MIXED GRADIENT-TIKHONOV METHODS FOR SOLVING NON-LINEAR ILL-POSED PROBLEMS IN BANACH SPACES

FÁBIO MARGOTTI

Abstract. Tikhonov regularization is a very useful and widely used method for finding stable solutions of ill-posed problems. A good choice of the penalization functional as well as a careful selection of the topologies of the involved spaces is fundamental to the quality of the reconstructions. These choices can be combined with some *a priori* information about the solution in order to preserve desired characteristics like sparsity constraints for example.

To prove convergence and stability properties of this method, one usually has to assume that a minimizer of the Tikhonov functional is known. In practical situations however, the exact computation of a minimizer is very difficult and even finding an approximation can be a very challenging and expensive task if the involved spaces have poor convexity or smoothness properties.

In this paper we propose a method to attenuate this gap between theory and practice, applying a gradient-like method to a Tikhonov functional in order to approximate a minimizer. Using only available information, we explicitly calculate a maximal step-size which ensures a monotonically decreasing error. The resulting algorithm performs only finitely many steps and terminates using the discrepancy principle. In particular the knowledge of a minimizer or even its existence does not need to be assumed.

Under standard assumptions, we prove convergence and stability results in relatively general Banach spaces, and subsequently, test its performance numerically, reconstructing conductivities with sparsely located inclusions and different kinds of noise in the 2D Electrical Impedance Tomography.

1. Introduction

We intent to find a solution of the non-linear and ill-posed inverse problem

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

where F operates between *Banach spaces* X and Y, that is, $F\colon \mathrm{D}(F) \subset X \to Y$, with $\mathrm{D}(F)$ denoting the domain of definition of F. We assume to have full knowledge of this operator, but only an approximation y^δ for y satisfying

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

is available. The *noise level* $\delta > 0$ is assumed to be known as well.

In the last few years, a large variety of methods such as gradient-based and Newton-like methods have been employed to solve the above inverse problem in relative general Banach spaces [21, 13, 15, 19]. Another popular and widely used method to stably approximate a solution of (1) is Tikhonov regularization. This method consists in

Date: January 8, 2016.

The author acknowledges support by Coordenação de Aperfeiçoamento de Pessoal do Nível Superior (CAPES) and Instituto Nacional de Matemática Pura e Aplicada (IMPA).

finding (in case it exists) a minimizer x_α of the *Tikhonov functional*

$$T_\alpha(x) := \frac{1}{r}\left\|F(x) - y^\delta\right\|^r + \alpha\Phi(x), \quad \alpha > 0, \tag{3}$$

where $r > 1$ and $\Phi\colon X \to \mathbb{R}_0^+$ is a subdifferentiable functional. Choosing the regularization parameter α depending on the noise level δ, the challenge is now proving the *regularization property*:

$$x_{\alpha(\delta)} \to x^+ \text{ as } \delta \to 0,$$

where $x^+ \in X$ is a solution of (1) . A vast literature concerning Tikhonov regularization methods is available [23, 2, 10, 14]. Probably the biggest trouble in this approach is the necessity of solving an optimization problem. Finding a minimizer of (3) exactly is almost impossible in practical situations. Even finding only an approximate minimizer for (3) is not a trivial task and it becomes a very difficult and expensive problem (in terms of computational effort) if the involved Banach spaces have poor smoothness or convexity properties, see e.g. [3, 12].

An alternative for stably solving (1) , but exempted from the obligation of solving an optimization problem, can be obtained employing a *gradient method.* If a functional $T\colon X \to \mathbb{R}$ is at least Gâteaux-differentiable (in short G-differentiable) at x and X is a Hilbert space, then the opposite of its G-derivative, $-\nabla T(x)$, is a descent direction for T from x , which means that $T(x - \lambda\nabla T(x)) < T(x)$ provided $\lambda > 0$ small enough. Motivated by this result, the gradient methods in *Hilbert spaces* are defined by the iteration:

$$x_{n+1} = x_n - \lambda_n \nabla T(x_n),$$

where $\lambda_n > 0$ and T is the residual-functional

$$T(x) := \frac{1}{2}\left\|F(x) - y^\delta\right\|^2. \tag{4}$$

Consequently, $\left\|F(x_{n+1}) - y^\delta\right\| < \left\|F(x_n) - y^\delta\right\|$ for λ_n small enough. The Fréchet-derivative (in short F-derivative) of T exists whenever F is F-differentiable. In this case, the chain-rule yields

$$\nabla T(x_n) = F'(x_n)^*\left(F(x_n) - y^\delta\right).$$

It is well-known that for ill-posed problems, the iteration needs to be terminated in time in order to avoid error amplification. A usual stop-criterion is the so-called *discrepancy principle*: stop in the first iteration $n = N(\delta)$ satisfying

$$\left\|y^\delta - F\left(x_{N(\delta)}\right)\right\| \leq \tau\delta, \tag{5}$$

with $\tau > 1$ being a pre-defined constant. The regularization property now reads:

$$x_{N(\delta)} \to x^+ \text{ as } \delta \to 0. \tag{6}$$

The choice of the *step-size* λ_n is a crucial factor for the success of this kind of method. Different choices of λ_n result in different gradient methods. Some classical examples are: *Landweber method* (constant step-size), *Steepest Descent method* (step-size minimizes the residual-functional (4)) and the *Minimal Error method* (where the step-size minimizes the error-functional $E(x) := \|x - x^+\|^2$), see [20] for more details.

Mixed methods are defined applying a gradient-like method to the Tikhonov functional (3) instead of applying it in the residual-functional (4) . The resulting iteration in Hilbert spaces is given by

$$x_{n+1} = x_n - \lambda_n \nabla T_\alpha(x_n),$$

with $\lambda_n > 0$. Of course this iteration is well-defined only if T_α is at least G-differentiable. Although the sequence generated by this iteration in some sense approximates a minimizer of T_α ($T_\alpha(x_{n+1}) < T_\alpha(x_n)$ for λ_n small enough), we *do not* intend to find this minimizer exactly. Instead, the iteration is terminated using the discrepancy principle (5). The regularization property reads exactly as in (6) in this situation.

In this paper we propose a mixed method in Banach spaces, combining a gradient-like method with a Tikhonov functional. Using only available information, we explicitly compute a maximal step-size $\lambda_{\max} = \lambda_{\max}(n, \alpha)$ such that $\lambda_n \in (0, \lambda_{\max}]$ implies the new method has a monotonically decreasing error. In order to have a non-empty interval for selecting λ_n, we choose either a small constant (in n) regularization parameter $\alpha = \alpha(\delta)$ or allow it to vary during the iteration, i.e., $\alpha = \alpha_n(\delta)$. The first variation is similar to Tikhonov-like methods, because the last iterate $x_{N(\delta)}$ in some sense approximates a minimizer $x_{\alpha(\delta)}$ of (3), however, it is not necessary to find or even assume the existence of a minimizer of (3). The second strategy in turn, results in a method similar to gradient-like methods, but with the significant advantage of having an extra stability term, which is automatically adjusted, gathering information on the current iterate, and accordingly conferring more stability to the whole iteration process. In both cases, an appropriate penalization functional can be utilized for incorporating *a priori* information of the exact solution and preserving desired features in the reconstructions. The analysis includes as particular cases, non-linear versions of the gradient methods: Landweber, Steepest Descent and Minimal Error method.

This work is structured as following: the next section collects some necessary results from [8, 7, 22] concerning the geometry of Banach spaces. In Section 3, we present the mixed method and in Section 4, a complete convergence analysis is carried out. Finally, we test the efficiency of our method in the non-linear and highly ill-posed inverse problem of 2D Electrical Impedance Tomography in Section 5.

2. A little about geometry of Banach spaces

In this section we collect needed facts about the geometry of Banach spaces. For proofs and more details we refer to the book of Cioranescu [8] and to [7, 22]. The notation $\lesssim$, which we will use in the following sections will be used with following meaning: $a(x) \lesssim b(x)$ if and only if there exists a positive constant C independent of x such that $a(x) \leq C\, b(x)$ for all x. Further, we assume w.l.o.g. that X is a *real* Banach space, see [8, Remark 3.2].

The *modulus of smoothness* of the Banach space X is defined 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 $\lim_{\tau \to 0^+} \frac{\rho_X(\tau)}{\tau} = 0$ and *uniformly convex* if $\delta_X(\epsilon) > 0$ for all $0 < \epsilon \leq 2$. Let $s > 1$ be fixed. We call X $s-$*smooth* if $\rho_X(\tau) \lesssim \tau^s$ for all $\tau \geq 0$ and we call it $s-$*convex* if $\delta_X(\epsilon) \gtrsim \epsilon^s$ for all $0 \leq \epsilon \leq 2$. X uniformly smooth or uniformly convex implies X reflexive. Further, X is $s-$smooth (resp. uniformly smooth) iff X^* is s^*-convex (resp. uniformly convex) and vice-versa, with $\frac{1}{s} + \frac{1}{s^*} = 1$.

Fix $p > 1$ and define the norm-functional as

$$f(x) := \frac{1}{p}\|x\|^p, \ x \in X. \tag{7}$$

The Banach space X is called strictly convex (resp. locally uniformly convex) iff f is strictly convex (resp. locally uniformly strictly convex[1]) and it is called smooth (resp. locally uniformly smooth) iff f is G-differentiable (resp. F-differentiable) in X. The following implications hold:

$$s-\text{smoothness} \Rightarrow \text{uniform smoothness} \Rightarrow \text{locally uniform smoothness} \Rightarrow \text{smoothness},$$
$$s-\text{convexity} \Rightarrow \text{uniform convexity} \Rightarrow \text{locally uniform convexity} \Rightarrow \text{strict convexity},$$

and none reciprocal is true. Hilbert spaces are $2-$smooth and $2-$convex.

Important examples are the Lebesgue spaces $L^p(\Omega)$, the Sobolev spaces $W^{n,p}(\Omega)$ and the space of $p-$summable sequences $\ell^p(\mathbb{R})$. They are[2] $p \wedge 2-$smooth and $p \vee 2-$convex for $1 < p < \infty$. For $p = 1$ or $p = \infty$ they are neither smooth nor strictly convex.

A continuous and strictly increasing function $\varphi\colon \mathbb{R}_0^+ \to \mathbb{R}_0^+$ such that $\varphi(0) = 0$ and $\lim_{t\to\infty} \varphi(t) = \infty$ is called a *gauge function.* The *duality mapping* associated with the gauge function φ is the point-to-set mapping $J_\varphi\colon X \to 2^{X^*}$ defined by

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

where $\langle\cdot,\cdot\rangle\colon X^* \times X \to \mathbb{R}$ is the duality pairing. The duality mapping associated with the gauge function $t \longmapsto t^{p-1}$ 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 mapping J_2 is called the *normalized* duality mapping and as a consequence of Riesz Representation Theorem, it is the identity operator in Hilbert spaces. Furthermore, J_2 is linear only in Hilbert spaces. A *selection* $j_\varphi\colon X \to X^*$ of the duality mapping J_φ is a function which satisfies $j_\varphi(x) \in J_\varphi(x)$ for all $x \in X$. From definition, it follows immediately that each selection j_p of the duality mapping J_p satisfies the inner-product-like properties

$$\langle j_p(x), y\rangle \le \|x\|^{p-1}\|y\| \qquad \text{and} \qquad \langle j_p(x), x\rangle = \|x\|^p.$$

The very important Asplund's Theorem guarantees that if φ is a gauge function, then the functional

$$\psi(x) := \int_0^{\|x\|} \varphi(s)\,ds \tag{8}$$

is convex and its subdifferential coincides with the duality mapping:

$$J_\varphi(x) = \partial\psi(x). \tag{9}$$

In particular, $J_p(x) = \partial f(x)$, with f from (7). Thus, X smooth implies that $J_p\colon X \to X^*$ is single-valued and in this case, $J_p = \nabla f(\cdot)$ is the G-derivative of f. Further, X locally uniformly smooth implies the F-differentiability of f and consequently J_p is continuous. Analogous results hold true if J_p is replaced by J_φ. If X is additionally locally uniformly convex, then X is reflexive and J_φ is invertible, with a continuous inverse given by $J_\varphi^{-1} = J_{\varphi^{-1}}^*$. In particular,

$$J_p^{-1} = J_{p^*}^*\colon X^* \to X^{**} \cong X \tag{10}$$

[1]See the definition of a locally uniformly strictly convex function in [8, Def. 2.10 *ii*), Ch. II].

