \alpha_k I\right)^{-1} F'(x_k^\delta)^* (y^\delta - F(x_k^\delta)). \tag{1.3}$$

In the sequel we address some previous convergence analysis results:
(i) For exact data (i.e., $\delta = 0$) convergence is proved in Hanke (1997, Theorem 2.2), provided the operator F satisfies adequate regularity assumptions, and $\{\alpha_k\}$ satisfies the 'exact' condition

$$\|y^\delta - F(x_k^\delta) - F'(x_k^\delta) h_{k,\alpha_k}\|^2 \;=\; \theta\, \|y^\delta - F(x_k^\delta)\|^2, \tag{1.4}$$

where $h_{k,\alpha_k} = h_k(\alpha_k)$, is given by (1.3), and $\theta < 1$ is an appropriately chosen constant.[1] In the case of inexact data (i.e., $\delta > 0$) semi-convergence is proven if the iteration in (1.3) is stopped according to the discrepancy principle. The analysis presented in Hanke (1997) depends on a nonlinearity assumption on the operator F, namely the *strong tangential cone condition* (Kaltenbacher *et al.*, 2008).
(ii) In Baumeister *et al.* (2010) a convergence analysis for a Kaczmarz version of the LM method, using constant sequence $\{\alpha_k = \alpha\}$, is presented. The convergence proofs depend once again on a nonlinearity assumption on the operator F, namely the *weak tangential cone condition* (wTCC) (Hanke *et al.*, 1995; Engl *et al.*, 1996; Kaltenbacher *et al.*, 2008).
(iii) The algorithm `REGINN` is a Newton-like method for solving nonlinear inverse problems (Rieder, 1999). This iterative algorithm linearizes the forward operator around the current iterate, and subsequently applies a regularization technique in order to find an approximate solution to the linearized system, which in turn is added to the current iterate to provide an update. If wTCC holds true and the iteration is terminated by the discrepancy principle, then `REGINN` renders a regularization method in the sense of Engl *et al.* (1996). If Tikhonov regularization is used for approximating the solution of the linearized system, then `REGINN` becomes a variant of the LM method with a choice of the Lagrange multipliers performed *a posteriori*. In this case the resulting method is very similar to the one presented in Hanke (1997), but with the difference that the equality in (1.4) is replaced by an inequality.

1.3 *Criticism on the available choices of the Lagrange multipliers*

Although the proposed choice of $\{\alpha_k\}$ in Hanke (1997) is performed *a posteriori*, there is a severe drawback: the calculation of α_k in (1.4) cannot be performed explicitly. Moreover, computation of accurate numerical approximations for α_k is highly expensive.

For larger choices of the discrepancy constant alternative parameter choice rules are discussed in Hanke (1997), namely $\alpha_k = \alpha$ a positive constant or $\alpha_k := \|F'(x_k^\delta)\|^2$. However, the use of large values for discrepancy principle implies that the computation of small stopping indexes, meaning that LM iteration is interrupted before it can deliver the best possible approximate solution. On the other hand, the constant choice $\{\alpha_k = \alpha\}$ also has an intrinsic disadvantage: although the calculation of α demands no numerical effort, it does not lead to fast convergence of the sequence $\{x_k^\delta\}$ (this is observed in the numerical experiments presented in Baumeister *et al.*, 2010).

The Newton-type method proposed in Rieder (1999) also chooses the Lagrange multiplier within a range (see also Winkler & Rieder, 2015). However, differently from our criteria (2.5), this range is defined by a single inequality (Rieder, 1999, Inequality (2.6)). As a consequence, a regularization method (an inner iteration) is needed for the accurate computation of each multiplier.

In our method the computation of α_k requires knowledge about the noise level $\delta > 0$ and the wTCC constant $\eta \in [0, 1)$ (see Algorithm I). Other Newton-type methods (with *a posteriori* choice of α_k) also have this characteristic, e.g., see Rieder (1999, Lemma 3.2) and Hanke (1997, proof of Theorems 2.2 and 2.3).

1.4 *Outline of the manuscript*

In Section 2 we state the basic assumptions and introduce the range-relaxed criteria for choosing the Lagrange multipliers. The algorithm for the corresponding LM-type method is presented, and we prove some preliminary results, which guarantee that our method is well defined. In Section 3 we present the

[1] It is well known (cf Groetsch, 1984) that α_k is uniquely defined by (1.4).

main convergence analysis results, namely: convergence for exact data, stability and semiconvergence results. In Section 4 numerical experiments are presented for the EIT problem in a two-dimensional domain. We compare the performance of our method with other implementations of the LM method using classical (*a priori*) geometrical choices of the Lagrange multipliers. Section 5 is devoted to final remarks and conclusions.

2. Range-relaxed LM method

In this section we introduce a range-relaxed criteria for choosing the Lagrange multipliers in the LM method. Moreover, we present and discuss an algorithm for the resulting LM-type method, here called the *range-relaxed LM* (rrLM) method.

We begin this section by introducing the main assumptions used in this manuscript. It is worth mentioning that these assumptions are commonly used in the analysis of iterative regularization methods for nonlinear ill-posed problems (Scherzer, 1993; Engl *et al.*, 1996; Kaltenbacher *et al.*, 2008).

2.1 *Main assumptions*

Throughout this article we assume that the domain of definition $D(F)$ has nonempty interior and that the initial guess $x_0 \in X$ satisfies $B_\rho(x_0) \subset D(F)$ for some $\rho > 0$. Additionally,

(A1) The operator F and its Fréchet derivative F' are continuous. Moreover, there exists $C > 0$ such that

$$\|F'(x)\| \leq C, \qquad x \in B_\rho(x_0). \tag{2.1}$$

(A2) The wTCC holds at some ball $B_\rho(x_0)$, with $0 \leq \eta < 1$ and $\rho > 0$, i.e.,

$$\|F(\bar{x}) - F(x) - F'(x)(\bar{x} - x)\|_Y \leq \eta \, \|F(\bar{x}) - F(x)\|_Y, \qquad \forall \, x, \bar{x} \in B_\rho(x_0). \tag{2.2}$$

(A3) There exists $x^\star \in B_{\rho/2}(x_0)$ such that $F(x^\star) = y$, where $y \in Rg(F)$ are the exact data satisfying (1.2), i.e., $x^\star$ is an arbitrary solution (non-necessarily unique).

2.2 *An LM-type algorithm*

In what follows we introduce an iterative method, which derives from the choice of Lagrange multipliers proposed in this manuscript (see Step [3.1] of the Algorithm I).

Remark 2.1 Due to (A2) and (2.3) it follows $\tau > 1$. Moreover, $\big[\tau(1-\eta) - (1+\eta)\big](\eta\tau)^{-1} > 0$. Consequently, the interval used to define ε in (2.3) is nondegenerate.

Remark 2.2 For linear operators $F : X \to Y$, Assumption (A2) is trivially satisfied with $\eta = 0$. Thus, $c_k = \delta$, $d_k = p\delta + (1-p)\|F(x_k^\delta) - y^\delta\|$ and (2.5) reduces to

$$\|F x_k^\delta - y^\delta + F h_k^\delta\| = \|F x_{k+1}^\delta - y^\delta\| \in [c_k, d_k].$$

Consequently, the rrLM method in Algorithm I generalizes the *range-relaxed nonstationary iterated Tikhonov* method for linear ill-posed operator equations proposed in Boiger *et al.* (2020).

From now on we assume that $F'(x) \neq 0$ for $x \in B_\rho(x_0)$. Notice that this fact follows from Assumption (A2) provided F is nonconstant in $B_\rho(x_0)$.

Algorithm I: Range-relaxed LM.

[0] Choose an initial guess $x_0 \in X$; Set $k = 0$.

[1] Choose the positive constants τ, ε and p such that

$$\tau > \frac{1+\eta}{1-\eta}, \qquad 0 < \varepsilon < \frac{\tau(1-\eta)-(1+\eta)}{\eta\tau}, \qquad 0 < p < 1. \tag{2.3}$$

[2] If $\|F(x_0) - y^\delta\| \le \tau\delta$, then $k^* = 0$; Stop!

[3] For $k \ge 0$ do

[3.1] Compute $\alpha_k > 0$ and $h_k \in X$, such that

$$h_k = \left(F'(x_k^\delta)^* F'(x_k^\delta) + \alpha_k I\right)^{-1} F'(x_k^\delta)^* (y^\delta - F(x_k^\delta)) \tag{2.4}$$

$$\|y^\delta - F(x_k^\delta) - F'(x_k^\delta)h_k\| \ \in \ [c_k, d_k] \tag{2.5}$$

where

$$c_k = (1+\varepsilon)\eta\|F(x_k^\delta) - y^\delta\| + (1+\eta)\delta \tag{2.6}$$

$$d_k = p\,c_k + (1-p)\,\|F(x_k^\delta) - y^\delta\|. \tag{2.7}$$

[3.2] Set

$$x_{k+1}^\delta \ = \ x_k^\delta + h_k. \tag{2.8}$$

[3.3] If $\|F(x_k^\delta) - y^\delta\| \le \tau\delta$, then $k^* = k$; Stop!

[3.3] Else $k = k+1$; Go to Step [3].

The remainder of this section is devoted to verify that, under assumptions (A1), (A2) and (A3), Algorithm I is well defined (see Theorem 2.6). We open the discussion with Lemma 2.3, where a collection of preliminary results in functional and convex analysis is presented.

LEMMA 2.3 Suppose $A : X \to Y$ $(A \neq 0)$ is a continuous linear mapping, $\bar{z} \in X$, $b \in Y$ has a nonzero projection onto the closure of the range of A and define, for $\alpha > 0$,

$$z_\alpha = \arg\min_{z\in X} \|A(z-\bar{z}) - b\|^2 + \alpha\|z - \bar{z}\|^2. \tag{2.9}$$

The following assertions hold:

1. $z_\alpha = \bar{z} + (A^*A + \alpha I)^{-1}A^*b$;
2. $\alpha \mapsto \|A(z_\alpha - \bar{z}) - b\|$ is a continuous, strictly increasing function on $\alpha > 0$;
3. $\lim\limits_{\alpha\to 0} \|A(z_\alpha - \bar{z}) - b\| = \inf\limits_{z\in X} \|A(z-\bar{z}) - b\|$;

4. $\lim_{\alpha\to\infty} \|A(z_\alpha - \bar{z}) - b\| = \|b\|$;
5. $\|A(z_\alpha - \bar{z})\| \geq \|b\| - \|A(z_\alpha - \bar{z}) - b\| \geq 0$;
6. $\alpha \leq \|A^*b\|^2 \left[\|b\|(\|b\| - \|A(z_\alpha - \bar{z}) - b\|)\right]^{-1}$;
7. For $z \in X$ and $\alpha > 0$

$$\|z-\bar{z}\|^2 - \|z-z_\alpha\|^2 = \|z_\alpha - \bar{z}\|^2 + \frac{1}{\alpha}\left[\|A(z_\alpha - \bar{z}) - b\|^2 - \|A(z-\bar{z}) - b\|^2\right] + \frac{1}{\alpha}\|A(z-z_\alpha)\|^2; \tag{2.10}$$

8. For $z \in X$, $z \neq \bar{z}$ and $\alpha > 0$

$$\alpha \geq \frac{\|A(z_\alpha - \bar{z}) - b\|^2 - \|A(z - \bar{z}) - b\|^2}{\|z - \bar{z}\|^2}. \tag{2.11}$$

Proof. The proofs of items *1.* and *5.* are straightforward. For a proof of items *2.* to *4.* we refer the reader to Groetsch (1984). The proofs of items *6.* and *7.* are adaptations of proofs presented in Boiger *et al.* (2020) and item *8.* follows from item *7.* □

The next Lemma provides an auxiliary estimate, which is used in the proof of Proposition 2.5. This proposition is fundamental for establishing that, as long as the discrepancy is not reached (see Step [3.3] of Algorithm I), two key facts hold true: (i) it is possible to find a pair ($\alpha_k \in \mathbb{R}^+$, $h_k \in X$) solving (2.4), (2.5) in Step [3.1] of Algorithm I; (ii) for any sequence $\{x_k^\delta\}$ generated by Algorithm I, the *iteration error* $\|x^\star - x_k^\delta\|$ is monotonically decreasing in k.