[2]$a \wedge b := \min\{a, b\}$ and $a \vee b := \max\{a, b\}$.

holds true in a locally uniformly smooth and locally uniformly convex Banach space.

Suppose now that X is a smooth Banach space and φ is a fixed gauge function. Then J_φ is single-valued and we define the *Bregman distance* $\Delta_\varphi \colon X \times X \to [0,\infty)$ as

$$\Delta_\varphi(x,y) := \psi(x) - \psi(y) - \langle J_\varphi(y), x-y\rangle,$$

where ψ is the function from (8) . From Asplund's Theorem (9) , it follows that $\Delta_\varphi(x,y) \geq 0$. Of course $x = y$ implies $\Delta_\varphi(x,y) = 0$. Using the particular gauge function $\varphi(t) = t^{p-1}$ the Bregman distance becomes

$$\begin{aligned}\Delta_p(x,y) &= \frac{1}{p}\|x\|^p - \frac{1}{p}\|y\|^p - \langle J_p(y), x-y\rangle \\ &= \frac{1}{p}\|x\|^p - \langle J_p(y), x\rangle + \frac{1}{p^*}\|J_p(y)\|^{p^*}.\end{aligned}$$

This equality mimics the *polarization identity*

$$\frac{1}{2}\|x-y\|^2 = \frac{1}{2}\|x\|^2 - \langle y, x\rangle + \frac{1}{2}\|y\|^2, \tag{11}$$

which holds true in Hilbert spaces, and where also holds $\Delta_2(x,y) = \frac{1}{2}\|x-y\|^2$.

If X is strictly convex, the function f in (7) is strictly convex, which in turn implies the strict convexity of Δ_p on its first variable and consequently $\Delta_p(x,y) = 0$ iff $x = y$. The Bregman distance is not a metric, because it does not necessarily satisfy the symmetry for instance. But if X is reflexive, the symmetry-like property $\Delta_p(x,y) = \Delta_{p^*}(J_p(y), J_p(x))$ can be shown. A straightforward calculation shows the important *Three Points Identity*:

$$\Delta_p(z,y) - \Delta_p(z,x) = \Delta_p(x,y) + \langle J_p(y) - J_p(x), x-z\rangle, \tag{12}$$

for all $x, y, z \in X$. Moreover,

$$\Delta_p(x,y) \geq \frac{1}{p}\|x\|^p + \frac{1}{p^*}\|y\|^p - \|y\|^{p-1}\|x\|.$$

Now, if $(x_n)_{n\in\mathbb{N}} \subset X$ is an arbitrary sequence and $x \in X$ is a fixed vector, then the inequality $\Delta_p(x,x_n) \leq \rho$ implies

$$\|x_n\|^{p-1}\left(\frac{1}{p^*}\|x_n\| - \|x\|\right) \leq \rho.$$

Considering now the cases $\frac{1}{p^*}\|x_n\| - \|x\| \leq \frac{1}{2p^*}\|x_n\|$ and $\frac{1}{p^*}\|x_n\| - \|x\| > \frac{1}{2p^*}\|x_n\|$, one can prove the implication

$$\Delta_p(x,x_n) \leq \rho \implies \|x_n\| \leq 2p^*\left(\|x\| \vee \rho^{1/p}\right). \tag{13}$$

Therefore, $(x_n)_{n\in\mathbb{N}}$ is bounded whenever $\Delta_p(x,x_n)$ is bounded. A similar result can be proven if $\Delta_p(x_n,x) \leq \rho$.

If the duality mapping is single-valued and continuous (this is the case in a locally uniformly smooth Banach space) then the continuity is handed down to both arguments of the Bregman distance Δ_p.

The important Xu-Roach Theorems [25] imply that in a $s-$convex Banach space there exists, for each $p \leq s$, a positive constant $K_{p,s}$ such that

$$K_{p,s}\left(\|y\| \vee \|x-y\|\right)^{p-s}\|x-y\|^s \leq \Delta_p(x,y)$$

for all $x, y \in X$. In particular, if the sequence $(x_n)_{n\in\mathbb{N}} \subset X$ is bounded (or $p = s$), then there exists a constant $C > 0$ such that

$$\|x_n - x_m\|^s \leq C\Delta_p(x_n, x_m). \tag{14}$$

Similarly, in a $s-$smooth Banach space it holds, for each $p \geq s$,

$$\Delta_p(x, y) \leq C_{p,s}\left(\|y\|^{p-s}\|x - y\|^s \vee \|x - y\|^p\right) \tag{15}$$

for all $x, y \in X$, where $C_{p,s} > 0$ is a constant.

3. Mixed Gradient-Tikhonov methods

To properly introduce the ideas, some requirements on the inverse problem $F(x) = y$ are necessary:

Assumption 1. (*a*) *There exists a solution* $x^+ \in X$ *of equation* (1) *and a number* $\rho > 0$ *such that*

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

(*b*) *The function* F *is continuously F-differentiable in* $B_\rho(x^+, \Delta_p)$ *and its derivative satisfies*

$$\|F'(v)\| \leq M \text{ for all } v \in B_\rho\left(x^+, \Delta_p\right),$$

where $M > 0$ *is a constant.*

(*c*) *(Tangential Cone Condition (TCC)): There exists a constant* $0 \leq \eta < 1$ *such that*

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

for all $v, w \in B_\rho(x^+, \Delta_p)$.

Assuming Y is a locally uniformly smooth Banach space, the norm-functional $f(y) := \frac{1}{r}\|y\|_Y^r$ is F-differentiable and its F-derivative coincides with the duality mapping: $\nabla f(y) := J_r(y)$. Further, $J_r \colon Y \to Y^*$ is single-valued and continuous and since F is F-differentiable too, an application of chain rule yields

$$\nabla\left(\frac{1}{r}\left\|F(x) - y^\delta\right\|^r\right) = F'(x)^* J_r\left(F(x) - y^\delta\right).$$

However, in a general Banach space Y, the duality mapping J_r is not necessarily single-valued and the above inequality does not hold automatically true. For this reason, we define the point-to-set mapping $dT_\alpha \colon B_\rho(x^+, \Delta_p) \to 2^{X^*}$ as

$$dT_\alpha(x) := \left\{w^* + \alpha z^* : w^* \in F'(x)^* J_r\left(F(x) - y^\delta\right) \text{ and } z^* \in \partial\Phi(x)\right\} \subset X^*, \tag{16}$$

where $\Phi \colon X \to \mathbb{R}_0^+$ is a subdifferentiable functional. Observe that, if Y is locally uniformly smooth and Φ is G-differentiable (resp. F-differentiable) at x, then the Tikhonov functional T_α in (3) is G-differentiable (resp. F-differentiable) at x as well, and the set $dT_\alpha(x)$ is singleton. In this case, the unique element in $dT_\alpha(x)$ is the G-derivative (resp. F-derivative) of T_α at x, i.e., $dT_\alpha(x) = \{\nabla T_\alpha(x)\}$.

The plan now is applying a gradient-like method to T_α in order to approximate a minimizer (in case it exists): assume X is locally uniformly smooth and $s-$convex with $p \leq s$ and define the iteration[3]

$$J_p(x_{n+1}) := J_p(x_n) - \lambda_n\Psi_n, \qquad \Psi_n \in dT_\alpha(x_n), \tag{17}$$

[3] J_p is single-valued and invertible in this situation, see (10).

with $\lambda_n > 0$. Assuming now that x_n is well-defined and belongs to $B_\rho(x^+, \Delta_p)$ ($\subset \mathrm{D}(F)$) and adopting the notations

$$A_n := F(x_n),\ b_n^\delta := y^\delta - F(x_n) \text{ and } e_n := x^+ - x_n,$$

it follows from Three Points Identity (12), with $f(\lambda_n) := \Delta_p(x^+, x_{n+1}) - \Delta_p(x^+, x_n)$, that

$$\begin{aligned} f(\lambda_n) &= \Delta_p(x_n, x_{n+1}) + \left\langle J_p(x_{n+1}) - J_p(x_n), x_n - x^+ \right\rangle \\ &= \Delta_{p^*}(J_p(x_{n+1}), J_p(x_n)) - \lambda_n \left\langle A_n^* j_r\left(b_n^\delta\right), e_n \right\rangle + \lambda_n \alpha \left\langle z_n^*, e_n \right\rangle, \end{aligned} \tag{18}$$

with $z_n^* \in \partial\Phi(x_n)$. Our aspiration now is calculating a positive number $\lambda_{\max} = \lambda_{\max}(\alpha, n)$ such that the inequalities $0 < \lambda_n \leq \lambda_{\max}$ will imply the rightmost expression in (18) is negative, resulting in the monotonicity of the error:

$$\Delta_p\left(x^+, x_{n+1}\right) < \Delta_p\left(x^+, x_n\right)$$

and consequently $x_{n+1} \in B_\rho(x^+, \Delta_p) \subset \mathrm{D}(F)$.

To estimate the first term in the rightmost expression in (18), we make use of the $s-$convexity of X, which is equivalent to the s^*-smoothness of X^*. As $x_n \in B_\rho(x^+, \Delta_p)$, it follows from (13) that $\|x_n\| \leq C_{\rho,x^+}$, with

$$C_{\rho,x^+} := 2p^*\left(\left\|x^+\right\| \vee \rho^{1/p}\right). \tag{19}$$

Since $p^* \geq s^*$, inequality (15) together with definition (17) implies that

$$\begin{aligned} \Delta_p^*(J_p(x_{n+1}), J_p(x_n)) &\leq C_{p^*,s^*}\left(\|J_p(x_n)\|^{p^*-s^*}\|J_p(x_{n+1}) - J_p(x_n)\|^{s^*} \vee \|J_p(x_{n+1}) - J_p(x_n)\|^{p^*}\right) \\ &\leq C_{p^*,s^*}\left(C_{\rho,x^+}^{p-s^*(p-1)}\lambda_n^{s^*}\|\Psi_n\|^{s^*} \vee \lambda_n^{p^*}\|\Psi_n\|^{p^*}\right). \end{aligned}$$

The discrepancy principle (5) ensures that $\delta < \frac{1}{\tau}\left\|b_n^\delta\right\|$ for $n = 0, ..., N(\delta) - 1$. Then from TCC and (2) we obtain

$$\begin{aligned} \left\|b_n^\delta - A_n e_n\right\| &= \left\|y^\delta - F(x_n) - F'(x_n)\left(x^+ - x_n\right)\right\| \\ &\leq \delta + \left\|F\left(x^+\right) - F(x_n) - F'(x_n)\left(x^+ - x_n\right)\right\| \\ &\leq \delta + \eta\left\|F\left(x^+\right) - F(x_n)\right\| \\ &\leq \delta + \eta\left(\left\|y^\delta - F(x_n)\right\| + \delta\right) \leq \left(\eta + \frac{1+\eta}{\tau}\right)\left\|b_n^\delta\right\|. \end{aligned}$$

Therefore, the second term in the rightmost expression in (18) can be estimated by

$$\begin{aligned} -\lambda_n \left\langle A_n^* j_r\left(b_n^\delta\right), e_n \right\rangle &= \lambda_n\left[\left\langle j_r\left(b_n^\delta\right), b_n^\delta - A_n e_n\right\rangle - \left\langle j_r\left(b_n^\delta\right), b_n^\delta\right\rangle\right] \\ &\leq \lambda_n\left(\left\|b_n^\delta\right\|^{r-1}\left\|b_n^\delta - A_n e_n\right\| - \left\|b_n^\delta\right\|^r\right) \leq -\lambda_n C_0\left\|b_n^\delta\right\|^r, \end{aligned}$$

with

$$0 < C_0 \leq 1 - \eta - \frac{1+\eta}{\tau}, \tag{20}$$

which is well-defined provided $\tau > (1+\eta)/(1-\eta)$. Finally, if a bound for $\Phi(x^+)$ is known, i.e., if a number $C_1 > 0$ satisfying $\Phi(x^+) \leq C_1$ is available, it follows from definition of subdifferential that the third term in the rightmost expression in (18) can be bounded by

$$\lambda_n \alpha \left\langle z_n^*, e_n \right\rangle \leq \lambda_n \alpha\left(\Phi\left(x^+\right) - \Phi(x_n)\right) \leq \lambda_n \alpha C_1. \tag{21}$$

Putting all together, we arrive at $f(\lambda_n) \leq g(\lambda_n)$, with

$$g(\lambda_n) := \lambda_n \left[C_{p^*,s^*} \left(C_{\rho,x^+}^{p-s^*(p-1)} \lambda_n^{s^*-1} \|\Psi_n\|^{s^*} \vee \lambda_n^{p^*-1} \|\Psi_n\|^{p^*} \right) - C_0 \|b_n\|^r + \alpha C_1 \right].$$

Observe that $g(\lambda_n) < 0$ if $\lambda_n \in (0, \lambda_{\max}]$, where the calculable bound $\lambda_{\max}$ is given by[4]

$$\lambda_{\max} := C_2 \left\{ \frac{\left[C_0 \left\|b_n^\delta\right\|^r - \alpha C_1 \right]^{s-1}}{\|\Psi_n\|^s} \wedge \frac{\left[C_0 \left\|b_n^\delta\right\|^r - \alpha C_1 \right]^{p-1}}{\|\Psi_n\|^p} \right\}, \tag{22}$$

with $0 < C_2 < \left(C_{p^*,s^*}^{1-s} C_{\rho,x^+}^{p-s} \wedge C_{p^*,s^*}^{1-p} \right)$.

To have a positive $\lambda_{\max}$, we arrive with two possibilities:

- **Variation 1:** $\alpha = \alpha(\delta)$ is fixed during the whole iteration. In this case

$$\lambda_{\max} > 0 \Leftrightarrow \left\|b_n^\delta\right\|^r > \frac{C_1}{C_0} \alpha(\delta).$$

  Observe that if

$$0 \leq \alpha(\delta) \leq C_3 \delta^r, \tag{23}$$

  with $0 < C_3 < \frac{C_0}{C_1} \tau^r$, then

$$\frac{C_1}{C_0} \alpha(\delta) < (\tau\delta)^r < \left\|b_n^\delta\right\|^r$$

  as long as the discrepancy principle (5) is not satisfied. Therefore $\lambda_{\max}$ remains positive and it is possible to choose $\lambda_n \in (0, \lambda_{\max}]$.
- **Variation 2:** $\alpha = \alpha_n(\delta)$ is possibly updated during the iteration. In this situation,

$$\lambda_{\max} > 0 \Leftrightarrow 0 \leq \alpha_n(\delta) < \frac{C_0}{C_1} \left\|b_n^\delta\right\|^r.$$

  Moreover, choosing

$$0 \leq \alpha_n \leq C_4 \|b_n\|^r \tag{24}$$

  with $0 < C_4 < C_0/C_1$, it follows that

$$C_0 \left\|b_n^\delta\right\|^r - \alpha_n C_1 \geq C_0 \left\|b_n^\delta\right\|^r - C_1 C_4 \left\|b_n^\delta\right\|^r = (C_0 - C_1 C_4) \left\|b_n^\delta\right\|^r > 0. \tag{25}$$

  See Algorithm 1 for an implementation in pseudocode.

Due to the discrepancy principle (5), the strategy (23) for choosing the regularization parameter α is a particular case of (24). Hence, it suffices to carry out the convergence analysis only for the second variation.

[4]A simple calculation could replace $\lambda_{\max}$ with the larger bound

$$\lambda_{\max,2} = \frac{\left[C_0 \left\|b_n^\delta\right\|^r - \alpha(C_1 - \Phi(x_n)) \right]^{s-1}}{C_{p^*,s^*}^{s-1} C_{\rho,x^+}^{s-p} \|\Psi_n\|^s} \wedge \frac{\left[C_0 \left\|b_n^\delta\right\|^r - \alpha(C_1 - \Phi(x_n)) \right]^{p-1}}{C_{p^*,s^*}^{p-1} \|\Psi_n\|^p},$$

but since this expression is more complicated, we keep the first one for the ease of the presentation.

Algorithm 1 Mixed Method

Input: F; Φ; y^δ; δ; η; C_{p^*,s^*}; C_1; C_{ρ,x^+}; x_0;

Output: x_N with $\|y^\delta - F(x_N)\| \leq \tau\delta$;

Choose $\tau > \frac{1+\eta}{1-\eta}$; $0 < C_0 \leq 1 - \eta - \frac{1+\eta}{\tau}$; $0 < C_2 < \left(C_{p^*,s^*}^{s-1} C_{\rho,x^+}^{s-p} \wedge C_{p^*,s^*}^{p-1}\right)$;
$0 < C_4 < C_0/C_1$;
$n := 0$;
while $\|y^\delta - F(x_n)\| > \tau\delta$ **do**
Choose $\Psi_n \in F'(x_n)^* J_r(y^\delta - F(x_n)) + \alpha\partial\Phi(x_n)$; $0 \leq \alpha \leq C_4\|y^\delta - F(x_n)\|^r$;
$\lambda_{\max} := C_2\left\{\frac{[C_0\|y^\delta - F(x_n)\|^r - \alpha C_1]^{s-1}}{\|\Psi_n\|^s} \wedge \frac{[C_0\|y^\delta - F(x_n)\|^r - \alpha C_1]^{p-1}}{\|\Psi_n\|^p}\right\}$;
Choose $0 < \lambda_n \leq \lambda_{\max}$;
$x_{n+1} := J_{p^*}^*(J_p(x_n) - \lambda_n\Psi_n)$;
$n := n + 1$;
end while
$x_N := x_n$;

4. Convergence Analysis

We formalize the results of last section in the next theorem. Note that in principle, $N(\delta) = \infty$ is possible, which means that termination of Algorithm 1 is not guaranteed yet.

Theorem 2. *Let X be locally uniformly smooth and $s-$convex for some $s \geq p > 1$. Fix $r > 1$ and suppose that Assumption 1 holds true. If $x_0 \in B_\rho(x^+, \Delta_p)$, $\tau > (1+\eta)/(1-\eta)$ in (5) and $\lambda_n \in (0, \lambda_{\max}]$, with $\lambda_{\max}$ defined in (22), then*

- *all the iterates are well-defined and belong to $B_\rho(x^+, \Delta_p)$ as long as the discrepancy principle (5) is not satisfied;*
- *the iterates have the monotonically decreasing error behavior*

$$\Delta_p\left(x^+, x_{n+1}\right) < \Delta_p\left(x^+, x_n\right) \tag{26}$$

for $n = 0, ..., N(\delta) - 1$;

- *the inequality*

$$\lambda_n \left\|b_n^\delta\right\|^r \lesssim \Delta_p\left(x^+, x_n\right) - \Delta_p\left(x^+, x_{n+1}\right) \tag{27}$$

holds for $n = 0, ..., N(\delta) - 1$;

- *the generated sequence is uniformly bounded:*

$$\|x_n\| \leq C_{\rho,x^+} \text{ for all } \delta > 0 \text{ and } n = 0, ..., N(\delta), \tag{28}$$

where $C_{\rho,x^+} > 0$ is defined in (19).

Proof. We employ an inductive argument: suppose that for some $n < N(\delta)$, all the iterates $x_0, ..., x_n$ are well defined and belong to $B_\rho(x^+, \Delta_p)$. Since $x_n \in B_\rho(x^+, \Delta_p)$, $\tau > (1+\eta)/(1-\eta)$ and $\lambda_n \leq \lambda_{\max}$, the reasoning in last section shows that

$$\begin{aligned}\Delta_p\left(x^+, x_{n+1}\right) - \Delta_p\left(x^+, x_n\right) \leq \lambda_n C_{p^*,s^*}\left(C_{\rho,x^+}^{p-s^*(p-1)} \lambda_{\max}^{s^*-1} \|\Psi_n\|^{s^*} \vee \lambda_{\max}^{p^*-1} \|\Psi_n\|^{p^*}\right) \\ - \lambda_n\left(C_0\|b_n\|^r + \alpha_n C_1\right)\end{aligned}$$

$$\leq -\lambda_n (1 - C_2) [C_0 \|b_n\|^r - \alpha_n C_1] \overset{(25)}{\lesssim} -\lambda_n \left\|b_n^\delta\right\|^r.$$

Thus, (27) holds true. The result also implies (26) and consequently

$$\Delta_p\left(x^+, x_{n+1}\right) < \Delta_p\left(x^+, x_n\right) < \rho.$$

Therefore $x_{n+1} \in B_\rho(x^+, \Delta_p) \subset \mathrm{D}(F)$ and the induction proof is complete. □

In order to prove that Algorithm 1 terminates, we first prove that $\lambda_{\max} \geq \lambda_{\min} > 0$ with

$$\lambda_{\min} := c \|b_n\|^t, \ t \geq -r, \tag{29}$$

where $c > 0$ is a constant independent of n and δ, and then restrict λ_n to the interval $[\lambda_{\min}, \lambda_{\max}]$. We start assuming the subdifferentiable functional Φ verifies the following assumption:

Assumption 3. *The sequence $(z_n^*)_{n\in\mathbb{N}} \subset X^*$ is uniformly bounded on n and δ whenever $(x_n)_{n\in\mathbb{N}} \subset X$ is uniformly bounded on the same variables, where $z_n^* \in \partial\Phi(x_n)$ for each $n \in \mathbb{N}$.*

Many usual choices of Φ satisfy above assumption. For instance:

- $\Phi(x) := \int_0^{\|x\|} \varphi(s)\, ds$, where φ is a gauge function, see Section 2. In this case, due to Asplund's Theorem (9), $\partial\Phi(x) = J_\varphi(x)$ and therefore
$$z_n^* \in \partial\Phi(x_n) \Rightarrow \|z_n^*\| = \varphi(\|x_n\|).$$
Since φ is continuous, the desired result follows. For the particular choice $\varphi(t) = t^{p-1}$ we have the well-known and widely used penalization functional $\Phi(x) = \frac{1}{p}\|x\|^p$. Of course the choice $\Phi(x) = \frac{1}{p}\|x - x_0\|^p$ is also permitted.
- $\Phi(x) := \Delta_p(x, \overline{x})$, where $\overline{x} \in X$ is a fixed vector. Then $\partial\Phi(x) = J_p(x) - J_p(\overline{x})$ and $\|z_n^*\| \leq \|x_n\|^{p-1} + \|\overline{x}\|^{p-1}$.
See that, if $\overline{x} = x_0$, then the upper bound C_1 in (21) can be chosen equals ρ because $\Phi(x^+) = \Delta_p(x^+, x_0) < \rho$.
- $\Phi(x) := \|x\|$ is also possible because, see [8, Prop. 3.4, Ch. I],
$$\partial\|x\| = \{z^* \in X^* : \langle z^*, x\rangle = \|x\| \text{ and } \|z^*\| = 1\}, \ x \neq 0.$$
Therefore, $\|z_n^*\| \leq 1$ for all $n \in \mathbb{N}$.
- $\Phi(x) := \Phi_n(x) = \Delta_p(x, x_{n-1})$. Here, $\partial\Phi_n(x) = J_p(x) - J_p(x_{n-1})$. Thus, $\|z_n^*\| \leq \|x_n\|^{p-1} + \|x_{n-1}\|^{p-1}$.
Observe that in this case,
$$\Phi_n\left(x^+\right) = \Delta_p\left(x^+, x_{n-1}\right) \leq \Delta_p\left(x^+, x_0\right) < \rho.$$
for all $n \in \mathbb{N}$, and therefore, $C_1 = \rho$ can be chosen again.

Remark 4. *The penalization functional Φ has a strong influence in the quality of the reconstructed solution. A suitable selection of Φ can help to preserve desired characteristics of the solution, whereas a bad choice of this functional may destroy these features.*

If the sought-for solution has a sparse representation in a fixed basis for example, this characteristic is not completely damaged during the iteration assuming that a suitable penalization functional Φ is chosen. Sparsity constraints are preserved for example, if the functionals $\Phi(x) = \frac{1}{p}\|x - x_0\|^p$ or $\Phi(x) = \Delta_p(x, x_0)$, for $p \approx 1$, are used.