Lemma 2.4 Let Assumptions (A2) and (A3) hold. Then for $x^\star$ as in (A3) it holds

$$\|F(x) - y^\delta + F'(x)(x^\star - x)\| \;\leq\; \eta\|F(x) - y^\delta\| + (1+\eta)\delta, \;\forall x \in B_\rho(x_0).$$

Proof. Since $x, x^\star \in B_\rho(x_0)$ it follows from (A2) that:

$$\begin{aligned}\|F(x) - y^\delta + F'(x)(x^\star - x)\| &= \|F(x) - F(x^\star) + F'(x)(x^\star - x) + F(x^\star) - y^\delta\| \\ &\leq \eta\|F(x) - F(x^\star)\| + \|F(x^\star) - y^\delta\| \\ &\leq \eta\big(\|F(x) - y^\delta\| + \|y^\delta - F(x^\star)\|\big) + \|F(x^\star) - y^\delta\|.\end{aligned}$$

The conclusion follows from this inequality, (A3) and (1.2). □

Proposition 2.5 Let Assumptions (A2) and (A3) hold. Given $x \in B_\rho(x_0)$, define

$$(0,+\infty) \ni \alpha \;\mapsto\; \xi_\alpha := arg\,min_{\xi\in X}\|F(x) - y^\delta + F'(x)(\xi - x)\|^2 + \alpha\|\xi - x\|^2 \in X. \tag{2.12}$$

1. For every $\alpha > 0$ it holds

$$\|F'(x)\|\;\|\xi_\alpha - x\| \geq \|F(x) - y^\delta\| - \|F(x) - y^\delta + F'(x)(\xi_\alpha - x)\|. \tag{2.13}$$

Additionally, if $\|F(x) - y^\delta\| > \tau\delta$, define the scalars

$$c := (1+\varepsilon)\eta\|F(x) - y^\delta\| + (1+\eta)\delta,$$
$$d := p[(1+\varepsilon)\eta\|F(x) - y^\delta\| + (1+\eta)\delta] + (1-p)\|F(x) - y^\delta\|,$$

and the set $J := \{\alpha > 0 : \|F(x) - y^\delta + F'(x)(\xi_\alpha - x)\| \in [c, d]\}$. Then 2. J is a nonempty, nondegenerate interval; 3. For $\alpha \in J$ and $x^\star$ as in (A3) it holds

$$\|x^\star - x\|^2 - \|x^\star - \xi_\alpha\|^2 \geq \|\xi_\alpha - x\|^2. \tag{2.14}$$

Proof. We adopt the notation: $z = x^\star$, $z_\alpha = \xi_\alpha$, $\bar{z} = x$, $b = y^\delta - F(x)$ and $A = F'(x)$.
Add 1.: Equation (2.13) follows from Lemma 2.3 (item 5.).
Add 2.: From the definition of ε and τ in (2.3) it follows that:

$$c < \left[\eta\tau + \tau(1-\eta) - (1+\eta)\right]\tau^{-1}\|F(x) - y^\delta\| + (1+\eta)\delta \leq \|F(x) - y^\delta\|$$

(the last inequality follows from $\delta \leq \tau^{-1}\|F(x) - y^\delta\|$). Since d is a proper convex combination of c and $\|F(x) - y^\delta\|$, we have

$$c < d < \|F(x) - y^\delta\|. \tag{2.15}$$

On the other hand, it follows from Lemma 2.4 that

$$\|F(x) - y^\delta + F'(x)(x^\star - x)\| \leq \eta\|F(x) - y^\delta\| + (1+\eta)\delta < c. \tag{2.16}$$

From (2.15), (2.16) it follows that:

$$\inf_z \|F(x) - y^\delta + F'(x)(z - x)\| < c < d < \|F(x) - y^\delta\|.$$

Assertion 2. follows from this inequality and Lemma 2.3 (items 2., 3. and 4.).
Add 3.: From (2.16) and the assumption $\alpha \in J$, we conclude that

$$\|F(x) - y^\delta + F'(x)(x^\star - x)\| < c \leq \|F(x) - y^\delta + F'(x)(\xi_\alpha - x)\|.$$

Assertion 3. follows from this inequality and Lemma 2.3 (item 7.). □

We are now ready to state and prove the main result of this section.

THEOREM 2.6 Let Assumptions (A1), (A2) and (A3) hold. Then Algorithm I is well defined, i.e., for $k < k^*$ (the stopping index defined in Step [3.3]) there exists a pair $(\alpha_k \in \mathbb{R}^+,\ h_k \in X)$ solving (2.4), (2.5). Moreover, k^* is finite and any sequence $\{x_k^\delta\}$ generated by this algorithm satisfies

$$\|x^\star - x_k^\delta\|^2 - \|x^\star - x_{k+1}^\delta\|^2 \geq \|x_k^\delta - x_{k+1}^\delta\|^2,\ 0 \leq k < k^*. \tag{2.17}$$

Proof. Let step $k = 0$ of Algorithm I be its initialization. We may assume $\|F(x_0) - y^\delta\| > \tau\delta$ (otherwise the algorithm stops with $k^* = 0$, and the theorem is trivial).

We use induction for proving this result. For $k = 0$ it follows from Proposition 2.5 (item 2.) with $x = x_0$, the existence of ($\alpha_0 \in \mathbb{R}^+$, $h_0 \in X$) solving (2.4), (2.5). Moreover, it follows from Proposition 2.5 (item 3.) with $x = x_0$ that (2.17) holds for $k = 0$.

Assume by induction that Algorithm I is well defined up to step $k_0 > 0$ and that (2.17) holds for $k = 0, \dots k_0 - 1$. There are two possible scenarios to consider:

- Case I: $\|F(x^\delta_{k_0}) - y^\delta\| \le \tau\delta$.
  In this case the algorithm terminates at iteration $k^* = k_0 \ge 1$, concluding the proof.
- Case II: $\|F(x^\delta_{k_0}) - y^\delta\| > \tau\delta$.
  Due to the inductive assumption, $\|x^\star - x^\delta_{k_0}\| \le \|x^\star - x^\delta_{k_0-1}\| \le \cdots \le \|x^\star - x^\delta_0\|$. From (A3) follows:

$$\|x^\delta_{k_0} - x^\delta_0\| \le \|x^\delta_{k_0} - x^\star\| + \|x^\star - x^\delta_0\| \le 2\|x^\star - x^\delta_0\| < \rho.$$

Hence, $x^\delta_{k_0} \in B_\rho(x_0)$. Proposition 2.5 (item 2.) with $x = x^\delta_{k_0}$, guarantees the existence of a pair ($\alpha_{k_0} \in \mathbb{R}^+$, $h_{k_0} \in X$) solving (2.4), (2.5) as well as the existence of $x^\delta_{k_0+1} \in X$. The validity of (2.17) for $k = k_0$ follows from Proposition 2.5 (item 3.) with $x = x^\delta_{k_0}$.

In order to verify the finiteness of the stopping index k^* notice that, from Proposition 2.5 (item 1.) with $x = x^\delta_k$, $\alpha = \alpha_k$ and $\xi_\alpha = x^\delta_{k+1}$, it follows:

$$\|F'(x^\delta_k)\|\, \|x^\delta_{k+1} - x^\delta_k\| \ge \|F(x^\delta_k) - y^\delta\| - \|F(x^\delta_k) - y^\delta + F'(x^\delta_k)(x^\delta_{k+1} - x^\delta_k)\|, \; k = 0, \dots k^* - 1.$$

From this inequality and the definition of c_k and d_k in Step [3.1], it follows that:

$$\begin{aligned}\|F'(x^\delta_k)\|\, \|x^\delta_{k+1} - x^\delta_k\| &\ge \|F(x^\delta_k) - y^\delta\| - d_k = p\big[\|F(x^\delta_k) - y^\delta\| - c_k\big] \\ &= p\big[(1 - (1+\varepsilon)\eta)\, \|F(x^\delta_k) - y^\delta\| - (1+\eta)\delta\big], \; k = 0, \dots k^* - 1.\end{aligned}$$

Since $\|F(x^\delta_k) - y^\delta\| > \tau\delta$, $0 \le k < k^*$ and $\varepsilon < \frac{1}{\eta} - 1$ (see (2.3)), we obtain from the last inequality

$$\|F'(x^\delta_k)\|\, \|x^\delta_{k+1} - x^\delta_k\| \ge p\big[(1 - (1+\varepsilon)\eta)\,\tau - (1+\eta)\big]\delta = p\delta\,\eta\tau\Big[\frac{\tau(1-\eta) - (1+\eta)}{\eta\tau} - \varepsilon\Big],$$

for $k = 0, \dots k^* - 1$. Now, Assumption (A1) implies

$$\|x^\delta_{k+1} - x^\delta_k\| \ge \frac{p\delta\,\eta\tau}{C}\Big[\frac{\tau(1-\eta) - (1+\eta)}{\eta\tau} - \varepsilon\Big] =: \Psi > 0, \; k = 0, \dots k^* - 1. \tag{2.18}$$

Adding up inequality (2.17) for $k = 0, \dots\ k^* - 1$ and using (2.18), we finally obtain

$$\|x^\star - x_0\|^2 > \|x^\star - x_0\|^2 - \|x^\star - x^\delta_{k^*}\|^2 > \textstyle\sum_{k=0}^{k^*-1} \|x^\delta_k - x^\delta_{k+1}\|^2 > k^*\Psi^2,$$

from where the finiteness of the stopping index k^* follows. □

REMARK 2.7 Assumption (A1) is used only once in the proof of Theorem 2.6, namely in the derivation of (2.18), which is used to prove finiteness of the stopping index k^*.

Corollary 2.8 Let Assumptions (A1), (A2) and (A3) hold, and assume the data is exact, i.e., $\delta = 0$. Then any sequence $\{x_k\}$ generated by Algorithm I satisfies

$$\textstyle\sum_{k=0}^{\infty} \|x_k - x_{k+1}\|^2 < \infty. \tag{2.19}$$

Proof. Adding up inequality (2.17), we obtain

$$\|x^\star - x_0\|^2 - \|x^\star - x_{n+1}\|^2 > \textstyle\sum_{k=0}^{n} \|x_k - x_{k+1}\|^2, \ \forall n > 0$$

and the assertion follows. □

We conclude this section obtaining an estimate for the Lagrange multipliers $\{\alpha_k\}$ defined in Step [3.1] of Algorithm I.

Proposition 2.9 Let Assumptions (A2) and (A3) hold. Then the Lagrange multipliers $\{\alpha_k\}$ in Algorithm I satisfy

$$\alpha_k \geq \rho^{-2}\varepsilon\eta\, \|F(x_k^\delta) - y^\delta\| \left[(1+\varepsilon)\eta\|F(x_k^\delta) - y^\delta\| + (1+\eta)\delta\right]. \tag{2.20}$$

Proof. Take $\alpha = \alpha_k$, $z_\alpha = x_{k+1}^\delta$, $\bar{z} = x_k^\delta$, $z = x^\star$, $b = y^\delta - F(x_k^\delta)$ and $A = F'(x_k^\delta)$. Arguing as in the proof of Lemma 2.4, we obtain

$$\|A(z - \bar{z}) - b\| \leq \eta\|b\| + (1+\eta)\delta. \tag{2.21}$$

On the other hand, it follows from Step [3.1] that

$$\|A(z_\alpha - \bar{z}) - b\| \geq (1+\varepsilon)\eta\|b\| + (1+\eta)\delta. \tag{2.22}$$

From (2.21) and (2.22) we obtain $\|A(z_\alpha - \bar{z}) - b\| - \|A(z - \bar{z}) - b\| \geq \varepsilon\eta\|b\|$. This last inequality together with (2.11) allow us to estimate

$$\begin{aligned}
\alpha_k &\geq \rho^{-2}\left[\|A(z_\alpha - \bar{z}) - b\|^2 - \|A(z - \bar{z}) - b\|^2\right] \\
&\geq \rho^{-2}\left[\|A(z_\alpha - \bar{z}) - b\| + \|A(z - \bar{z}) - b\|\right]\varepsilon\eta\,\|b\| \\
&\geq \rho^{-2}\varepsilon\eta\,\|b\|\,\|A(z_\alpha - \bar{z}) - b\|.
\end{aligned}$$

Estimate (2.20) follows from this inequality together with (2.22). □

3. Convergence analysis

We open this section obtaining an estimate, which is similar in spirit to Lemma 2.3 (item 7.).