In these cases, $\Phi(x)$ has a small value for those vectors x, whose representation of $x - x_0$ in the fixed basis is sparse. This means that these functionals penalize vectors with a non-sparse representation more severely and accordingly, give priority for sparse reconstructions, see [9, 12] *and the numerical experiments in Section 5.*

From (22) and (25) it follows that

$$\lambda_{\max} \gtrsim \frac{\left\|b_n^\delta\right\|^{r(s-1)}}{\|\Psi_n\|^s} \wedge \frac{\left\|b_n^\delta\right\|^{r(p-1)}}{\|\Psi_n\|^p}. \tag{30}$$

The uniform bound (28) together with Assumption 3 implies that the sequences $(z_n^*)_{0\le n\le N(\delta)}$ are uniformly bounded on n and δ. Moreover, from inequality $\alpha \lesssim \left\|b_n^\delta\right\|^r$ and Assumption $1(b)$, it follows that

$$\|\Psi_n\| \le M\left\|b_n^\delta\right\|^{r-1} + \alpha\|z_n^*\| \lesssim \left\|b_n^\delta\right\|^{r-1} + \left\|b_n^\delta\right\|^r \lesssim \left\|b_n^\delta\right\|^{r-1} \vee \left\|b_n^\delta\right\|^r. \tag{31}$$

This implies in view of (30) that

$$\lambda_{\max} \gtrsim \left\|b_n^\delta\right\|^{s-r} \wedge \left\|b_n^\delta\right\|^{-r} \wedge \left\|b_n^\delta\right\|^{p-r}, \tag{32}$$

which proves that there exists $c > 0$, independent of n and δ and $t \ge -r$ in (29) such that $0 < \lambda_{\min} \le \lambda_{\max}$.

We are now well-prepared to prove termination of Algorithm 1.

Theorem 5 (Termination)**.** *Assume all the hypotheses of Theorem 2 and assume additionally that Assumption 3 holds true as well as $\lambda_n \in [\lambda_{\min}, \lambda_{\max}]$, where $\lambda_{\max}$ is defined in* (22) *and $\lambda_{\min}$ in* (29)*. Then Algorithm 1 terminates.*

Proof. From (32) it follows that the interval $[\lambda_{\min}, \lambda_{\max}]$ is non-empty provided the constant c in (29) small enough. Since $\lambda_n \ge \lambda_{\min} \gtrsim \left\|b_n^\delta\right\|^t$, with $t \ge -r$, it follows from (5) that for any $\ell \le N(\delta)$,

$$\begin{aligned}(\tau\delta)^{t+r}\,\ell &\le \sum_{n=0}^{\ell-1}\left\|b_n^\delta\right\|^{t+r} \lesssim \sum_{n=0}^{\ell-1}\lambda_n\left\|b_n^\delta\right\|^r \overset{(27)}{\lesssim} \sum_{n=0}^{\ell-1}\left(\Delta_p\left(x^+, x_n\right) - \Delta_p\left(x^+, x_{n+1}\right)\right)\\ &= \Delta_p\left(x^+, x_0\right) - \Delta_p\left(x^+, x_\ell\right) \le \Delta_p\left(x^+, x_0\right) < \infty,\end{aligned}$$

which can only be possible if $\ell < \infty$ and consequently $N(\delta) < \infty$. □

Weak convergence follows from (28) and reflexivity of X. More precisely, let $(\delta_j)_{j\in\mathbb{N}}$ be a zero-sequence and assume all the hypotheses of last Theorem. If the operator F is weakly closed, then each subsequence of $\left(x_{N(\delta_j)}\right)_{j\in\mathbb{N}}$ has itself a subsequence which converges weakly to a solution of (1). If this solution is unique, then $x_{N(\delta)} \rightharpoonup x^+$ as $\delta \to 0$, see e.g. [17, Cor. 3.5].

Remark 6. *The reasoning in last Theorem shows that the sequences of the residuals $\left(b_n^\delta\right)_{0\le n\le N(\delta)}$ are uniformly bounded. Thus, from* (32)*,*

$$\lambda_{\max} \gtrsim \left\|b_n^\delta\right\|^{s-r}. \tag{33}$$

Additionally, inequality (31) *guarantees that the sequences $(\Psi_n)_{0\le n\le N(\delta)}$ are uniformly bounded too. Hence, in view of* (25)*, the definition of $\lambda_{\max}$ in* (22) *could be changed by the much simpler expression*

$$\overline{\lambda}_{\max} := C_5\frac{\left\|b_n^\delta\right\|^{r(s-1)}}{\|\Psi_n\|^s}, \tag{34}$$

with $C_5 > 0$ being a small constant. In this case, $\overline{\lambda}_{\max} \leq \lambda_{\max}$ and the property

$$0 < \lambda_n \leq \overline{\lambda}_{\max} \Longrightarrow \Delta_p\left(x^+, x_{n+1}\right) < \Delta_p\left(x^+, x_n\right)$$

would still hold true.

Some obvious possibilities for choosing a step-size satisfying $\lambda_n \in [\lambda_{\min}, \lambda_{\max}]$ are: $\lambda_n = c\left\|b_n^\delta\right\|^{s-r}$, for a small constant $c > 0$, $\lambda_n = \overline{\lambda}_{\max}$ and $\lambda_n = \lambda_{\max}$, see (33) and (34). Further, λ_n can be randomically chosen in the interval $[\lambda_{\min}, \lambda_{\max}]$.

Due to (33) we are allowed to change the definition of $\lambda_{\min}$ in (29) to

$$\lambda_{\min} := c\left\|b_n\right\|^t, \ t > -r \tag{35}$$

and we still have $\lambda_{\max} \geq \lambda_{\min}$ provided $c > 0$ small enough. This more restrictive definition of $\lambda_{\min}$ will be important for proving that $F(x_n) \to y$ in the noiseless case, see (40).

Remark 7. *If $\alpha \equiv 0$ is chosen, definition* (17) *becomes*

$$J_p(x_{n+1}) := J_p(x_n) - \lambda_n F'(x_n)^* j_r\left(F(x_n) - y^\delta\right),$$

which is a gradient-like method applied directly to the non-linear equation (1)*. In this case, the step-size* (34) *is transformed in*

$$\lambda_{ME} = C_5 \frac{\left\|b_n^\delta\right\|^{r(s-1)}}{\left\|A_n^* J_r\left(b_n^\delta\right)\right\|^s}$$

and the resulting gradient method is a variation of the Minimal Error method, see [20] *and the first section of this paper. Further,*

$$\begin{aligned}\left\|A_n^* J_r\left(b_n^\delta\right)\right\|^{p^* r^*(s-1)} &= \left\langle J_{p^*}^*\left(A_n^* J_r\left(b_n^\delta\right)\right), A_n^* J_r\left(b_n^\delta\right)\right\rangle^{r^*(s-1)} \\ &\leq \left\|A_n J_{p^*}^*\left(A_n^* J_r\left(b_n^\delta\right)\right)\right\|^{r^*(s-1)} \left\|b_n^\delta\right\|^{r(s-1)},\end{aligned}$$

and therefore, the step-size

$$\lambda_{SD} := C_5 \frac{\left\|A_n^* J_r\left(b_n^\delta\right)\right\|^{p^* r^*(s-1)-s}}{\left\|A_n J_{p^*}^*\left(A_n^* J_r\left(b_n^\delta\right)\right)\right\|^{r^*(s-1)}} \tag{36}$$

satisfies $\lambda_{SD} \leq \lambda_{ME} \leq \lambda_{\max}$. Moreover, using Assumption 1(b) and the boundedness of the residual sequence, it follows that

$$\lambda_{SD} \geq C_5 \frac{\left\|A_n^* J_r\left(b_n^\delta\right)\right\|^{r^*(s-1)-s}}{M^{r^*(s-1)}} \geq C_6, \tag{37}$$

whenever $s \leq r$, where $C_6 > 0$ is a small constant. Thus $\lambda_{SD} \in [\lambda_{\min}, \lambda_{\max}]$. Observe that the step-size (36) *reduces to*

$$\lambda_{SD} := \frac{\left\|A_n^* b_n^\delta\right\|^2}{\left\|A_n A_n^* b_n^\delta\right\|^2}$$

in Hilbert spaces whenever $C_5 = 1$ and $p = s = r = 2$. This means that this variation of the Steepest Descent method in Banach spaces is also included as particular case in our convergence analysis[5]*. Finally, inequality* (37) *implies that the Landweber method*

[5]In Hilbert spaces, $C_5 \leq 2\left(1 - \eta - \frac{1+\eta}{\tau}\right)$. Thus $C_5 = 1$ can be chosen provided $\eta < \frac{1}{2}$ and $\tau \geq 2\frac{1+\eta}{1-2\eta}$, see again [20].

satisfies $\lambda_{LW} \in [\lambda_{\min}, \lambda_{\max}]$ *with the constant step-size* $\lambda_{LW} = C_6$ *provided* $s \leq r$. *In this case, the inequalities*

$$0 < C_6 \left\| b_n^\delta \right\|^0 = \lambda_{\min} = \lambda_{LW} \leq \lambda_{SD} \leq \lambda_{ME} = \overline{\lambda}_{\max} \leq \lambda_{\max}$$

hold.

4.1. **The noiseless case.** In this subsection and in the next one, we must clearly differ between the noisy ($\delta > 0$) and the noise-free ($\delta = 0$) cases. Therefore, a notation with a superscript δ will be adopted whenever data is contamined with noise: x_n^δ, b_n^δ, A_n^δ, etc., and in contrast, x_n, b_n, A_n, etc. always originate from exact data. The initial iterate is chosen independently of δ: $x_0^\delta = x_0$.

Algorithm 1 is well-defined in the noiseless situation if the inequalities in (20) are replaced with $0 < C_0 \leq 1 - \eta$, and in this case, all the results of Theorem 2 still hold true. Notice that, although in the noise-free case, the stop-criterion (5) is well-defined and reads $\|b_n\| = 0$, in which case x_n is a solution of (1) , Algorithm 1 generally does not stop, but generates a sequence which converges strongly to a solution of (1) as we will prove in Theorem 8 below.

Theorem 8 (Convergence without noise)**.** *Let* X *be locally uniformly smooth and* $s-$*convex for some* $s \geq p > 1$. *Fix* $r > 1$ *and suppose that Assumptions 1 and 3 hold true. If* $\delta = 0$, $x_0 \in B_\rho(x^+, \Delta_p)$ *and* $\lambda_n \in [\lambda_{\min}, \lambda_{\max}]$, *where* $\lambda_{\min}$ *and* $\lambda_{\max}$ *are defined in* (35) *and* (22) *respectively, then Algorithm 1 either terminates with a solution of* (1) *or generates a sequence which converges to a solution of this equation. In particular, if the solution of* (1) *is unique in* $B_\rho(x^+, \Delta_p)$, *then* $x_n \to x^+$ *as* $n \to \infty$.

Proof. If Algorithm 1 stops after $n \in \mathbb{N}$ iterations, then the current iterate is a solution of (1) because $\|y - F(x_n)\| = \|b_n\| = 0$. Otherwise, $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, as we will prove now. Let $m, l \in \mathbb{N}$ with $m \leq l$ and choose $z \in \{m, \ldots, l\}$ such that

$$\|b_z\| \leq \|b_n\| \text{ for all } n \in \{m, \ldots, l\}. \tag{38}$$

As X is $s-$convex and the sequence $(x_n)_{n\in\mathbb{N}}$ is bounded, see (28), it follows from (14) that

$$\|x_m - x_l\|^s \leq 2^{s-1}\left(\|x_m - x_z\|^s + \|x_z - x_l\|^s\right) \lesssim \Delta_p(x_z, x_m) + \Delta_p(x_z, x_l).$$

Three Points Identity (12) implies now that

$$\|x_m - x_l\|^s \lesssim \beta_{m,z} + \beta_{l,z} + f(z, m, l) \tag{39}$$

with

$$\beta_{m,z} := \Delta_p\left(x^+, x_m\right) - \Delta_p\left(x^+, x_z\right)$$

and

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

By monotonicity (26), we conclude that $\Delta_p(x^+, x_n) \to \gamma \geq 0$ as $n \to \infty$. Thus, $\beta_{m,z}$ and $\beta_{l,z}$ converge to zero as $m \to \infty$ (which causes $z \to \infty$ and $l \to \infty$). Further, using $z_n^* \in \partial\Phi(x_n)$, it follows from definition (17) that

$$\begin{aligned} f(z, m, l) &\leq \sum_{n=m}^{l-1} \left|\left\langle J_p(x_{n+1}) - J_p(x_n), x_z - x^+\right\rangle\right| \\ &\leq \sum_{n=m}^{l-1} \lambda_n \left|\left\langle A_n^* J_r(b_n) + \alpha z_n^*, x_z - x^+\right\rangle\right| \end{aligned}$$

$$\leq \sum_{n=m}^{l-1} \lambda_n \left(\|b_n\|^{r-1} \left\|A_n\left(x_z - x^+\right)\right\| + \alpha \|z_n^*\| \left\|x_z - x^+\right\| \right).$$

As the sequence $(x_n)_{n\in\mathbb{N}}$ is bounded, it follows from Assumption 3 that $(z_n^*)_{n\in\mathbb{N}}$ is bounded too and since $0 \leq \alpha \lesssim \|b_n\|^r$,

$$\alpha \|z_n^*\| \left\|x_z - x^+\right\| \lesssim \|b_n\|^r.$$

From assumption $1(c)$,

$$\begin{aligned}
\left\|A_n\left(x_z - x^+\right)\right\| &\leq \left\|A_n\left(x^+ - x_n\right)\right\| + \|A_n\left(x_z - x_n\right)\| \\
&\leq \|b_n\| + \left\|b_n - A_n\left(x^+ - x_n\right)\right\| + \|F\left(x_z\right) - F\left(x_n\right)\| \\
&+ \|F\left(x_z\right) - F\left(x_n\right) - F'\left(x_n\right)\left(x_z - x_n\right)\| \\
&\leq (\eta+1)\left(\|b_n\| + \|F\left(x_z\right) - F\left(x_n\right)\|\right) \\
&\leq (\eta+1)\left(2\|b_n\| + \|b_z\|\right) \overset{(38)}{\leq} 3(\eta+1)\|b_n\|.
\end{aligned}$$

Putting all together, we arrive at

$$f(z,m,l) \lesssim \sum_{n=m}^{l-1} \lambda_n \|b_n\|^r \overset{(27)}{\lesssim} \sum_{n=m}^{l-1} \left(\Delta_p\left(x^+, x_n\right) - \Delta_p\left(x^+, x_{n+1}\right)\right) = \beta_{m,l},$$

which together with (39) implies that $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. Since X is complete, it converges to some $x_\infty \in X$. Now, as $\lambda_n \geq \lambda_{\min}$,

$$\sum_{n=0}^{\infty} \|b_n\|^{r+t} \lesssim \sum_{n=0}^{\infty} \lambda_n \|b_n\|^r \overset{(27)}{<} \infty. \tag{40}$$

Since $r > -t$, it follows that $\|y - F\left(x_n\right)\| = \|b_n\| \to 0$ as $n \to \infty$ and since F is continuous, we have $y = F\left(x_\infty\right)$. If (1) has only one solution in $B_\rho\left(x^+, \Delta_p\right)$ then $x_\infty = x^+$. □

4.2. **Regularization Property.** Our next step is proving a stability property, which together with the convergence result of Theorem 8 will ensure the regularization property of Theorem 11 below. To this end, we require the following assumption:

Assumption 9. *Let $(\delta_j)_{j\in\mathbb{N}}$ be a zero-sequence and $\overline{n} := \limsup_{j\to\infty} N(\delta_j)$ (which can possibly be ∞). Then, for every $n \leq \overline{n}$,*

$$\lim_{j\to\infty} \lambda_n^{\delta_j} =: \overline{\lambda}_n \in \mathbb{R} \qquad \text{and} \qquad \lim_{j\to\infty} \alpha_n^{\delta_j} =: \overline{\alpha}_n \in \mathbb{R}.$$

Of course the above assumption is true for instance, if λ_n^δ and α_n^δ depend continuously on δ. This assumption, combined with appropriate restrictions on the data-space Y and on the functional Φ, ensures stability, as next lemma shows.

Lemma 10. *Assume all the hypotheses of Theorem 5 and fix a zero-sequence $(\delta_j)_{j\in\mathbb{N}}$. Assume additionally that Assumptions 3 and 9 hold true, Y is locally uniformly smooth and Φ is continuously G-differentiable in $B_\rho\left(x^+, \Delta_p\right)$. Then the mapping $dT\alpha$ defined in (16) is single-valued, $\overline{\lambda}_n \in [\lambda_{\min}, \lambda_{\max}]$ and $\overline{\alpha}_n \in [0, C_4\|b_n\|^r]$ (see Assumption 9) and for each $n \leq \overline{n}$, it holds that*

$$x_n^{\delta_j} \to x_n \text{ as } j \to \infty,$$

where the noise-free sequence $(x_n)_{n\in\mathbb{N}}$ *is the one associated with the step-size sequence* $\left(\overline{\lambda}_n\right)_{n\in\mathbb{N}}$ *and the regularization sequence* $(\overline{\alpha}_n)_{n\in\mathbb{N}}$.

Proof. We prove the statement by induction. Assume that for a fixed number $n \in \mathbb{N}$ satisfying $n < \overline{n}$, it holds

$$x_n^{\delta_j} \to x_n \text{ as } j \to \infty.$$

From (2) it follows that $y^{\delta_j} \to y$ as $j \to \infty$ and since F is a continuous function, $\left\|b_n^{\delta_j}\right\| \to \|b_n\|$ as $j \to \infty$. Consequently, $\lambda_{\min}^{\delta_j} \to \lambda_{\min}$ and since $\alpha_n^{\delta_j}$ converges to $\overline{\alpha}_n$ as $j \to \infty$, it follows from $0 \leq \alpha_n^{\delta_j} \leq C_4 \left\|b_n^{\delta_j}\right\|^r$ that $\overline{\alpha}_n \in [0, C_4 \|b_n\|^r]$ and accordingly, $\lambda_{\max}^{\delta_j} \to \lambda_{\max}$. Therefore the inequality $\lambda_{\min}^{\delta_j} \leq \lambda_n^{\delta_j} \leq \lambda_{\max}^{\delta_j}$ implies that $\overline{\lambda}_n \in [\lambda_{\min}, \lambda_{\max}]$, which proves that x_{n+1} is well-defined. Now, the locally uniform smoothness of Y implies the duality mapping J_r is single-valued and continuous and since F' is continuous too,

$$F'\left(x_n^{\delta_j}\right)^* J_r\left(F\left(x_n^{\delta_j}\right) - y^{\delta_j}\right) \to F'(x_n)^* J_r(F(x_n) - y)$$

as $j \to \infty$. Since Φ is continuously G-differentiable, $\partial\Phi$ coincides with the G-derivative $\nabla\Phi$ and it is therefore single-valued and continuous. Then $\nabla\Phi\left(x_n^{\delta_j}\right) \to \nabla\Phi(x_n)$ and

$$\Psi_n^{\delta_j} = dT_\alpha\left(x_n^{\delta_j}\right) = F'\left(x_n^{\delta_j}\right)^* J_r\left(F\left(x_n^{\delta_j}\right) - y^{\delta_j}\right) + \alpha_n^{\delta_j} \nabla\Phi\left(x_n^{\delta_j}\right)$$

converges to

$$\Psi_n := dT_\alpha(x_n) = F'(x_n)^* J_r(F(x_n) - y) + \overline{\alpha}_n \nabla\Phi(x_n)$$

as $j \to \infty$. Finally, the continuity of $J_{p^*}^*$ implies that

$$x_{n+1}^{\delta_j} = J_{p^*}^*\left(J_p\left(x_n^{\delta_j}\right) - \lambda_n^{\delta_j}\Psi_n^{\delta_j}\right) \to J_{p^*}^*\left(J_p(x_n) - \overline{\lambda}_n\Psi_n\right) = x_{n+1}$$

as $j \to \infty$. □

From Theorem 8 and Lemma 10 we obtain the regularization property.

Theorem 11 (Regularization Property)**.** *Assume all the hypotheses of last Lemma and fix a zero-sequence* $(\delta_j)_{j\in\mathbb{N}}$. *Then the sequence* $\left(x_{N(\delta_j)}^{\delta_j}\right)_{j\in\mathbb{N}}$ *converges to a solution of* (1) *in* $B_\rho(x^+, \Delta_p)$. *In particular, if there is a unique solution of this equation in* $B_\rho(x^+, \Delta_p)$, *then*

$$\lim_{j\to\infty} x_{N(\delta_j)}^{\delta_j} = x^+.$$

Proof. Assume first the sequence $(N(\delta_j))_{j\in\mathbb{N}}$ is bounded. In this case, $\overline{n} < \infty$, and since $(N(\delta_j))_{j\in\mathbb{N}} \subset \mathbb{N}$, we conclude that there exists a number $J \in \mathbb{N}$ such that $\|N(\delta_j)\| \leq \overline{n}$ for all $j \geq J$. Thus, the sequence $\left(x_{N(\delta_j)}^{\delta_j}\right)_{j\geq J}$ splits into at most $\overline{n}$ subsequences having the form $\left(x_n^{\delta_{j_k}}\right)_{k\in\mathbb{N}}$, where $N(\delta_{j_k}) = n \leq \overline{n}$. From Theorem 10, $x_n^{\delta_{j_k}} \to x_n$ as $k \to \infty$ and therefore

$$\lim_{k\to\infty} x_{N\left(\delta_{j_k}\right)}^{\delta_{j_k}} = \lim_{k\to\infty} x_n^{\delta_{j_k}} = x_n.$$

Now, x_n is a solution of (1) because from (2) and (5),

$$\|y - F(x_n)\| = \lim_{k\to\infty} \left\|y - F\left(x_n^{\delta_{j_k}}\right)\right\|$$

$$\leq \lim_{k\to\infty}\left(\left\|y-y^{\delta_{j_k}}\right\|+\left\|y^{\delta_{j_k}}-F\left(x_{N\left(\delta_{j_k}\right)}^{\delta_{j_k}}\right)\right\|\right)$$
$$\leq \lim_{k\to\infty}(1+\tau)\,\delta_{j_k}=0.$$

But, $x_{n+k}=x_n$ for all $k\in\mathbb{N}$ whenever x_n is a solution of (1), because in this case $\|b_n\|=0$ and consequently, $\lambda_{\max}=0$ in (22). Thus, all the subsequences of $\left(x_{N(\delta_j)}^{\delta_j}\right)_{j\in\mathbb{N}}$ converge to the same solution.

Suppose now that $N(\delta_j)\to\infty$ as $j\to\infty$ and let $\epsilon>0$ be given. As the Bregman distance is a continuous function in both arguments, there exists $\gamma=\gamma(\epsilon)>0$ such that

$$\Delta_p\left(x,x_n^{\delta_j}\right)<\frac{\epsilon^s}{C}\text{ whenever }\left\|x-x_n^{\delta_j}\right\|<\gamma, \tag{41}$$

where $C>0$ is the constant from (14). From Theorem 8, there exists $M\in\mathbb{N}$ and a solution x_∞ of (1) such that

$$\|x_\infty-x_n\|<\frac{\gamma}{2}\text{ for all }n\geq M.$$

As $N(\delta_j)\to\infty$, there exists $J_1\in\mathbb{N}$ such that $N(\delta_j)\geq M$ for all $j\geq J_1$. Finally, from Lemma 10, there exists $J_2\geq J_1$ such that

$$\left\|x_M-x_M^{\delta j}\right\|<\frac{\gamma}{2}\text{ for all }j\geq J_2.$$

Thus, for any $j\geq J_2$,

$$\left\|x_\infty-x_M^{\delta_j}\right\|\leq\|x_\infty-x_M\|+\left\|x_M-x_M^{\delta_j}\right\|<\gamma$$

and consequently,

$$\left\|x_\infty-x_{N(\delta_j)}^{\delta_j}\right\|^s\overset{(14)}{\leq}C\Delta_p\left(x_\infty,x_{N(\delta_j)}^{\delta_j}\right)\overset{(26)}{\leq}C\Delta_p\left(x_\infty,x_M^{\delta_j}\right)\overset{(41)}{<}\epsilon^s$$

□

5. Numerical Experiments

In this section we test our method for the non-linear and severely ill-posed inverse problem of EIT (Electrical Impedance Tomography) introduced by Calderón [5]. We focus only on the main ideas and refer the interested reader to the survey article [4] and the close related work [6] for more details.

Let $\Omega\subset\mathbb{R}^2$ be a bounded and simply connected Lipschitz domain. Applying electric currents $g\colon\partial\Omega\to\mathbb{R}$ on its boundary[6] and reading the resulting voltages $f\colon\partial\Omega\to\mathbb{R}$ on its boundary as well, we aim to reconstruct the electric conductivity $\gamma\colon\Omega\to\mathbb{R}$ in the whole of Ω.