LEMMA 3.1 Let Assumptions (A2) and (A3) hold. Then for $x^\star$ as in (A3) it holds

$$\|x^\star - x_k^\delta\|^2 - \|x^\star - x_{k+1}^\delta\|^2 \geq \|x_k^\delta - x_{k+1}^\delta\|^2 + 2\varepsilon\eta\alpha_k^{-1}\|F'(x_k)(x_{k+1}^\delta - x_k^\delta) + F(x_k^\delta) - y^\delta\|\,\|F(x_k^\delta) - y^\delta\|, \quad (3.1)$$

for $k = 0, \ldots, k^* - 1$.

Proof. The polarization identity yields

$$\|x^\star - x_k^\delta\|^2 - \|x^\star - x_{k+1}^\delta\|^2 = \|x_k^\delta - x_{k+1}^\delta\|^2 - 2\langle x_{k+1}^\delta - x_k^\delta, x_{k+1}^\delta - x^\star\rangle. \quad (3.2)$$

Adopting the notation $A := F'(x_k^\delta)$, $b := y^\delta - F(x_k^\delta)$ it follows from (2.4) and (2.8):

$$\begin{aligned}
-\langle x_{k+1}^\delta - x_k^\delta,\ x_{k+1}^\delta - x^\star\rangle &= \alpha_k^{-1}\langle A^*(Ah_k - b),\ x_{k+1}^\delta - x^\star\rangle \\
&= \alpha_k^{-1}\langle Ah_k - b,\ A[h_k - (x^\star - x_k^\delta)]\rangle \\
&= \alpha_k^{-1}\Big[\langle Ah_k - b, Ah_k - b\rangle - \langle Ah_k - b, A(x^\star - x_k^\delta) - b\rangle\Big] \\
&\geq \alpha_k^{-1}\Big[\|Ah_k - b\|^2 - \|Ah_k - b\|\,\|A(x^\star - x_k^\delta) - b\|\Big] \\
&= \alpha_k^{-1}\|Ah_k - b\|\Big[\|Ah_k - b\| - \|A(x^\star - x_k^\delta) - b\|\Big]. \quad (3.3)
\end{aligned}$$

However, from Lemma 2.4 (with $x = x_k^\delta$) and Algorithm I (see (2.6) and (2.5)), it follows:

$$\|A(x^\star - x_k^\delta) - b\| \ \leq\ \eta\,\|b\| + (1+\eta)\,\delta \ =\ c_k - \varepsilon\eta\,\|b\| \ \leq\ \|Ah_k - b\| - \varepsilon\eta\,\|b\|. \quad (3.4)$$

Thus, inequality (1) follows substituting (3.3) and (3.4) in (3.2). □

The following results are devoted to the analysis of the residuals $y^\delta - F(x_k^\delta)$ for a sequence $\{x_k^\delta\}$ generated by Algorithm I. In Proposition 3.2 we estimate the decay rate of the residuals. Moreover, in Proposition 3.4 we prove the summability of the series of squared residuals.

PROPOSITION 3.2 Let Assumptions (A2) and (A3) hold. Then for any sequence $\{x_k^\delta\}$ generated by Algorithm I, we have

$$\|y^\delta - F(x_{k+1}^\delta)\| \ \leq\ \Lambda\,\|y^\delta - F(x_k^\delta)\|, \quad (3.5)$$

for $k = 0, \ldots, k^* - 1$. Here $\Lambda := (C_1 + \eta)(1-\eta)^{-1}$, $C_1 := p(C_0 - 1) + 1$ and $C_0 := (1+\varepsilon)\eta + (1+\eta)\tau^{-1} < 1$. Additionally, if

$$\eta \ <\ \frac{p + \frac{p}{\tau}}{2 + p(1+\varepsilon) - \frac{p}{\tau}}, \quad (3.6)$$

then $\Lambda < 1$, from where it follows $k^* = O(|\ln\delta| + 1)$.[2]

[2] Here k^* is the stopping index defined in Step [3.3] of Algorithm I.

Proof. From Algorithm I (see (2.7)) and (A2), it follows:

$$\begin{aligned}\|y^\delta - F(x_{k+1}^\delta)\| &\le \|y^\delta - F(x_k^\delta) - F'(x_k^\delta)h_k\| + \|F(x_k^\delta) + F'(x_k^\delta)h_k - F(x_{k+1}^\delta)\| \\ &\le d_k + \eta\|F(x_k^\delta) - F(x_{k+1}^\delta)\| \\ &\le d_k + \eta\big(\|y^\delta - F(x_k^\delta)\| + \|y^\delta - F(x_{k+1}^\delta)\|\big),\ 0 \le k < k^*. \end{aligned} \tag{3.7}$$

On the other hand, Algorithm I (see (2.6)) implies $c_k \le C_0\|y^\delta - F(x_k^\delta)\|$, $0 \le k < k^*$. Consequently, $d_k \le C_1\|y^\delta - F(x_k^\delta)\|$, $0 \le k < k^*$. Substituting this inequality in (3.7) we obtain the estimate (3.5).

To prove the last assertion observe that $\Lambda < 1$ iff (3.6) holds true. Moreover, from Algorithm I (see Step [3.3]) and (3.5) follows $\tau\delta \le \|y^\delta - F(x_{k^*-1}^\delta)\| \le \Lambda^{k^*-1}\|y^\delta - F(x_0)\|$. Consequently, (3.6) implies $k^* \le (\ln\Lambda)^{-1}\ln\big(\tau\delta/\|y^\delta - F(x_0)\|\big) + 1$, completing the proof. □

REMARK 3.3 Inequality (3.6) holds true if $\eta < 1/3$, p is sufficiently close to 1, ε is sufficiently close to zero and τ is large enough. Notice that the condition $\eta < 1/3$ is not necessary for the convergence analysis devised in this manuscript.

PROPOSITION 3.4 Let Assumptions (A1), (A2) and (A3) hold. Suppose that no noise is present in the data (i.e., $\delta = 0$). Then for any sequence $\{x_k\}$ generated by Algorithm I we have

$$\sum_{k=0}^{\infty}\|y - F(x_k)\|^2 < \infty. \tag{3.8}$$

Proof. From Lemma 2.3 (item 6.) with $\alpha = \alpha_k$, $\bar z = x_k$, $z_\alpha = x_{x+1}$, $b = y - F(x_k)$ and $A = F'(x_k)$, follows:

$$\begin{aligned}\frac{1}{\alpha_k} &\ge \frac{\|y - F(x_k)\|\big(\|y - F(x_k)\| - \|F(x_k) - y - F'(x_k)h_k\|\big)}{\|F'(x_k)^*(y - F(x_k))\|^2} \\ &\ge \frac{\|y - F(x_k)\| - \|F(x_k) - y - F'(x_k)h_k\|}{C^2\|y - F(x_k)\|}\end{aligned} \tag{3.9}$$

(the last inequality follows from (A1)). Moreover, it follows from Algorithm I (see (2.5))

$$\|y - F(x_k)\| - \|F(x_k) - y - F'(x_k)h_k\| \ge \|y - F(x_k)\| - d_k \ge p\big(1 - (1+\varepsilon)\eta\big)\,\|y - F(x_k)\|$$

(notice that $(1 - (1+\varepsilon)\eta) > 0$ due to (2.3)). From this inequality (3.9) and (1) follows:

$$\begin{aligned}\|x^\star - x_0\|^2 &\ge \sum_{k=0}^{m}\tfrac{2\eta}{\alpha_k}\|F'(x_k)(x_{k+1} - x_k) + F(x_k) - y\|\,\|F(x_k) - y\| \\ &\ge \tfrac{2\eta p(1-(1+\varepsilon)\eta)}{C^2}\sum_{k=0}^{m}\|F'(x_k)(x_{k+1} - x_k) + F(x_k) - y\|\,\|F(x_k) - y\|,\end{aligned} \tag{3.10}$$

for all $m \in \mathbb{N}$. Finally, (3.8) follows from (3.10) and the inequality $\|F'(x_k)(x_{k+1} - x_k) + F(x_k) - y\| \ge c_k = (1+\varepsilon)\eta\|F(x_k) - y\|$ (see Algorithm I, (2.5) and (2.6)). □

REMARK 3.5 An immediate consequence of Proposition 3.4 is the fact that $\|F(x_k)-y\| \to 0$ as $k \to \infty$. It is worth noticing that (3.10) and Algorithm I also imply the summability of the series

$$\sum_{k=0}^{\infty} \|F\prime(x_k)(x_{k+1}-x_k)+F(x_k)-y\|^2 \text{ and } \sum_{k=0}^{\infty} \|F'(x_k)(x_{k+1}-x_k)+F(x_k)-y\|\,\|F(x_k)-y\|$$

(compare with Baumeister *et al.*, 2010, inequalities (18a), (18b), (18c)).

In the sequel we address the first main result of this section (see Theorem 3.7), namely convergence of Algorithm I in the exact data case (i.e., $\delta = 0$). To state this theorem we need the concept of x_0*−minimal-norm solution* of (1.1), i.e., the unique $x^\dagger \in X$ satisfying $\|x^\dagger - x_0\| := \inf\{\|x^* - x_0\| : F(x^*) = y \text{ and } x^* \in B_\rho(x_0)\}$.

REMARK 3.6 Due to (A2), given $x^* \in B_{\rho/2}(x_0)$ a solution of (1.1) and $z \in N(F'(x^*))$, the element $x^* + tz \in B_\rho(x_0)$ is also a solution of (1.1) for all $t \in (-\frac{\rho}{2}, \frac{\rho}{2})$.[3]

Due to (A3), $x^\dagger \in B_{\rho/2}(x_0)$. Thus, the inequality $\|x^\dagger - x_0\|^2 \leq \|(x^\dagger + tz) - x_0\|^2$ holds for all $t \in (-\frac{\rho}{2}, \frac{\rho}{2})$ and all $z \in N(F'(x^\dagger))$, from where we conclude[4]

$$x^\dagger - x_0 \in N(F'(x^\dagger))^\perp. \tag{3.11}$$

THEOREM 3.7 Let Assumptions (A1), (A2) and (A3) hold. Suppose that no noise is present in the data (i.e., $\delta = 0$). Then any sequence $\{x_k\}$ generated by Algorithm I either terminates after finitely many iterations with a solution of (1.1) or it converges to a solution of this equation as $k \to \infty$. Moreover, if

$$N(F'(x^\dagger)) \subset N(F'(x)),\ \forall x \in B_\rho(x_0) \tag{3.12}$$

holds, then $x_k \to x^\dagger$ as $k \to \infty$.

Proof. In what follows we adopt the notation $A_k := F'(x_k)$, $b_k := y - F(x_k)$. If for some $k \in \mathbb{N}$, $\|y - F(x_k)\| = 0$, then x_k is a solution and Algorithm I stops with $k^* = k$. Otherwise, $\{x_k\}_{k\in\mathbb{N}}$ is a Cauchy sequence. Indeed, fix $m < n$ and choose $\overline{k} \in \{m, \ldots, n\}$ s.t.

$$\|b_{\overline{k}}\| \leq \|b_k\| \text{ for all } k \in \{m, \ldots, n\}. \tag{3.13}$$

[3] Indeed, due to (A2) we have

$$\|F(x^* + tz) - y\| = \|F(x^* + tz) - F(x^*)\| \leq \frac{1}{1-\eta}\|F'(x^*)(x^* + tz - x^*)\| = \frac{|t|}{1-\eta}\|F'(x^*)\,z\| = 0.$$

[4] The conclusion follows from the fact that $\|x^\dagger - x_0\|^2 \leq \|(x^\dagger + tz) - x_0\|^2, \forall t \in (-\epsilon, \epsilon)$, implies $\langle x^\dagger - x_0, z\rangle = 0$.