Once fixed the electric current $g\in L_\diamondsuit^2(\partial\Omega):=\left\{v\in L^2(\partial\Omega):\int_{\partial\Omega}v=0\right\}$, there exists a unique electric potential $u\in H_\diamondsuit^1(\Omega):=\left\{v\in H^1(\Omega):\int_{\partial\Omega}v=0\right\}$ satisfing the variational equation

$$\int_\Omega\gamma\nabla u\nabla\varphi=\int_{\partial\Omega}g\varphi\text{ for all }\varphi\in H_\diamondsuit^1(\Omega), \tag{42}$$

[6]The symbol ∂ was adopted in the previous sections as the subdifferential of a convex function. In this section however, we also use it to represent the boundary of a set. Its meaning will be made clear by the context.

provided $\gamma \in L^\infty_+(\Omega) := \{v \in L^\infty(\Omega) : v \geq C \text{ a.e.}\}$, where C is a positive constant. Further, since $u \in H^1_\diamondsuit(\Omega)$, its trace belongs to $L^2_\diamondsuit(\partial\Omega)$ and it is just the voltage on the boundary: $f = u|_{\partial\Omega}$.[7]

For each fixed conductivity $\gamma \in L^\infty_+(\Omega)$, we define a bounded linear operator $\Lambda_\gamma \colon L^2_\diamondsuit(\partial\Omega) \to L^2_\diamondsuit(\partial\Omega)$, $g \mapsto f$, which we call *Neumann-to-Dirichlet* map (in short NtD). The *forward operator* associated with EIT is now defined by $F \colon L^\infty_+(\Omega) \subset L^\infty(\Omega) \to \mathcal{L}(L^2_\diamondsuit(\partial\Omega), L^2_\diamondsuit(\partial\Omega))$,

$$F(\gamma) := \Lambda_\gamma.$$

Finding γ in above equation for a given Λ_γ is the EIT inverse problem we want to solve. Astala and Päivärinta proved in [1] the uniqueness of solutions of this inverse problem provided Λ_γ operates between $H^{-1/2}_\diamondsuit(\partial\Omega)$ and $H^{1/2}_\diamondsuit(\partial\Omega)$.

In practical situations, the NtD map is not completely available. Only partial data can be observed and the best one can do is applying $d \in \mathbb{N}$ currents $g_j \in L^2_\diamondsuit(\partial\Omega)$, $j = 1, \ldots, d$, and then recording the resulting voltages $f_j = \Lambda_\gamma g_j$. We thus fix the vector $G := (g_1, ..., g_d) \in \left(L^2_\diamondsuit(\partial\Omega)\right)^d$ and introduce the operator $F_G \colon L^\infty_+(\Omega) \subset L^\infty(\Omega) \to \left(L^2_\diamondsuit(\partial\Omega)\right)^d$, $\gamma \mapsto (\Lambda_\gamma g_1, ..., \Lambda_\gamma g_d)$, which is F-differentiable[8], see, e.g. [16].

Since an analytical solution of (42) is not available in general, the inverse problem needs to be solved with help of a computer. For this reason, we construct a Delaunay triangulation for Ω, $\mathcal{T} = \{T_i : i = 1, \ldots, M\}$, with $M = 1684$ triangles (see the third picture in Figure 1 below) and approximate γ by piecewise constant conductivities: define the finite dimensional space $V := \operatorname{span}\{\chi_{T_1}, \ldots, \chi_{T_M}\} \subset L^\infty(\Omega)$ and search conductivities in V, which means that our reconstructions always have the form $\sum_{i=1}^M \theta_i \chi_{T_i}$, with $(\theta_1, ..., \theta_M) \in \mathbb{R}^M$. Unfortunately, the Banach space $L^\infty(\Omega)$ is not regular enough to be included in the convergence analysis of last section (it is not locally uniformly smooth for instance). We thus fix $p \in (1, \infty)$ and equip the space V with the L^p-norm: $V_p := \left(V, \|\cdot\|_{L^p(\Omega)}\right)$. The spaces $L^p(\Omega)$, $1 < p < \infty$, are $p\wedge 2-$smooth and $p\vee 2-$convex, and the duality mapping $J_p \colon L^p(\Omega) \to L^{p^*}(\Omega)$ is computed by

$$J_p(f) = |f|^{p-1} \operatorname{sgn}(f) \tag{43}$$

pointwise. With this new framework our forward operator reads

$$F_G \colon \overline{V} \subset V_p \to \left(L^2(\partial\Omega)\right)^d, \quad \gamma \mapsto (\Lambda_\gamma g_1, ..., \Lambda_\gamma g_d), \tag{44}$$

where $\overline{V} = L^\infty_+(\Omega) \cap V$. Since all the norms are equivalent in finite dimensional spaces, F_G remains F-differentiable. We stress the fact that this restriction in the solution space X is reasonable, because the necessity of using a computer always forces the introduction of a finite dimensional space. Moreover, it is possible to recover only finitely many degrees of freedom of the conductivity from finitely many measurements.

Above framework guarantees the forward operator F_G in (44) satisfies the tangential cone condition, Assumption 1(c), at least in a small ball around a solution, see [16, Theorem 3.4]. Further, its F-derivative, $F'_G \colon \operatorname{int}(\overline{V}) \to \mathcal{L}(V_p, (L^2(\partial\Omega))^d)$, satisfies

[7] The conditions $\int_{\partial\Omega} g = 0$ and $\int_{\partial\Omega} f = 0$ are sufficient to respectively prove existence and uniqueness of solutions. They are physically interpreted as the law of conservation of charge and the grounding of the potential.

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

$F'_G(\gamma)\,h = (w_1|_{\partial\Omega}, ..., w_d|_{\partial\Omega})$, where $w_j \in H^1_\diamondsuit(\Omega)$ is the unique solution of

$$\int_\Omega \gamma \nabla w_j \nabla \varphi = -\int_\Omega h \nabla u_j \nabla \varphi \quad \text{for all } \varphi \in H^1_\diamondsuit(\Omega) \tag{45}$$

with u_j solving (42) for $g = g_j$.

Intending to compute the adjoint operator $F'_G(\gamma)^*\colon (L^2(\partial\Omega))^d \to V_{p^*}$, we fix the vectors $z := (z_1, ..., z_d) \in (L^2(\partial\Omega))^d$ and $h \in V_p$ and apply the following procedure: for each $j = 1, ..., d$, let u_j and ψ_{z_j} be the unique solutions of (42) for $g = g_j$ and $g = z_j$ respectively. Then,

$$\begin{aligned}
\langle F'_G(\gamma)^* z, h\rangle = \langle z, F'_G(\gamma)h\rangle &= \sum_{j=1}^d \left\langle z_j,\, w_j|_{\partial\Omega}\right\rangle \\
&= \sum_{j=1}^d \int_{\partial\Omega} z_j\; w_j|_{\partial\Omega} \overset{(42)}{=} \sum_{j=1}^d \int_\Omega \gamma \nabla \psi_{z_j} \nabla w_j \\
&= \sum_{j=1}^d \int_\Omega \gamma \nabla w_j \nabla \psi_{z_j} \overset{(45)}{=} -\sum_{j=1}^d \int_\Omega h \nabla u_j \nabla \psi_{z_j} = -\left\langle \sum_{j=1}^d \nabla u_j \nabla \psi_{z_j}, h \right\rangle.
\end{aligned}$$

Hence,

$$F'_G(\gamma)^* z = -\sum_{j=1}^d \nabla u_j \nabla \psi_{z_j}. \tag{46}$$

In our tests, the set Ω is defined as the unit square $(0,1)\times(0,1)$ and we supply the current-set G with $d = 12$ independent currents: identifing the faces of Ω with the numbers $m = 0, 1, 2, 3$, we apply the currents

$$g_{3m+k}(x) = \begin{cases} \cos(2k\pi x) & : \text{ on the face } m \\ 0 & : \text{ elsewhere on } \partial\Omega \end{cases}$$

for $k = 1, 2, 3$. The exact solution γ^+ consists of a constant background conductivity $\overline{\gamma} \equiv 0.1$ and an inclusion $B \subset \Omega$ with conductivity 1. Denoting the centroid of T_i by ξ_i, we approximate γ^+ in the space V_p by

$$\widetilde{\gamma}^+ = \overline{\gamma} + \sum_{i=1}^M \sigma_i v_i, \quad \text{with } \sigma_i = \begin{cases} 0.9 & : \ \xi_i \in B \\ 0 & : \text{ otherwise} \end{cases}.$$

Figure 1 illustrates in the images on the left and in the middle, two examples of γ^+, when B is a ball and a four-squares inclusion respectively. The rightmost picture in this figure shows the triangulation $\mathcal{T}$, used to reconstruct the conductivities in our experiments.

The data

$$y := (\Lambda_{\gamma^+} g_1, ..., \Lambda_{\gamma^+} g_d) \tag{47}$$

corresponding to the exact solution γ^+ have been computed using the finite element method (FEM). The problems (42) and (45) have been solved by FEM as well, but using a much coarser discretization mesh than the one used to generate the data. Further, we assume the background conductivity is known and start with the initial iterate $\gamma_0 = \overline{\gamma} \equiv 0.1$. Observe that in this case, the vector $\widetilde{\gamma}^+ - \gamma_0 = \sum_{i=1}^M \sigma_i v_i$ has a

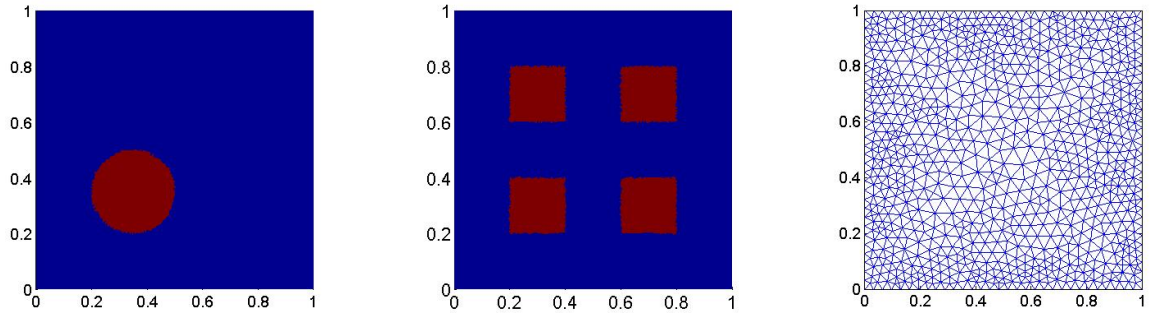


FIGURE 1. Left and middle: two examples of searched-for conductivities with sparsely distributed inclusions. Right: triangulation $\mathcal{T}$ used to reconstruct the conductivities.

sparse representation in the fixed basis of V_p because the searched coefficients σ_i are non-zero only in B.[9] The *relative* L^p-*error* of the $n-$th iterate γ_n,

$$E_n^p := 100\,\frac{\|\gamma_n - \gamma^+\|_{L^p(\Omega)}}{\|\gamma^+\|_{L^p(\Omega)}}$$

is utilized to compare the quality of the reconstructions.

In order to avoid undesirable instability effects, which can arise from an unfavorable selection of the geometry of the mesh, we propose employing a strategy using a *weight-function* $\omega\colon \Omega \to \mathbb{R}$ to define the weighted-space $L^p_\omega(\Omega) := \left\{f\colon \Omega \to \mathbb{R} : \int_\Omega |f|^p\,\omega < \infty\right\}$. The main results have been collected from [24] and [18, Subsection 5.1.2], we only clarify the most important ideas. See that $\|f\|_{L^p_\omega(\Omega)} := \left\|f\omega^{1/p}\right\|_{L^p(\Omega)}$ defines a norm in $L^p_\omega(\Omega)$ if $\omega > 0$, and this norm becomes equivalent to the regular L^p-norm whenever $\omega_{\min} \le \omega \le \omega_{\max}$ with $\omega_{\min}$ and $\omega_{\max}$ being positive constants. In this case, the spaces $L^p(\Omega)$ and $L^p_\omega(\Omega)$ become isomorphic and $L^p_\omega(\Omega)$ inherits all the properties from $L^p(\Omega)$. In particular, $L^p_\omega(\Omega)$ is a $p\vee 2-$convex (thus $s = p\vee 2$ in Theorem 2) and $p\wedge 2-$smooth Banach space, $\left(L^p_\omega\right)^* = L^{p^*}_\omega$, and the duality mapping $J_p\colon L^p_\omega(\Omega) \to L^{p^*}_\omega(\Omega)$ is exactly the same as (43). Following ideas from [24], we further choose

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

where $|T_i|$ is the area of triangle T_i. Finally, we alter the topology of space V_p, changing the norm $L^p(\Omega)$ to $L^p_\omega(\Omega)$,[10] i.e., from now on, $V_p = \left(V, \|\cdot\|_{L^p_\omega(\Omega)}\right)$.