From the triangle inequality and the polarization identity, it follows that for any $x^\star$ as in (A3):

$$\begin{aligned}\tfrac{1}{2}\,\|x_n - x_m\|^2 &\le \|x_n - x_{\overline{k}}\|^2 + \|x_m - x_{\overline{k}}\|^2 \\ &= \big(\|x^\star - x_n\|^2 - \|x^\star - x_{\overline{k}}\|^2 + 2\langle x_n - x_{\overline{k}}\, x_{\overline{k}} - x^\star\rangle\big) \\ &\quad + \big(\|x^\star - x_m\|^2 - \|x^\star - x_{\overline{k}}\|^2 + 2\langle x_m - x_{\overline{k}}\, x_{\overline{k}} - x^\star\rangle\big). \end{aligned} \tag{3.14}$$

Since the sequence $\{\|x^\star - x_n\|\}_{n\in\mathbb{N}}$ is non-negative and nonincreasing (see (2.17)) it converges. Therefore, the difference $\|x^\star - x_n\|^2 - \|x^\star - x_{\overline{k}}\|^2$ as well as $\|x^\star - x_m\|^2 - \|x^\star - x_{\overline{k}}\|^2$ both converge to zero as $m \to \infty$. It remains to estimate the inner products in (3.14). Notice that

$$\begin{aligned}\big|\langle x_n - x_{\overline{k}}\, x_{\overline{k}} - x^\star\rangle + \langle x_m - x_{\overline{k}}\, x_{\overline{k}} - x^\star\rangle\big| &\le \big|\langle x_n - x_m\, x_{\overline{k}} - x^\star\rangle\big| \\ &\le \sum_{k=m}^{n-1} \big|\langle x_{k+1} - x_k\, x_{\overline{k}} - x^\star\rangle\big| \\ &= \sum_{k=m}^{n-1} \tfrac{1}{\alpha_k}\big|\langle A_k^*(A_k h_k - b_k)\, x_{\overline{k}} - x^\star\rangle\big| \\ &\le \sum_{k=m}^{n-1} \tfrac{1}{\alpha_k}\|A_k h_k - b_k\|\,\|A_k(x_{\overline{k}} - x^\star)\|, \end{aligned} \tag{3.15}$$

with h_k as in (2.4). However, from (3.13) and (A2) follows:

$$\begin{aligned}\|A_k(x_{\overline{k}} - x^\star)\| &\le \|A_k(x_{\overline{k}} - x_k)\| + \|A_k(x^\star - x_k)\| \\ &\le \|F(x_{\overline{k}}) - F(x_k) - A_k(x_{\overline{k}} - x_k)\| + \|F(x_{\overline{k}}) - F(x_k)\| \\ &\quad + \|F(x^\star) - F(x_k) - A_k(x^\star - x_k)\| + \|F(x^\star) - F(x_k)\| \\ &\le (\eta+1)\,\|F(x_{\overline{k}}) - F(x_k)\| + (\eta+1)\,\|y - F(x_k)\| \\ &\le 2(\eta+1)\,\|y - F(x_k)\| + (\eta+1)\|y - F(x_{\overline{k}})\| \\ &\le 3(\eta+1)\,\|y - F(x_k)\|. \end{aligned}$$

Substituting this last inequality in (3.15) and using (1) (with $x_k^\delta = x_k$, $y^\delta = y$), we obtain

$$\begin{aligned}\big|\langle x_n - x_{\overline{k}}\, x_{\overline{k}} - x^\star\rangle + \langle x_m - x_{\overline{k}}\, x_{\overline{k}} - x^\star\rangle\big| &\le 3(\eta+1)\sum_{k=m}^{n-1} \tfrac{1}{\alpha_k}\|A_k h_k - b_k\|\,\|b_k\| \\ &\le \frac{3(\eta+1)}{2\varepsilon\eta}\sum_{k=m}^{n-1}\big(\|x^\star - x_k\|^2 - \|x^* - x_{k+1}\|^2\big) \\ &= \frac{3(\eta+1)}{2\varepsilon\eta}\big[\|x^\star - x_m\|^2 - \|x^\star - x_n\|^2\big] \to 0 \end{aligned}$$

as $m \to \infty$. Thus, it follows from (3.14) that $\|x_n - x_m\| \to 0$ as $m \to \infty$, proving that $\{x_k\}_{k\in\mathbb{N}}$ is indeed a Cauchy sequence.

Since X is complete $\{x_k\}$ converges to some $x_\infty \in X$ as $k \to \infty$. On the other hand, $\|y - F(x_k)\| \to 0$ as $k \to \infty$ (see Remark 3.5). Consequently, x_∞ is a solution of (1.1) proving the first assertion.

In order to prove the last assertion notice that, if (3.12) hold, then

$$x_{k+1} - x_k = \alpha_k^{-1} A_k^*(A_k h_k - b_k) \in R(F'(x_k)^*) \subset N(F'(x_k))^\perp \subset N(F'(x^\dagger))^\perp, \; k = 0, 1, \ldots,$$

from where we conclude that $x_k - x_0 \in N(F'(x^\dagger))^\perp$, $k \in \mathbb{N}$. Since $x^\dagger - x_0 \in N(F'(x^\dagger))^\perp$ (see (3.11)) it follows that $x_k - x^\dagger \in N(F'(x^\dagger))^\perp$, $k \in \mathbb{N}$. Consequently, $x_\infty - x^\dagger = \lim_k x_k - x^\dagger \in N(F'(x^\dagger))^\perp$. However, (A2) implies $\|F'(x^\dagger)(x_\infty - x^\dagger)\| \le (1+\eta)\|F(x_\infty) - F(x^\dagger)\| = 0$, from what follows $x_\infty - x^\dagger \in N(F'(x^\dagger))$. Thus, $x_\infty - x^\dagger = 0$. □

We conclude this section addressing the last two main results, namely: stability (Theorem 3.9) and semi-convergence (Theorem 3.10). The following definition is quintessential for the discussion of these results.

DEFINITION 3.8 A vector $z \in X$ is a **successor** of x_k^δ if

- $k < k^*$.
- There exists $(\alpha_k > 0,\ h_k \in X)$ satisfying (2.4), (2.5), such that $z = x_k^\delta + h_k$.

Notice that Theorem 3.7 guarantees that the sequence $\{x_k\}_{k\in\mathbb{N}}$ converges to a solution of $F(x) = y$ whenever x_{k+1} is a successor of x_k for every $k \in \mathbb{N}$. In this situation we call $\{x_k\}_{k\in\mathbb{N}}$ a *noiseless sequence*.

THEOREM 3.9 (Stability). Let Assumptions (A1), (A2) and (A3) hold, and $\{\delta_j\}_{j\in\mathbb{N}}$ be a positive zero-sequence. Assume that the (finite) sequences $\{x_k^{\delta_j}\}_{0\le k\le k^*(\delta_j)}$, $j \in \mathbb{N}$, are fixed,[5] where $x_{k+1}^{\delta_j}$ is a successor of $x_k^{\delta_j}$. Then there exists a noiseless sequence $\{x_k\}_{k\in\mathbb{N}}$ such that, for every fixed $k \in \mathbb{N}$, there exists a subsequence $\{\delta_{j_m}\}_{m\in\mathbb{N}}$ (depending on k), satisfying

$$x_\ell^{\delta_{j_m}} \to x_\ell \text{ as } m \to \infty, \qquad \text{for } \ell = 0, \ldots, k.$$

Proof. We use an inductive argument. Since $x_0^\delta = x_0$ for every $\delta \ge 0$ the assertion is clear for $k = 0$. Our main argument consists of repeatedly choosing a subsequence of the current subsequence. In order to avoid a notational overload we denote a subsequence of $\{\delta_j\}_j$ again by $\{\delta_j\}_j$.

Suppose by induction that the assertion holds true for some $k \in \mathbb{N}$, i.e., that there exists a subsequence $\{\delta_j\}_j$ and $\{x_\ell\}_{\ell=0}^k$, satisfying

$$x_\ell^{\delta_j} \to x_\ell \text{ as } j \to \infty, \qquad \text{for } \ell = 0, \ldots, k,$$

where $k < k^*(\delta_j)$ and $x_{\ell+1}$ is a successor of x_ℓ, for $\ell = 0, \ldots, k-1$. Since $x_{k+1}^{\delta_j}$ is a successor of $x_k^{\delta_j}$ (for each δ_j) there exists (for each δ_j) a positive number $\alpha_k^{\delta_j}$ such that $x_{k+1}^{\delta_j} = x_k^{\delta_j} + h_k^{\delta_j}$, with $h_k^{\delta_j}$ as in (2.4) and

$$\left\| F(x_k^{\delta_j}) - y^{\delta_j} + F'(x_k^{\delta_j}) h_k^{\delta_j} \right\| \in [c_k^{\delta_j}, d_k^{\delta_j}]. \tag{3.16}$$

[5] Notice that the stopping index k^* in Step [3.3] depends on δ, i.e., $k^* = k^*(\delta)$.

Our next goal is to prove the existence of a successor x_{k+1} of x_k and of a subsequence $\{\delta_j\}_j$ of the current subsequence such that $x_{k+1}^{\delta_j} \to x_{k+1}$ as $j \to \infty$, ensuring that

$$x_\ell^{\delta_j} \to x_\ell \text{ as } j \to \infty, \quad \text{for } \ell = 0, \dots, k+1, \tag{3.17}$$

and completing the inductive argument. We divide this proof in four steps as follows:

Step 1. We find a vector $z \in X$ such that, for some subsequence $\{\delta_j\}_j$

$$h_k^{\delta_j} \rightharpoonup z \text{ as } j \to \infty. \tag{3.18}$$

Step 2. We define

$$\alpha_k := \liminf_{j\to\infty} \alpha_k^{\delta_j} \tag{3.19}$$

and prove that $\alpha_k > 0$, which permit us to define h_k as in (2.4) as well as $x_{k+1} := x_k + h_k$.

Step 3. We show that $h_k = z$, which ensures that $h_k^{\delta_j} \rightharpoonup h_k$.

Step 4. We validate that

$$\|h_k^{\delta_j}\| \to \|h_k\|, \text{ as } j \to \infty, \tag{3.20}$$

which together with $h_k^{\delta_j} \rightharpoonup h_k$ proves that $h_k^{\delta_j} \to h_k$ and, consequently, $x_{k+1}^{\delta_j} \to x_{k+1}$.

Finally, we prove that x_{k+1} is a successor of x_k, which validates (3.17).

Proof of Step 1. Since the sequence $\{h_k^{\delta_j}\}_{j\in\mathbb{N}}$ is bounded (see (2.17)) there exists a subsequence $\{\delta_j\}$ of the current subsequence and a vector $z \in X$ such that (3.18) holds. Consequently,

$$A_k^{\delta_j} h_k^{\delta_j} - b_k^{\delta_j} \;\rightharpoonup\; A_k z - b_k, \text{ as } j \to \infty \tag{3.21}$$

(here $A_k^{\delta_j} = F'(x_k^{\delta_j})$, $A_k = F'(x_k)$, $b_k^{\delta_j} = y^{\delta_j} - F(x_k^{\delta_j})$, $b_k = y - F(x_k)$).