For our first experiment, we observe the exact conductivity γ^+ shown in the leftmost picture of Figure 1. As $Y = \left(L^2(\partial\Omega)\right)^d$ is a Hilbert space, the normalized duality mapping is the identity operator, as a result, we fix $r = 2$ and consequently $J_r(f) = f$ for all $f \in Y$. Because $\widetilde{\gamma}^+ - \gamma_0$ has a sparse representation in the basis $\chi_{T_1}, ..., \chi_{T_M}$, we choose the penalization functional $\Phi(\gamma) = \frac{1}{p}\|\gamma - \gamma_0\|^p$, see Remark 4. Since the value of $\Phi(\gamma^+)$ is known in our experiments, but unknown in practical situations, this

[9] Jin, Khan and Maass proposed a different method to solve the EIT problem in a similar situation, see [11].

[10] The introduction of the weight-function ω changes the evaluation of the adjoint operator in (46) slightly. From [18], we see that $F'_G(\gamma)^* z = -\frac{1}{\omega}\sum_{j=1}^{d} \nabla u_j \nabla \psi_{z_j}$.

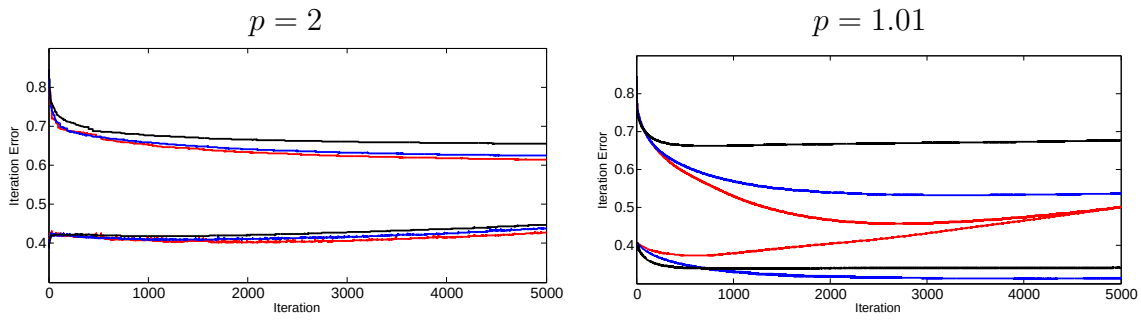


FIGURE 2. Experiments performed without addition of artificial noise. Relative iteration errors E_n^2 (curves from above) and E_n^1 (curves from below) versus the iteration index n in different spaces and with different regularization parameters $\overline{\alpha}$. Red line: $\overline{\alpha} = 0$, blue line: $\overline{\alpha} = 0.1$ and black line: $\overline{\alpha} = 0.9$.

quantity has been overestimated with an error of 50%, this is, we fix $C_1 = 1.5\Phi\left(\gamma^+\right)$, which we consider a reasonable guess. Next, we compare the reconstructions with respect to different spaces $X = V_2$ $(p = 2)$ and $X = V_{1.01}$ $(p = 1.01)$. At the same time, we test the convergence results from Theorem 8 and examine the effect of the regularization parameter α_n. From polarization identity (11) follows that the constant C_{p^*,s^*} in (15) can be set to $1/2$ in Hilbert spaces, and since $C_2 < C_{p^*,s^*}^{1-s}$, we fix $C_2 = 1.9$ for the case $p = 2$. As the constant C_2 depends on both C_{p^*,s^*} and C_{ρ,x^+} in the space $V_{1.01}$, it is much harder to estimate. We try $C_2 = 0.001$ for the case $p = 1.01$, which worked satisfactorily in our experiments. Further, the estimate $\eta = 0.5$ has been utilized and the remaining parameters of Algorithm 1 have been chosen as $C_0 = 1 - \eta$, $\lambda_k = \lambda_{\max}$ and

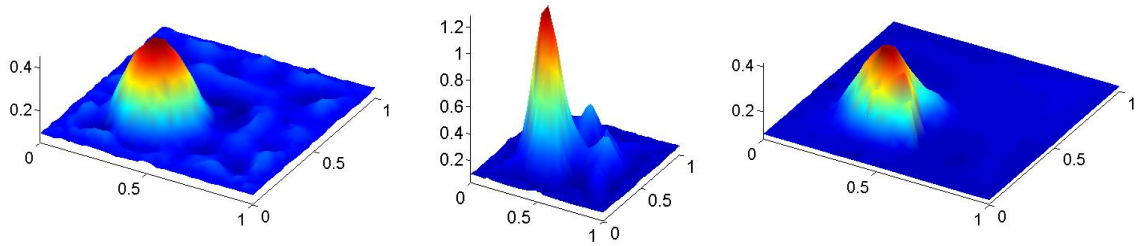
$$\alpha_n := \overline{\alpha}\frac{C_0}{C_1}\left\|y - F_G\left(\gamma_n\right)\right\|^r,$$

with $\overline{\alpha} \in [0, 1)$. The purpose of parameter $\overline{\alpha}$ is regulating on which level the effect of the penalization functional Φ is incorporated into the reconstructions. The larger $\overline{\alpha}$ is, the bigger is the influence of Φ in the iterates. In view of (22) however, one can see that large values of $\overline{\alpha}$ lead to small step-sizes, which can possibly slow down the convergence of the method. Figure 2 confronts the errors E_n^1 and E_n^2 with the iteration index n for $n = 0, ..., 5000$, using three different regularization levels: $\overline{\alpha} = 0$ (red lines), $\overline{\alpha} = 0.1$ (blue lines) and $\overline{\alpha} = 0.9$ (black lines), where no artificial noise is added $(\delta = 0)$.[11]

It is clear that different values of $\overline{\alpha}$ do not affect the quality of the reconstructions in the Hilbert space V_2 very strongly. Moreover, the sequence $\left(E_n^1\right)_{n\in\mathbb{N}}$ does not decrease in this space, which means that the errors in the L^1-norm are not improved. However, this is not the case in $V_{1.01}$, where $\overline{\alpha}$ has an important impact on the reconstructions. See that in this Banach space, the choice $\overline{\alpha} = 0$ produces a fast decrease of the sequence $\left(E_n^2\right)_{n\in\mathbb{N}}$ although, large errors in the L^1-norm are observed after a small improvement. On the other hand, both sequences have a monotonically decreasing behavior for a relatively long period if the value of $\overline{\alpha}$ is increased.

Complementary to Figure 2, we display in Figure 3 a linear interpolation of the iterate $n = 5000$ using the following configurations: leftmost picture: $p = 2$ and $\overline{\alpha} = 0$

[11]Of course some sources of noise such as noise from discretization and computer work precision cannot be avoided and are always present.



$E_N^2 = 61,40\%$, $E_N^1 = 42,64\%$ | $E_N^2 = 49,99\%$, $E_N^1 = 50,08\%$ | $E_N^2 = 67,57\%$, $E_N^1 = 34,10\%$

FIGURE 3. Linear interpolation of iterate γ_{5000} in the noiseless situation, comparing different spaces and regularization levels. Left: $p = 2$ and $\overline{\alpha} = 0$. Middle: $p = 1.01$ and $\overline{\alpha} = 0$. Right: $p = 1.01$ and $\overline{\alpha} = 0.9$.

(this structure corresponds to the red line in the left picture of Figure 2), picture in the middle: $p = 1.01$ and $\overline{\alpha} = 0$ (corresponds to the red line in the right picture of Figure 2) and rightmost picture: $p = 1.01$ and $\overline{\alpha} = 0.9$ (black line in the right picture of Figure 2). The pictures are in different scales and below each of them, the L^1 and L^2 relative errors E_N^1 and E_N^2 for $N = 5000$ are shown. As expected, the Hilbert space V_2 has produced in the first picture, an oversmoothed reconstruction with a very oscillatory background. The picture displayed in the middle exhibits a thin and high inclusion, as predicted for reconstructions in Banach spaces V_p with p close to 1, however, the use of a very small regularization parameter ($\overline{\alpha} = 0$) impedes that the effect of the penalization functional Φ is assimilated by the iterates, leading to a large oscillation in the background and a consequent large error in the L^1-norm. Finally, in the rightmost picture, the large value of $\overline{\alpha}$ fully incorporates the effect of the penalization functional Φ into the reconstruction, which combined with the small value of p, promotes a sparse reconstruction for $\gamma_n - \gamma_0$. As a result, a small error in the L^1-norm is observed and a very stable background is produced. The price to be paid is a large error in the L^2-norm and a slow convergence, due to the resulting small step-sizes, see (22) .

Aiming the examination of the performance of our method when data is corrupted by noise, we contaminate the simulated data y in (47) with artificially generated random noise, with a *relative* noise level $\delta > 0$,

$$y^\delta = y + \delta \,\mathrm{noi}\|y\|_{(L^2(\partial\Omega))^d}, \tag{48}$$

where noi is a uniformly distributed random variable such that $\|\mathrm{noi}\|_{(L^2(\partial\Omega))^d} = 1$. Next, we set $\tau = 1.2\,(1+\eta)\,/\,(1-\eta)$, $\overline{\alpha} = 0.5$ and $\delta = 0.1\%$, and fix all the remaining parameters as in last experiment, except for C_0, which needs to be replaced by

$$C_0 = 1 - \eta - \frac{1+\eta}{\tau} \approx 0.17\,(1-\eta)\,.$$

Finally, we try an alternative step-size setting $\lambda_n = \overline{\lambda}_{\max}$ with $C_5 = 0.3C_2$ in (34) . In our first experiment with noisy data, we investigate the influence of different spaces in the reconstructions and therefore, compare again the spaces $X = V_p$ for $p = 2$ and $p = 1.01$. The results are exhibited in Figure 4, where each row represents a different sought-for conductivity. The first column displays the exact solutions γ^+, while the second and the third one show the reconstructions for $p = 2$ and $p = 1.01$ respectively. After the third column, a colorbar related to the reconstructions is displayed on each

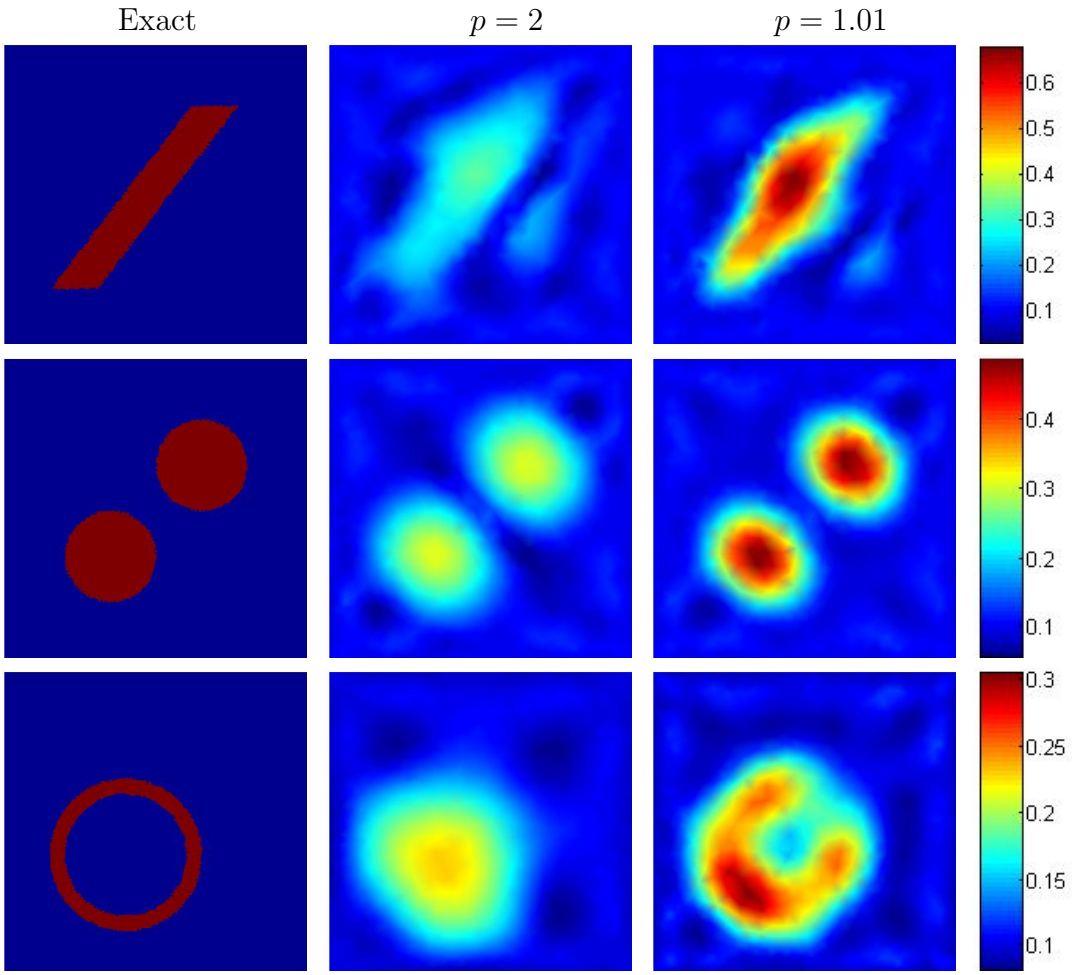


FIGURE 4. Comparison of reconstructions in different spaces and with the same noise level $\delta = 0.1\%$. First column: sought-for conductivities. Second column: reconstructions in Hilbert space V_2. Third column: reconstructions in Banach space $V_{1.01}$.

row (the exact solutions are in a different color scale). In all cases, the inclusions were located, but a clear improvement is perceived when the Hilbert space V_2 is changed by the more suitable Banach space $V_{1.01}$. Observe that the values of the conductivities are much closer to the real ones if the space $X = V_{1.01}$ is considered. Moreover, the shapes of the inclusions are more realistic in this space (see the last row).