Proof of Step 2. If α_k in (3.19) is not positive, we conclude from (A2)

$$\begin{aligned}
\liminf_j \|A_k^{\delta_j} h_k^{\delta_j} - b_k^{\delta_j}\|^2 &\le \liminf_j T_{k,\delta_j,\alpha_k^{\delta_j}}(h_k^{\delta_j}) \le \liminf_j T_{k,\delta_j,\alpha_k^{\delta_j}}(x^\dagger - x_k) \\
&= \liminf_j \left(\|A_k^{\delta_j}(x^\dagger - x_k) - b_k^{\delta_j}\|^2 + \alpha_k^{\delta_j}\|x^\dagger - x_k\|^2\right) \\
&= \|A_k(x^\dagger - x_k) - b_k\|^2 \;\le\; \eta^2 \|b_k\|^2
\end{aligned}$$

(here $T_{k,\delta,\alpha}(h) := \|F'(x_k^\delta)h - y^\delta + F(x_k^\delta)\|^2 + \alpha\|h\|^2$). This leads to the contradiction

$$c_k = \lim_j c_k^{\delta_j} \le \liminf_j \|A_k^{\delta_j} h_k^{\delta_j} - b_k^{\delta_j}\| \le \eta\|b_k\| < c_k.$$

Thus, $\alpha_k > 0$ holds. We define $T_{k,\alpha}(h) := T_{k,\delta,\alpha}(h)$ with $\delta = 0$, $h_k := \arg\min_{h\in X} T_{k,\alpha_k}(h)$ and $x_{k+1} := x_k + h_k$. In order to prove that x_{k+1} is a successor of x_k it is necessary to prove that

$$c_k \leq \|A_k h_k - b_k\| \leq d_k. \tag{3.22}$$

We first prove that $h_k = z$ (see Step 3).
Proof of Step 3. From (3.18), (3.21) and (3.19), it follows:

$$T_{k,\alpha_k}(z) = \|A_k z - b_k\|^2 + \alpha_k \|z\|^2 \;\leq\; \liminf_j \left(\left\| A_k^{\delta_j} h_k^{\delta_j} - b_k^{\delta_j} \right\|^2 + \alpha_k^{\delta_j} \left\| h_k^{\delta_j} \right\|^2 \right)$$
$$= \liminf_j T_{k,\delta_j,\alpha_k^{\delta_j}}(h_k^{\delta_j}) \;\leq\; \liminf_j T_{k,\delta_j,\alpha_k^{\delta_j}}(h_k) \;=\; T_{k,\alpha_k}(h_k).$$

Since h_k is the unique minimizer of T_{k,α_k} we conclude that $h_k = z$. Thus, $h_k^{\delta_j} \rightharpoonup h_k$ as $j \to \infty$. The last inequalities also ensure that $\liminf_j T_{k,\delta_j,\alpha_k^{\delta_j}}(h_k^{\delta_j}) = T_{k,\alpha_k}(h_k)$. This guarantees the existence of a subsequence, satisfying

$$\lim_{j\to\infty} T_{k,\delta_j,\alpha_k^{\delta_j}}(h_k^{\delta_j}) \;=\; T_{k,\alpha_k}(h_k). \tag{3.23}$$

Proof of Step 4. The goal is to validate (3.20), which, together with $h_k^{\delta_j} \rightharpoonup h_k$, imply $h_k^{\delta_j} \to h_k$. Consequently, (3.22) follows from (3.16). This ensures that x_{k+1} is a successor of x_k and validates (3.17), completing the proof of the theorem.

We first prove the existence of a constant $\alpha_{\max,k}$ such that

$$\alpha_k^{\delta_j} \leq \alpha_{\max,k} \quad \text{for all } j \in \mathbb{N}.$$

Indeed, if such a constant did not exist, we could find a subsequence satisfying $\alpha_k^{\delta_j} \to \infty$ as $j \to \infty$. Thus, since

$$\alpha_k^{\delta_j} \|h_k^{\delta_j}\|^2 \;\leq\; T_{k,\delta_k,\alpha_k^{\delta_j}}(h_k^{\delta_j}) \;\leq\; T_{k,\delta_k,\alpha_k^{\delta_j}}(0) \;=\; \|b_k^{\delta_j}\|^2,$$

we would have,

$$\lim_{j\to\infty} \alpha_k^{\delta_j} \|h_k^{\delta_j}\|^2 \;\leq\; \|b_k\|^2 \;<\; \infty,$$

which would imply $h_k^{\delta_j} \to 0$. Consequently,

$$\lim_{j\to\infty} \|A_k^{\delta_j} h_k^{\delta_j} - b_k^{\delta_j}\| \;=\; \|b_k\| \;\; d_k \;=\; \lim_{j\to\infty} d_k^{\delta_j},$$

which would imply the contradiction $\|A_k^{\delta_j} h_k^{\delta_j} - b_k^{\delta_j}\| \;\; d_k^{\delta_j}$, for j large enough.

Now we validate (3.20). This proof follows the lines of Margotti & Rieder (2015, Lemma 5.2). Define

$$a_j := \|h_k^{\delta_j}\|^2, \quad a := \limsup a_j, \quad c := \|h_k\|^2, \quad re_j := \|A_k^{\delta_j} h_k^{\delta_j} - b_k^{\delta_j}\|^2, \quad re := \liminf re_j.$$

As $\|h_k\| \leq \liminf \|h_k^{\delta_j}\|$ it suffices to prove that $a \leq c$. Assume the contrary. From (3.23) there exists a number $N_1 \in \mathbb{N}$ such that

$$j \geq N_1 \implies T_{k,\delta_j,\alpha_k^{\delta_j}}(h_k^{\delta_j}) < T_{k,\alpha_k}(h_k) + \alpha_k \frac{a-c}{2}. \tag{3.24}$$

From definition of lim inf there exist constants $N_2, N_3 \in \mathbb{N}$ such that

$$j \geq N_2 \implies re_j \geq re - \alpha_k(a-c)/6 \tag{3.25}$$

and

$$j \geq N_3 \implies \alpha_k^{\delta_j} \geq \alpha_k - \alpha_k(a-c)/6a. \tag{3.26}$$

Moreover, from definition of lim sup, we conclude that for each $M \in \mathbb{N}$ fixed, there exists an index $j \geq M$ such that

$$a_j \geq a - \alpha_k(a-c)/(6\alpha_{\max,k}). \tag{3.27}$$

Therefore, for $M := \max\{N_1, N_2, N_3\}$, there exists an index $j \geq M$ such that

$$\begin{aligned}
T_{k,\alpha_k}(h_k) &\leq re + \alpha_k c = re + (\alpha_k - \alpha_k^{\delta_j})a + \alpha_k^{\delta_j}(a - a_j) + \alpha_k^{\delta_j} a_j - \alpha_k(a-c) \\
&\leq (re_j + \alpha_k \tfrac{1}{6}(a-c)) + \alpha_k \tfrac{1}{6}(a-c) + \alpha_k \tfrac{1}{6}(a-c) + \alpha_k^{\delta_j} a_j - \alpha_k(a-c) \\
&= re_j + \alpha_k^{\delta_j} a_j - \alpha_k \tfrac{1}{2}(a-c) = T_{k,\delta_j,\alpha_k^{\delta_j}}(h_k^{\delta_j}) - \alpha_k \tfrac{1}{2}(a-c) < T_{k,\alpha_k}(h_k),
\end{aligned}$$

where the second inequality follows from (3.25), (3.26), (3.27), while the last inequality follows from (3.24). This leads to the obvious contradiction $T_{k,\alpha_k}(h_k) < T_{k,\alpha_k}(h_k)$, proving that $a \leq c$ as desired. Thus, (3.20) holds and the proof is complete. □

THEOREM 3.10 (Regularization). Let Assumptions (A1), (A2) and (A3) hold, and $\{\delta_j\}_{j\in\mathbb{N}}$ be a positive zero-sequence. Assume that the (finite) sequences $\{x_k^{\delta_j}\}_{0\leq k\leq k^*(\delta_j)}$, $j \in \mathbb{N}$, are fixed, where $x_{k+1}^{\delta_j}$ is a successor of $x_k^{\delta_j}$. Then every subsequence of $\{x_{k^*(\delta_j)}^{\delta_j}\}_{j\in\mathbb{N}}$ has itself a subsequence converging strongly to a solution of (1.1).

Proof. Since any subsequence of $\{\delta_j\}_{j\in\mathbb{N}}$ is itself a positive zero-sequence, it suffices to prove that $\{x_{k^*(\delta_j)}^{\delta_j}\}_{j\in\mathbb{N}}$ has a subsequence converging to a solution. We consider two cases:

Case 1. The sequence $\{k^*(\delta_j)\}_{j\in\mathbb{N}}$ is bounded. Thus, there exists a constant $M \in \mathbb{N}$ such that $k^*(\delta_j) \leq M$ for all $j \in \mathbb{N}$. Thus, the sequence $\{x_{k^*(\delta_j)}^{\delta_j}\}_{j\in\mathbb{N}}$ splits into at most $M + 1$ subsequences having the form $\{x_m^{\delta_{j_n}}\}_{n\in\mathbb{N}}$, with fixed $m \leq M$. Pick one of these subsequences. From Theorem 3.9 this subsequence has itself a subsequence (again denoted by $\{x_m^{\delta_{j_n}}\}_{n\in\mathbb{N}}$) converging to some $x_m \in X$, i.e.,

$$\lim_{n\to\infty} x_{k^*(\delta_{j_n})}^{\delta_{j_n}} = \lim_{n\to\infty} x_m^{\delta_{j_n}} = x_m.$$

Notice that x_m is a solution of (1.1). Indeed,

$$\begin{aligned}\|y - F(x_m)\| &= \lim_{n\to\infty} \|y - F(x^{\delta_{j_n}}_{k^*(\delta_{j_n})})\| \\ &\le \lim_{n\to\infty} \left(\|y - y^{\delta_{j_n}}\| + \|y^{\delta_{j_n}} - F(x^{\delta_{j_n}}_{k^*(\delta_{j_n})})\|\right) \\ &\le \lim_{n\to\infty} (\tau + 1)\, \delta_{j_n} \;=\; 0.\end{aligned}$$

Case 2. The sequence $\{k^*(\delta_j)\}_{j\in\mathbb{N}}$ is not bounded. Thus, there is a subsequence such $k^*(\delta_j) \to \infty$ as $j \to \infty$. Let $\varepsilon > 0$ be given and consider the noiseless sequence $\{x_k\}_{k\in\mathbb{N}}$ constructed in last theorem. Since x_{k+1} is a sucessor of x_k for all $k \in \mathbb{N}$, $\{x_k\}_{k\in\mathbb{N}}$ converges to some solution x^* of (1.1) (see Theorem 3.7). Then there exists $M = M(\varepsilon) \in \mathbb{N}$ such that

$$\|x_M - x^*\| \;<\; \tfrac{1}{2}\,\varepsilon.$$

On the other hand, there exists $J \in \mathbb{N}$ such that $k^*(\delta_j) \ge M$, for $j \ge J$. Consequently, it follows from the monotonicity of the iteration error (see Theorem 2.6) that:

$$j \ge J \implies \|x^{\delta_j}_{k^*(\delta_j)} - x^*\| \le \|x^{\delta_j}_M - x^*\|.$$

Moreover, it follows from Theorem 3.9, the existence of a subsequence $\{\delta_{j_m}\}$ (depending on $M(\varepsilon)$) and the existence of $N \in \mathbb{N}$ such that

$$m \ge N \implies \|x^{\delta_{jm}}_M - x_M\| \;<\; \tfrac{1}{2}\varepsilon.$$

Consequently, for $m \ge \max\{J, N\}$ (which simultaneously guarantees $j_m \ge m \ge J$ and $m \ge N$), it holds

$$\|x^{\delta_{jm}}_{k^*(\delta_{jm})} - x^*\| \le \|x^{\delta_{jm}}_M - x^*\| \le \|x^{\delta_{jm}}_M - x_M\| + \|x_M - x^*\| \;<\; \varepsilon. \tag{3.28}$$

Notice that the subsequence $\{\delta_{j_m}\}$ depends on ε. We now construct an ε-independent subsequence using a diagonal argument: for $\varepsilon = 1$ in (3.28), there is a subsequence of $\{\delta_j\}$ (called again $\{\delta_j\}$) and $j_1 \in \mathbb{N}$ such that

$$\|x^{\delta_{j_1}}_{k^*(\delta_{j_1})} - x^*\| \;<\; 1.$$

Now, for $\varepsilon = 1/2$, there exists a subsequence $\{\delta_j\}$ of the previous one, and $j_2 > j_1$ such that

$$\|x^{\delta_{j_2}}_{k^*(\delta_{j_2})} - x^*\| \;<\; 2^{-1}.$$

Arguing in this way we construct a subsequence $\{\delta_{j_n}\}_{n\in\mathbb{N}}$, satisfying

$$\|x^{\delta_{jn}}_{k^*(\delta_{jn})} - x^*\| \;<\; n^{-1},$$

from what follows $\lim x^{\delta_{jn}}_{k^*(\delta_{jn})} = x^*$. □

REMARK 3.11 If the solution of (1.1) referred to in Theorem 3.9 were independent of the chosen subsequence, then any subsequence of $\{x^{\delta_j}_{k^*(\delta_j)}\}_{j\in\mathbb{N}}$ would have itself a subsequence converging to the same solution. This would be enough to ensure that the whole sequence $\{x^{\delta_j}_{k^*(\delta_j)}\}_{j\in\mathbb{N}}$ converges to x^*.