In the last experiment, we test the efficiency of Algorithm 1 when dealing with *impulsive noise*. In contrast with uniformly distributed noise, this kind of perturbation has a very sparse distribution, and we can take advantage of it equipping the data-space Y with a different norm. In Figure 5 we present side by side, two different kinds of noise used in our experiments. The picture on the left shows uniformly distributed noise and the one on the right exhibits an example of impulsive noise, with uniform noise superimposed by some *outliers*[12], present in about 5% of the points. Both noises type are scaled in order to have the same L^2-norm.

[12]Outliers are highly inconsistent data points and may arise from procedural measurement errors.

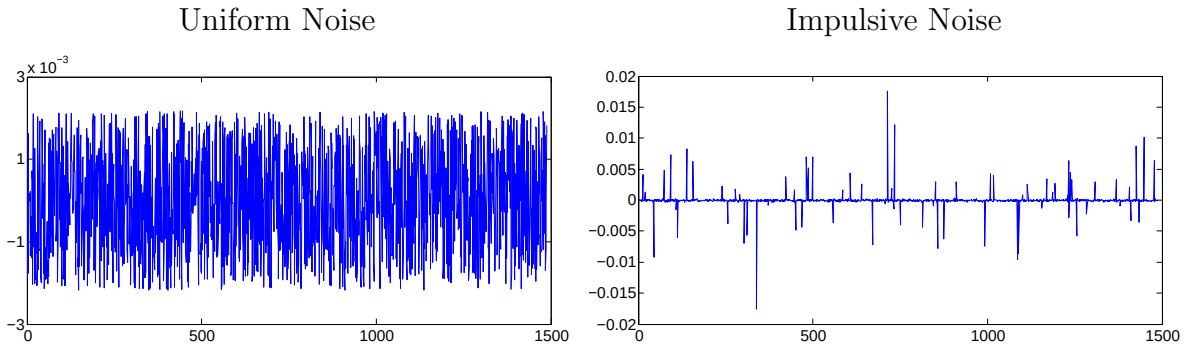


FIGURE 5. Different kinds of noise. Left: uniformly distributed noise. Right: impulsive noise. Both noises have the same L^2-norm.

Although the L^2-norm is the same for both noises shown in Figure 5, the $L^{1.01}-$norm of the second kind of noise amounts to only 25% of the first one, which means that in this norm, the data corrupted by impulsive noise furnishes much more information and is more precise than the one contamined with uniform noise. For this reason, we change the norm in data-space Y to $(L^r(\partial\Omega))^d$ for some $r > 1$ but close to 1 and consider the new operator

$$F_G \colon V_+ \subset V_p \to (L^r(\partial\Omega))^d, \quad 1 < r \leq 2.$$

As $\|w\|_{(L^r(\partial\Omega))^d} \lesssim \|w\|_{(L^2(\partial\Omega))^d}$ for every $w \in Y$, the new operator F_G is still F-differentiable and its derivative remains the same. To perform our experiment under impulsive noise, we have chosen the exact solution γ^+ with a four-squares inclusion, displayed in the middle of Figure 1. The noises have been scaled such that the resulting relative L^2-noise in (48) is $\delta = 0.3\%$ in both cases. The results are illustrated in Figure 6, where the first row shows the reconstructions when uniform noise is considered and the second one displays the results when the data is contamined with impulsive noise. The first and second column represent the Hilbert space $Y = (L^2(\partial\Omega))^d$ and the Banach space $Y = (L^{1.01}(\partial\Omega))^d$ respectively. Below each picture, the number of performed iterations until termination, $N = N(\delta)$, and the relative error in the L^2-norm, E_N^2, are highlighted. As is promptly seen, the reconstructions do not show a relevant variation when uniform noise is considered, but under the effect of impulsive, the space $Y = (L^{1.01}(\partial\Omega))^d$ provides a much better reconstruction, showing a smaller L^2-error and exhibiting inclusions with more precise locations, values and shapes. However, a disadvantage becomes evident: a much larger number of iterations needs to be performed in this situation.

REFERENCES

[1] Kari Astala and Lassi Päivärinta. Calderón's inverse conductivity problem in the plane. *Ann. of Math. (2)*, 163(1):265–299, 2006.

[2] A. B. Bakushinskiĭ. On a convergence problem of the iterative-regularized Gauss-Newton method. *Zh. Vychisl. Mat. i Mat. Fiz.*, 32(9):1503–1509, 1992.

[3] Thomas Bonesky, Kamil S. Kazimierski, Peter Maass, Frank Schöpfer, and Thomas Schuster. Minimization of Tikhonov functionals in Banach spaces. *Abstr. Appl. Anal.*, pages Art. ID 192679, 19, 2008.

[4] Liliana Borcea. Electrical impedance tomography. *Inverse Problems*, 18(6):R99–R136, 2002.

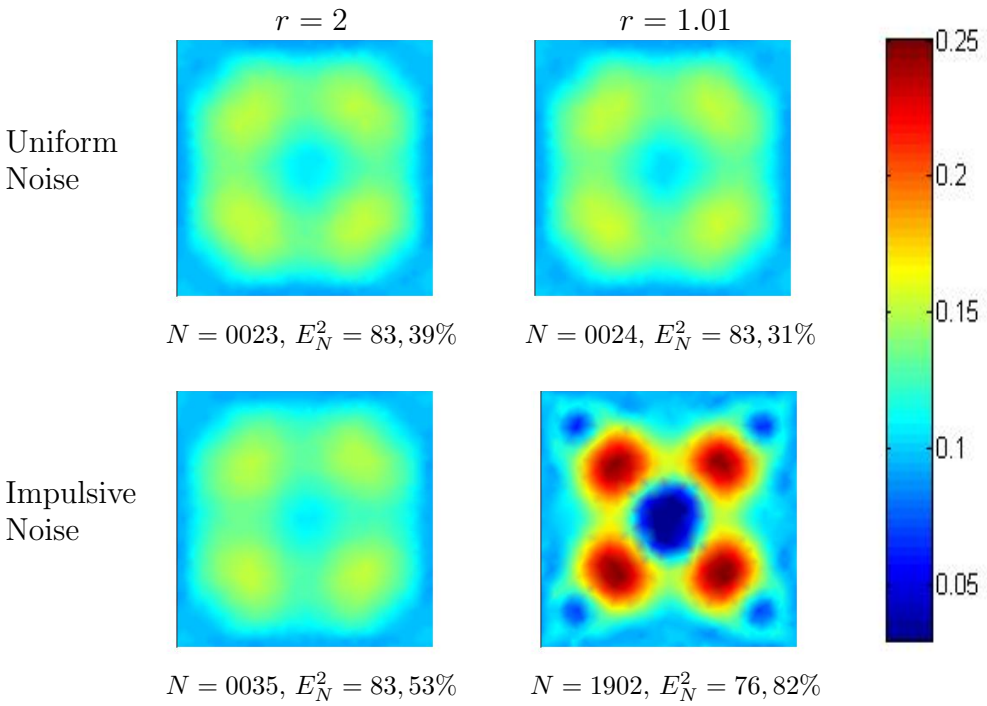


FIGURE 6. Reconstructed conductivities obtained under the use of uniform and impulsive noise. The columns represent different spaces and the rows, different types of noise.

[5] Alberto-P. Calderón. On an inverse boundary value problem. In *Seminar on Numerical Analysis and its Applications to Continuum Physics (Rio de Janeiro, 1980)*, pages 65–73. Soc. Brasil. Mat., Rio de Janeiro, 1980.

[6] Margaret Cheney, David Isaacson, and Jonathan C. Newell. Electrical impedance tomography. *SIAM Rev.*, 41(1):85–101 (electronic), 1999.

[7] Charles Chidume. *Geometric properties of Banach spaces and nonlinear iterations*, volume 1965 of *Lecture Notes in Mathematics*. Springer-Verlag London, Ltd., London, 2009.

[8] Ioana Cioranescu. *Geometry of Banach spaces, duality mappings and nonlinear problems*, volume 62 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1990.

[9] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57(11):1413–1457, 2004.

[10] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[11] Bangti Jin, Taufiquar Khan, and Peter Maass. A reconstruction algorithm for electrical impedance tomography based on sparsity regularization. *Internat. J. Numer. Methods Engrg.*, 89(3):337–353, 2012.

[12] Bangti Jin and Peter Maass. Sparsity regularization for parameter identification problems. *Inverse Problems*, 28(12):123001, 70, 2012.

[13] Qinian Jin. Inexact Newton-Landweber iteration for solving nonlinear inverse problems in Banach spaces. *Inverse Problems*, 28(6):065002, 15, 2012.

[14] Barbara Kaltenbacher, Andreas Neubauer, and Otmar Scherzer. *Iterative regularization methods for nonlinear ill-posed problems*, volume 6 of *Radon Series on Computational and Applied Mathematics*. Walter de Gruyter GmbH & Co. KG, Berlin, 2008.

[15] Barbara Kaltenbacher and Ivan Tomba. Convergence rates for an iteratively regularized Newton-Landweber iteration in Banach space. *Inverse Problems*, 29(2):025010, 18, 2013.

[16] Armin Lechleiter and Andreas Rieder. Newton regularizations for impedance tomography: convergence by local injectivity. *Inverse Problems*, 24(6):065009, 18, 2008.
[17] Armin Lechleiter and Andreas Rieder. Towards a general convergence theory for inexact Newton regularizations. *Numer. Math.*, 114(3):521–548, 2010.
[18] Fábio Margotti. *On inexact Newton methods for inverse problems in Banach spaces*. PhD Thesis. Karlsruher Institut für Technologie, Karlsruhe, 2015.
[19] Fábio Margotti and Andreas Rieder. An inexact Newton regularization in Banach spaces based on the nonstationary iterated Tikhonov method. *Journal of inverse and Ill-posed Problems*, 23(4):373–392, 2014.
[20] A. Neubauer and O. Scherzer. A convergence rate result for a steepest descent method and a minimal error method for the solution of nonlinear ill-posed problems. *Z. Anal. Anwendungen*, 14(2):369–377, 1995.
[21] F. Schöpfer, A. K. Louis, and T. Schuster. Nonlinear iterative methods for linear ill-posed problems in Banach spaces. *Inverse Problems*, 22(1):311–329, 2006.
[22] Thomas Schuster, Barbara Kaltenbacher, Bernd Hofmann, and Kamil S. Kazimierski. *Regularization methods in Banach spaces*, volume 10 of *Radon Series on Computational and Applied Mathematics*. Walter de Gruyter GmbH & Co. KG, Berlin, 2012.
[23] A. N. Tikhonov. On the solution of incorrectly put problems and the regularisation method. In *Outlines Joint Sympos. Partial Differential Equations (Novosibirsk, 1963)*, pages 261–265. Acad. Sci. USSR Siberian Branch, Moscow, 1963.
[24] Robert Winkler and Andreas Rieder. Model-aware newton-type inversion scheme for electrical impedance tomography. *Inverse Problems*, 31(4):045009, 2015.
[25] Zong Ben Xu and G. F. Roach. Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces. *J. Math. Anal. Appl.*, 157(1):189–210, 1991.

Department of Mathematics, Federal University of Santa Catarina
E-mail address: fabiomarg@gmail.com