However, x^* in the above proof depends on the noiseless sequence $\{x_n\}_{n\in\mathbb{N}}$ (whose existence is guaranteed by Theorem 3.9), which in turn depends on the fixed sequences $\{x^{\delta_j}_k\}_{0\le k\le k^*(\delta_j)}$, $j \in \mathbb{N}$. Consequently, if different subsequences of $\{\delta_j\}_{j\in\mathbb{N}}$ are chosen, the solution of (1.1) referred to in Theorem 3.9 can be different.

COROLLARY 3.12 Under the assumptions of Theorem 3.9 the following assertions hold true:

1. The sequence $\{x^{\delta_j}_{k^*(\delta_j)}\}_{j\in\mathbb{N}}$ splits into convergent subsequences, each one converges to a solution of (1.1).
2. If $x^\star$ in (A3) is the unique solution of (1.1) in $B_\rho(x_0)$, then $x^{\delta_j}_{k^*(\delta_j)} \to x^\star$ as $j \to \infty$.
3. If the null-space condition (3.12) holds, then $\{x^{\delta_j}_{k^*(\delta_j)}\}$ converges to the x_0-minimal-norm solution $x^\dagger$ as $j \to \infty$.

Proof. The proof of Assertion 1. is straightforward. Assertion 2. follows from the fact that if $x^\star$ is the unique solution of (1.1) in $B_\rho(x_0)$, then any subsequence of $\{x^{\delta_j}_{k^*(\delta_j)}\}_{j\in\mathbb{N}}$ has itself a subsequence converging to $x^\star$. To prove Assertion 3. notice that if (3.12) holds, then any noiseless sequence $\{x_n\}_{n\in\mathbb{N}}$ converges to $x^\dagger$ (Theorem 3.7). Thus, any subsequence of $\{x^{\delta_j}_{k^*(\delta_j)}\}_{j\in\mathbb{N}}$ has itself a subsequence converging to $x^\dagger$ and the proof follows. □

4. Numerical experiments

4.1 *The model problem and its discretization*

We test the performance of our method applying it to the nonlinear and ill-posed inverse problem of EIT introduced by Calderón (1980). A survey article concerning this problem is Borcea (2002).

Let $\Omega \subset \mathbb{R}^2$ be a bounded and simply connected Lipschitz domain. The EIT problem consists in applying different configurations of electric currents on the boundary of Ω and then reading the resulting voltages on the boundary of Ω as well. The objective is recovering the electric conductivity in the whole of set Ω. This problem is governed by the variational equation

$$\int_\Omega \gamma \nabla u \nabla \varphi = \int_{\partial\Omega} g\varphi \quad \text{for all } \varphi \in H^1_\diamond(\Omega), \tag{4.1}$$

where $g\colon \partial\Omega \to \mathbb{R}$ represents the electric current, $\gamma : \Omega \to \mathbb{R}$ is the electric conductivity and $u\colon \Omega \to \mathbb{R}$ represents the electric potential. Employing the Lax–Milgram lemma one can prove that for each $g \in L^2_\diamond(\partial\Omega) := \{v \in L^2(\partial\Omega) : \int_{\partial\Omega} v = 0\}$ and $\gamma \in L^\infty_+(\Omega) := \{v \in L^\infty(\Omega) : v \ge c > 0 \text{ a.e. in } \Omega\}$ fixed, there exists a unique $u \in H^1_\diamond(\Omega) := \{v \in H^1(\Omega) : \int_{\partial\Omega} v = 0\}$ satisfying (4.1). The voltage $f\colon \partial\Omega \to \mathbb{R}$ is the trace of the potential u ($f = u|_{\partial\Omega}$), which belongs to $L^2_\diamond(\partial\Omega)$.

For a fixed conductivity $\gamma \in L^\infty_+(\Omega)$ the bounded linear operator $\Lambda_\gamma : L^2_\diamond(\partial\Omega) \to L^2_\diamond(\partial\Omega), g \mapsto f$, which associates the electric current with the resulting voltage, the so-called *Neumann-to-Dirichlet* map

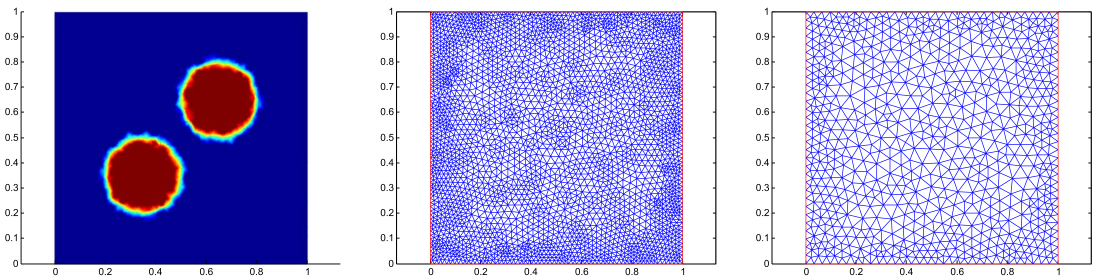

FIG. 1. Left: sought solution. Middle: mesh used to generate the data. Right: mesh used to solve the inverse problem.

(in short NtD). The *forward operator* associated with EIT is defined by

$$\mathcal{F}(\gamma) = \Lambda_\gamma, \tag{4.2}$$

with $\mathcal{F}\colon L^\infty_+(\Omega) \subset L^\infty(\Omega) \to \mathcal{L}(L^2_\diamond(\partial\Omega), L^2_\diamond(\partial\Omega))$. The EIT inverse problem consists in finding γ in above equation for a given Λ_γ. However, in practical situations, only a part of the data can be observed and therefore the NtD map is not completely available. One has to apply $d \in \mathbb{N}$ currents $g_j \in L^2_\diamond(\partial\Omega)$, $j = 1, \ldots, d$ and then record the resulting voltages $f_j = \Lambda_\gamma g_j$. We thus fix the vector $(g_1, \ldots, g_d) \in (L^2_\diamond(\partial\Omega))^d$ and introduce the operator $F\colon L^\infty_+(\Omega) \subset L^\infty(\Omega) \to (L^2_\diamond(\partial\Omega))^d$, $\gamma \mapsto (\Lambda_\gamma g_1, \ldots, \Lambda_\gamma g_d)$, which is Fréchet-differentiable[6] (see, e.g., Lechleiter & Rieder, 2008).

Since an analytical solution of (4.1) is not available, in general, the inverse problem needs to be solved with help of a computer. For this reason we construct a triangulation for Ω, $\mathcal{T} = \{T_i : i = 1, \ldots, M\}$, with $M = 1476$ triangles (see the third picture in Fig. 1) and approximate γ by piecewise constant conductivities: define the finite dimensional space $V := (\operatorname{span}\{\chi_{T_1}, \ldots, \chi_{T_M}\}, \|\cdot\|_{L^2(\Omega)})$.[7] We now search the conductivity in V, which means that our reconstructions always have the form $\sum_{i=1}^M \theta_i \chi_{T_i}$, with $(\theta_1, \ldots, \theta_M) \in \mathbb{R}^M$. With this new framework our forward operator reads

$$F\colon \widetilde{V} \subset V \to (L^2(\partial\Omega))^d, \quad \gamma \mapsto (\Lambda_\gamma g_1, \ldots, \Lambda_\gamma g_d), \tag{4.3}$$

where $\widetilde{V} = L^\infty_+(\Omega) \cap V$.

It is still unclear whether the forward operator associated with the continuous model of EIT, defined in (4.2), satisfies the tangential cone condition (2.2), but the version presented in the restricted set (4.3) guarantees this result, at least in a small ball around a solution, see Lechleiter & Rieder (2008). The Fréchet derivative of F, $F'\colon \operatorname{int}(\widetilde{V}) \to \mathcal{L}(V, (L^2(\partial\Omega))^d)$, satisfies $F'(\gamma)h = (w_1|_{\partial\Omega}, \ldots, w_d|_{\partial\Omega})$, where $w_j \in H^1_\diamond(\Omega)$ is the unique solution of

$$\int_\Omega \gamma \nabla w_j \nabla \varphi = -\int_\Omega h \nabla u_j \nabla \varphi \quad \text{for all } \varphi \in H^1_\diamond(\Omega), \tag{4.4}$$

[6] Equipped with an inner product defined in a very natural way, induced by the inner product in $L^2(\partial\Omega)$, the space $(L^2_\diamond(\partial\Omega))^d$ is a Hilbert space.

[7] Notice that $\operatorname{span}\{\chi_{T_1}, \ldots, \chi_{T_M}\} \subset L^\infty(\Omega)$.

with u_j solving (4.1) for $g = g_j$. The adjoint operator $F'(\gamma)^*\colon (L^2(\partial\Omega))^d \to V$ is given by

$$F'(\gamma)^* z = -\sum_{j=1}^{d} \nabla u_j \nabla \psi_{z_j}, \tag{4.5}$$

where $z := (z_1, \ldots, z_d) \in (L^2(\partial\Omega))^d$ and for each $j = 1, \ldots, d$, the vectors u_j and ψ_{z_j} are the unique solutions of (4.1) for $g = g_j$ and $g = z_j$, respectively.

In our numerical simulations we define $\Omega := (0,1) \times (0,1)$ and supply the current vector $(g_1, \ldots, g_d)$ with $d = 8$ independent currents: identifying the faces of Ω with the numbers $m = 0, 1, 2, 3$, we apply the currents

$$g_{2m+k}(x) = \begin{cases} \cos(2k\pi x) & : \text{ on the face } m \\ 0 & : \text{ elsewhere on } \partial\Omega \end{cases}$$

for $k = 1, 2$. The exact solution γ^+ consists of a constant background conductivity 1 and an inclusion $B \subset \Omega$ with conductivity 2:

$$\gamma^+(x) := \begin{cases} 2 & : x \in B \\ 1 & : \text{otherwise.} \end{cases}$$

The set B models two balls with radii equal 0.15 and center at the points $(0.35, 0.35)$ and $(0.65, 0.65)$. The data,

$$y := (\Lambda_{\gamma^+} g_1, \ldots, \Lambda_{\gamma^+} g_d), \tag{4.6}$$

corresponding to the exact solution γ^+ are computed using the finite element method (FEM). The problems (4.1) and (4.4) have been solved by FEM as well, but using a much coarser discretization mesh than the one used to generate the data for avoiding inverse crimes, see Fig. 1.

It is well known that in this specific problem undesirable instability effects may arise from an unfavorable selection of the geometry of the mesh. For avoiding this problem we employ a strategy using a *weight-function* $\omega\colon \Omega \to \mathbb{R}$ to define the weighted-space $L^2_\omega(\Omega) := \{f\colon \Omega \to \mathbb{R} : \int_\Omega |f|^2\omega < \infty\}$. This alteration changes the evaluation of the adjoint operator (4.5) in the discretized setting, see Winkler & Rieder (2015) and Margotti (2015, Subsection 5.1.2) for details. In the mentioned references the authors use the weight-function

$$\omega := \sum_{i=1}^{M} \beta_i \chi_{T_i} \text{ with } \beta_i := \frac{\|F\prime(\gamma_0)\chi_{T_i}\|_{(L^2(\partial\Omega))^d}}{|T_i|},$$

where $|T_i|$ is the area of triangle T_i, and the initial iterate γ_0 is the constant 1 function.

In the notation of Section 1 we have $F : D(F) \subset X := (\mathrm{span}\{\chi_{T_1}, \ldots, \chi_{T_M}\}, \|\cdot\|_{L^2_\omega(\Omega)}) \to (L^2(\partial\Omega))^d =: Y$, where $D(F) = X \cap L^\infty_+(\Omega)$. We define the *relative error* in the kth iterate γ_k as

$$E_k := 100\,\frac{\|\gamma_k - \gamma^+\|_X}{\|\gamma^+\|_X}, \tag{4.7}$$

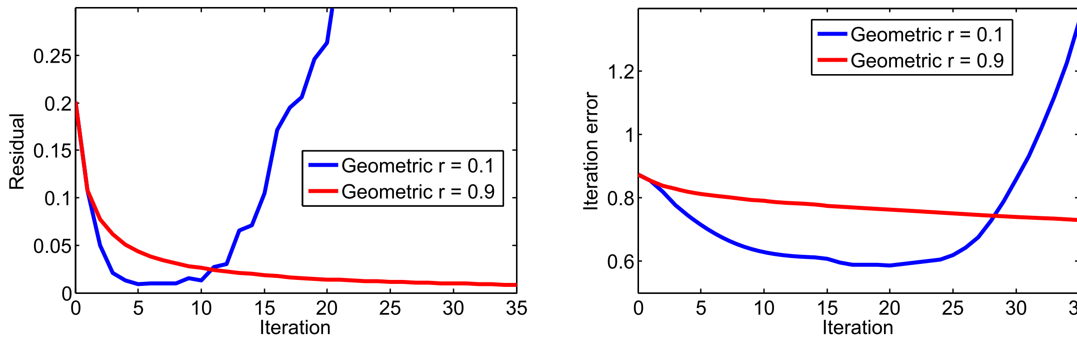

FIG. 2. Geometric choice of the parameters α_k for $r = 0.1$ and $r = 0.9$, with $\alpha_0 = 2$. Noise level $\delta = 0.1\%$, $\eta = 0.4$ and $\tau = 1.3(1+\eta)/(1-\eta)$. Left: residual, right: iteration error.

and use it to compare the quality of the reconstructions. Finally, we corrupt the simulated data y in (4.6) by adding artificially generated random noise with a *relative* noise level $\delta > 0$,

$$y^\delta = y + \delta \,\mathrm{noi}\, \|y\|_Y, \tag{4.8}$$

where $\mathrm{noi} \in Y$ is a uniformly distributed random variable such that $\|\mathrm{noi}\|_Y = 1$.

4.2 *Implementation of the range-relaxed LM method*

Now, we turn to the problem of finding a pair $(\alpha_k > 0, h_k \in X)$ in accordance to Step [3.1] of Algorithm I.

An usual choice for the parameters α_k is of geometric type, i.e., the parameters are defined *a priori* by the rule $\alpha_k = r\alpha_{k-1}$, where $\alpha_0 > 0$ and $0 < r < 1$ (the *decreasing ratio*) are given. This method is usually very efficient if a good guess for the constant r is available. However, big troubles may arise if the decreasing ratio r is chosen either too large or too small. Indeed, on the one hand, if the constant r is too large ($r \approx 1$), then the method becomes slow and the computational costs increase considerably; on the other hand, the LM method becomes unstable in case r is chosen too small ($r \approx 0$), see Fig. 2 above.

Notice that the α_k defined by the geometric choice does not necessarily satisfy the problem in Step [3.1]. We propose a strategy for choosing the decreasing ratio r in each step, so that the resulting parameter α_k (and the corresponding h_k) are in agreement with Step [3.1]. For the actual computation of the ratio r in the current step we use information on the current iteration and past iterations as well. This is described in the sequel.

We adopt the notation

$$H_k(\alpha) = \|y^\delta - F(\gamma_k) - F'(\gamma_k)h_\alpha\|,\ \alpha > 0,$$

where h_α is given by

$$h_\alpha = \left(F'(\gamma_k)^* F'(\gamma_k) + \alpha I\right)^{-1} F'(\gamma_k)^*(y^\delta - F(\gamma_k)). \tag{4.9}$$

According to Step [3.1] in Algorithm I we need to determine $\alpha_k > 0$ such that $H_k(\alpha_k) \in [c_k, d_k]$, where c_k and d_k are defined in (2.6) and (2.7), respectively. For doing that we have employed the *adaptive* *strategy* introduced in Machado *et al.* (2020). This algorithm is based on the geometric method, but allows adaptation of the decreasing ratio using *a posteriori* information. First, we define the constants

$$\widehat{c}_k = p_1 c_k + (1 - p_1) d_k \quad \text{and} \quad \widehat{d}_k = p_2 c_k + (1 - p_2) d_k, \tag{4.10}$$

where $0 < p_1 < p_2 < 1$. Notice that $[\widehat{c}_k, \widehat{d}_k] \subset [c_k, d_k]$.

Choose the initial parameter $\alpha_0 > 0$; compute $h_0 := h_{\alpha_0}$ and γ_1 according to Algorithm I.

Choose the initial decreasing ratio $0 < r_0 < 1$, define $\alpha_1 = r_0 \alpha_0$; compute $h_1 := h_{\alpha_1}$ and γ_2 according to Algorithm I.

For $k \geq 1$ we define $\alpha_{k+1} = r_k \alpha_k$, where

$$r_k = \begin{cases} a_1 r_{k-1}, & \text{if } c_{k-1} \leq H_{k-1}(\alpha_{k-1}) < \widehat{c}_{k-1} \\ a_2 r_{k-1}, & \text{if } \widehat{d}_{k-1} < H_{k-1}(\alpha_{k-1}) \leq d_{k-1} \\ r_{k-1}, & \text{if } H_{k-1}(\alpha_{k-1}) \in [\widehat{c}_{k-1}, \widehat{d}_{k-1}]. \end{cases} \tag{4.11}$$

Here the constants $0 < a_2 < 1 < a_1$ play the role of correction factors and are chosen *a priori*.

The idea of the *adaptive strategy* is to observe the behavior of the function H_k and try to determine how much the parameter α_k should be decreased in the next iteration. For example the number $H_k(\alpha_k)$ lying to the left of the smaller interval $[\widehat{c}_k, \widehat{d}_k]$ means that α_k was too small. We thus multiply the decreasing ratio r_{k-1} by the number $a_1 > 1$, in order to increase it, and, consequently, to decrease the parameter α_k slower than in the previous step, trying to hit $[\widehat{c}_k, \widehat{d}_k]$ in the next iteration. This algorithm is efficient in terms of computational cost: like the geometric choice for α_k, it requires only one minimization of a Tikhonov functional in each iteration. Further, the *adaptive strategy* has the additional advantage of correcting the decreasing ratio if this ratio is either too large or too small.

An attentive reader could object that, in some iterations, the evaluated parameter α_k may lead to a number $H_k(\alpha_k)$ that does not belong to the interval $[c_k, d_k]$ defined in Step [3.1]. This is indeed possible! In this situation we apply the secant method in order to recalculate α_k such that $H_k(\alpha_k) \in [c_k, d_k]$, before starting the next iteration. This is however an expensive task since each step of the secant method demands the additional minimizations of Tikhonov functionals.

It is worth noticing that this situation has been barely observed in our numerical experiments, occurring only in the cases when either the initial decreasing ratio r_0 or the initial guess α_0 are poorly chosen.

4.3 *Numerical realizations*

For the constant τ in (2.3) we use $\tau = 1.3(1 + \eta)/(1 - \eta)$, where $\eta = 0.4$ is the constant in (A2). Moreover, we choose $p = 0.1$ and $\varepsilon = 0.1[\tau(1 - \eta) - (1 + \eta)]/\eta\tau$ in (2.3). The constants in (4.10) are $p_1 = 1/3$ and $p_2 = 2/3$, while the constants in (4.11) are $a_1 = 2$ and $a_2 = 1/2$.

First test (one level of noise): The goal of this test is to investigate the performance of our rrLM method with adaptive strategy (*a posteriori*) for computing the parameters, with respect of different choices of initial decreasing ratio r_0.

As observed in Fig. 2 the performance of the LM method with geometric choice (*a priori*) of parameters is very sensitive to the choice of the (constant) decreasing ratio $r < 1$.

We implement the rrLM method (using *adaptive strategy*) with $r_0 = 0.1$ and $r_0 = 0.9$. In Fig. 3 the results of the rrLM method are compared with the LM method using geometric choice of parameters (see top-left, top-right and bottom-left pictures).

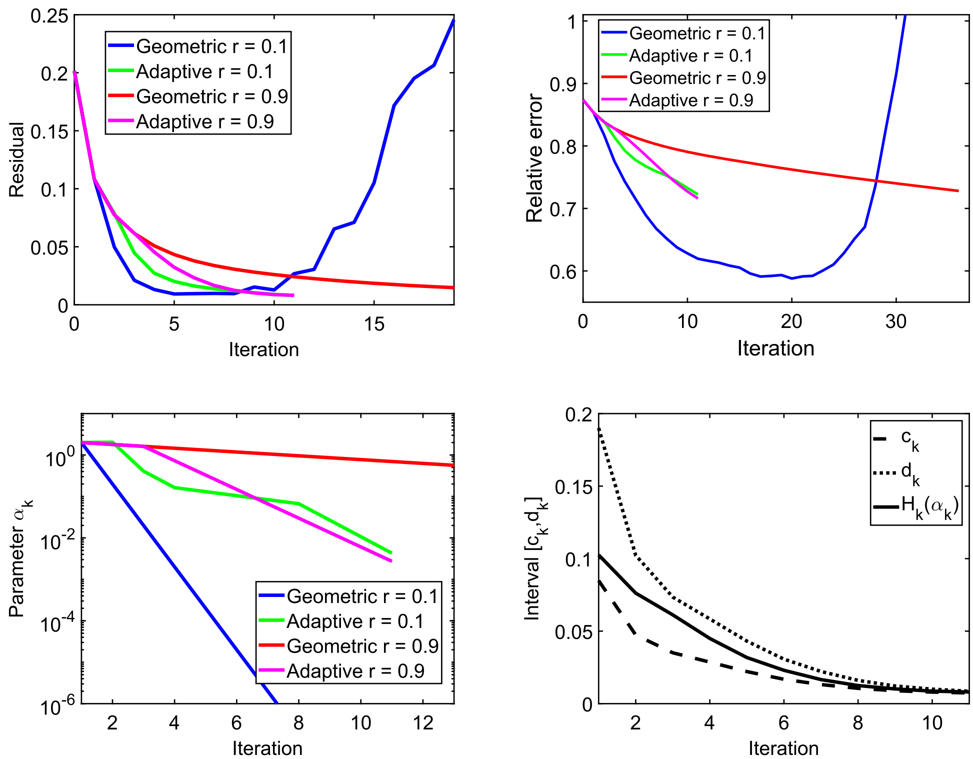

FIG. 3. First test: noisy data, $\delta = 0.1\%$. Top-left: residual. Top-right: relative iteration error. Bottom-left: parameter α_k. Bottom-right: linearized residual $H_k(\alpha_k)$ and the numbers c_k and d_k for the rrLM with $r_0 = 0.9$.

— [GREEN] rrLM with $r_0 = 0.1$, reaches discrepancy with $k^* = 11$ steps;

— [MAGENTA] rrLM with $r_0 = 0.9$, reaches discrepancy with $k^* = 11$ steps;

— [RED] LM with $r = 0.9$, reaches discrepancy with $k^* = 36$ steps;

— [BLUE] LM with $r = 0.1$, does not reach discrepancy.

The noise level is $\delta = 0.1\%$. All methods are started with $\alpha_0 = 2$. The last picture in Fig. 3 (bottom-right) shows the values of the linearized residual $H_k(\alpha_k)$ as well as the intervals $[c_k, d_k]$ (see (2.6) and (2.7)) for the rrLM with $r_0 = 0.9$.

From this first test we draw the following conclusions: • The rrLM method (using *adaptive strategy*) is robust with respect of the choice of the (initial) decreasing ratio. We tested two poor choices of initial decreasing ratios (namely $r_0 = 0.1$ and $r_0 = 0.9$); nevertheless, the performance of the rrLM method in both cases is stable and numerically efficient. For rrLM method the relative error obtained for $r_0 = 0.1$ is comparable to that obtained for $r_0 = 0.9$ (see top-right picture in Fig. 3).

- We also tested the rrLM method (using *adaptive strategy*) and the LM method (using geometric choice of parameters) for $r = r_0 = 0.5$, which seems to be the 'optimal' choice of constant decreasing ratio. In this case both methods performed similarly. Moreover, the performance of

TABLE 1 *Comparison between rrLM and LM methods: computational effort*

	$k^*(N_{k^*})$					
	$(r_0 = 0.9)$		$(r_0 = 0.5)$		$(r_0 = 0.1)$	
$\delta(\%)$	rrLM	LM	rrLM	LM	rrLM	LM
0.8	5(6)	3(3)	4(5)	3(3)	5(8)	3(3)
0.4	8(8)	8(8)	6(6)	4(4)	8(12)	4(4)
0.2	9(9)	18(18)	7(7)	7(7)	8(11)	Fails
0.1	11(11)	35(35)	10(10)	10(10)	11(14)	Fails

the rrLM method (number of iterations and numerical effort) was similar to the ones depicted in Fig. 3 using $r_0 = 0.1$ and $r_0 = 0.9$.

- The rrLM method (using *adaptive strategy*) ‘corrects’ eventual poor choices of the decreasing ratio. If r_0 is too small the *adaptive strategy* increases this ratio during the first iterations (GREEN curve in Fig. 3) preventing instabilities (compare with the LM method using geometric choice of parameters — BLUE curve in Fig. 3). On the other hand, if r_0 is large (close to one), the *adaptive strategy* decreases this ratio during the first iterations (MAGENTA curve in Fig. 3), preventing slow convergence (compare with the LM method using geometric choice of parameters — RED curve in Fig. 3).
- The last picture in Fig. 3 (bottom-right) shows that the linearized residual $H_k(\alpha_k)$, computed using the *adaptive strategy*, satisfies (2.5) in Step [3.1] of Algorithm I. Consequently, this strategy provides a numerical realization of Algorithm I, which is in agreement with the theory devised in this article.

Second test (several levels of noise): The goal of this test is twofold: (first) we validate the regularization property (see Theorem 3.10 and Corollary 3.12) by choosing different levels of noise $\delta > 0$ and observing what happens when the noise level decreases; (second) we compare the numerical effort of the rrLM method (with *adaptive strategy*) with the LM method (with geometric choice of parameters).

In what follows we present a set of experiments with four different levels of noise $\delta > 0$ namely, $\delta = 0.8\%$, $\delta = 0.4\%$, $\delta = 0.2\%$, $\delta = 0.1\%$. In each scenario above we implemented the rrLM method (with *adaptive strategy*) as well as the LM method (with geometric choice of parameters).

For the implementation of the LM method with geometric choice of parameters we use the constant decreasing ratios: $r_0 = 0.9$, $r_0 = 0.5$ and $r_0 = 0.1$. For the implementation of the rrLM method we used the same choices of r_0 as starting value for r together with the *adaptive strategy*. In all implementations $\alpha_0 = 2$ is used. Comparisons of these methods are presented in Tables 1 and 2. Three distinct indicators are used, namely:

- Number of iterations to reach discrepancy $k^* = k^*(\delta)$ (see Step [3.3]);
- Total number of Tikhonov functionals minimized for $k = 0, \ldots k^* - 1$, denoted by N_{k^*};[8]

[8] The numbers k^* and N_{k^*} are always the same in the geometric choice (LM method), but N_{k^*} may be larger than k^* in the *adaptive strategy* (rrLM method).

TABLE 2 *Comparison between rrLM and LM methods: relative iterative error at the final iteration*

	E_{k^*}					
	$(r_0 = 0.9)$		$(r_0 = 0.5)$		$(r_0 = 0.1)$	
$\delta(\%)$	rrLM	LM	rrLM	LM	rrLM	LM
0.8	82.6	82.7	82.8	81.5	82.8	80.9
0.4	79.7	79.7	79.5	79.7	79.6	79.5
0.2	76.5	76.5	76.3	76.6	76.4	Fails
0.1	71.5	72.9	71.6	71.7	72.1	Fails

– Relative iteration error at step $k = k^*$, denoted by E_{k^*} (see (4.7)).[9]

From this second test we draw the following conclusions:

- For both methods k^* increases and E_{k^*} decreases as δ becomes smaller (validating the regularization property).
- For each fixed noise level δ the values of E_{k^*} are similar for both methods.
- If the noise level is small ($\delta = 0.1\%$ and $\delta = 0.2\%$) the rrLM method is more efficient than the LM method for $r_0 = 0.9$. Both methods perform similarly for $r_0 = 0.5$. For $r_0 = 0.1$ the LM method fails to converge, while the rrLM method succeeds in reaching the stopping criterium.
- For higher levels of noise ($\delta = 0.4\%$ and $\delta = 0.8\%$) both methods perform similarly for $r_0 = 0.9$ and $r_0 = 0.5$. For $r_0 = 0.1$ the LM method converges faster than the rrLM method. This is due to the fact that rrLM needs to correct the initial guess for $\alpha_0 = 2$.
- For levels of noise higher than 0.8% the rrLM stops after two or less iterations (for different choices of r_0). Consequently, these experiments do not give relevant information about the performance of our method.
- For the rrLM method the values of k^* and N_{k^*} are identical in most of the scenarios of Table 1, i.e., only one Tikhonov functional is minimized in each step (this is the same numerical cost for one step of the LM method with geometric choice of parameters).

The last conclusion validates the *adaptive strategy* for computing the parameters α_k as an efficient alternative for the numerical implementation of Step [3.1] in Algorithm I.

5. Final remarks and conclusions

In this article we address the LM method for solving nonlinear ill-posed problems, and propose a novel range-relaxed criteria for choosing the Lagrange multipliers, namely: the new iterate is obtained as the projection of the current one onto a level set of the linearized residual function; this level belongs to an interval (or *range*), which is defined by the current nonlinear residual and by the noise level (see Step [3.1] of Algorithm I).

The main contributions in this article are:

[9] It is worth noticing that the initial iteration error is $E_0 = 87.39\%$ in all four scenarios above.

- We derive a complete convergence analysis: convergence (Theorem 3.7), stability (Theorem 3.9), semi-convergence (Theorem 3.10). We also prove monotonicity of iteration error (Theorem 2.6) and geometric decay of residual (Proposition 3.2). Moreover, we prove convergence to minimal-norm solution under additional null-space condition (3.12), in both exact and noisy data cases.
- We give a novel proof for the stability result, which uses nonstandard arguments. In the classical stability proof, since each Lagrange multiplier is uniquely defined by an (implicit) equation, the set of successors (Definition 3.8) of each x_k^δ is singleton. However, due to our range-relaxed criteria (2.5), each set of successors may contain infinitely many elements; consequently, the subsequences $\{\delta_{j_m}\}_{m\in\mathbb{N}}$ obtained in Theorem 3.9 do depend on the iteration index k.
- We devise a numerical algorithm, based on the *adaptive strategy* (see Subsection 4.2), for implementing the range-relaxed criteria proposed in this article. Its main features are:
  - Efficiency in terms of computational cost: like the LM with geometric (*a priori*) choice of parameters it (almost always) requires only one minimization of a Tikhonov functional in each iteration.
  - Correction of the decreasing ratio if this ratio is either too large or too small.
  - The computed pairs (α_k, h_k) satisfy (2.5) for all $k > 0$, i.e., this algorithm provides a numerical realization of Algorithm I.

Funding

Brazilian National Research Council CNPq (grant 311087/2017–5 to A.L.; grants 306247/2015-1, 430868/2018-9 to B.F.S). The Alexander von Humboldt Foundation AvH to A.L. The Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro FAPERJ, grant Cientistas de Nosso Estado E-26/203.318/2017 to B.F.S.

References

Bakushinsky, A. B. & Kokurin, M. Y. (2004) *Iterative Methods for Approximate Solution of Inverse Problems. Mathematics and Its Applications, vol. 577*. Dordrecht: Springer.

Baumeister, J., Kaltenbacher, B. & Leitão, A. (2010) On Levenberg–Marquardt–Kaczmarz iterative methods for solving systems of nonlinear ill-posed equations. *Inverse Probl. Imaging*, **4**, 335–350.

Boiger, R., Leitão, A. & Svaiter, B. F. (2020) Range-relaxed criteria for choosing the Lagrange multipliers in nonstationary iterated Tikhonov method. *IMA J. Numer. Anal.*, **40**, 606–627.

Borcea, L. (2002) Electrical impedance tomography. *Inverse Problems*, **18**, R99–R136.

Calderón, A.-P. (1980) On an inverse boundary value problem. *Seminar on Numerical Analysis and Its Applications to Continuum Physics (Rio de Janeiro, 1980)*. Rio de Janeiro: Soc. Brasil. Mat., pp. 65–73. MR 590275 (81k:35160).

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

Engl, H. W., Hanke, M. & Neubauer, A. (1996) *Regularization of Inverse Problems*. Dordrecht: Kluwer Academic Publishers.

Groetsch, C. W. (1984) *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*. *Research Notes in Mathematics, vol. 105*. Boston, MA: Pitman (Advanced Publishing Program).

Hanke, M. (1997) A regularizing Levenberg–Marquardt scheme, with applications to inverse groundwater filtration problems. *Inverse Problems*, **13**, 79–95.

Hanke, M., Neubauer, A. & Scherzer, O. (1995) A convergence analysis of Landweber iteration for nonlinear ill-posed problems. *Numer. Math.*, **72**, 21–37.

Kaltenbacher, B., Neubauer, A. & Scherzer, O. (2008) *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*. *Radon Series on Computational and Applied Mathematics, vol. 6*. KG, Berlin: Walter de Gruyter GmbH & Co.

Landweber, L. (1951) An iteration formula for Fredholm integral equations of the first kind. *Amer. J. Math.*, **73**, 615–624.

Lechleiter, A. & Rieder, A. (2008) Newton regularizations for impedance tomography: convergence by local injectivity. *Inverse Problems*, **24**, 065009, 18.

Levenberg, K. (1944) A method for the solution of certain non-linear problems in least squares. *Quart. Appl. Math.*, **2**, 164–168.

Machado, M., Margotti, F. & Leitão, A. (2020) On the choice of Lagrange multipliers in the iterated Tikhonov method for linear ill-posed equations in Banach spaces. *Inverse Probl. Sci. Eng.*, **28**, 796–826.

Margotti, F. (2015) *On inexact Newton methods for inverse problems in Banach spaces*. Ph.D. Thesis. Karlsruher Institut für Technologie, Karlsruhe.

Margotti, F. & Rieder, A. (2015) An inexact Newton regularization in Banach spaces based on the nonstationary iterated Tikhonov method. *J. Inverse Ill-Posed Probl.*, **23**, 373–392.

Marquardt, D. W. (1963) An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Indust. Appl. Math.*, **11**, 431–441.

Morozov, V. A. (1993) *Regularization Methods for Ill-Posed Problems*. Boca Raton: CRC Press.

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

Scherzer, O. (1993) Convergence rates of iterated Tikhonov regularized solutions of nonlinear ill-posed problems. *Numer. Math.*, **66**, 259–279.

Seidman, T. I. & Vogel, C. R. (1989) Well posedness and convergence of some regularisation methods for non-linear ill posed problems. *Inverse Problems*, **5**, 227–238.

Tikhonov, A. N. (1963) Regularization of incorrectly posed problems. *Soviet Math. Dokl.*, **4**, 1624–1627.

Tikhonov, A. N. & Arsenin, V. Y. (1977) *Solutions of Ill-Posed Problems*. Washington, D.C.: John Wiley & Sons. Translation editor: Fritz John.

Winkler, R. & Rieder, A. (2015) Model-aware Newton-type inversion scheme for electrical impedance tomography. *Inverse Problems*, **31**, 045009.