Taylor & Francis
Taylor & Francis Group

Check for updates

On the choice of Lagrange multipliers in the iterated Tikhonov method for linear ill-posed equations in Banach spaces

M. P. Machado[a], F. Margotti[b] and A. Leitão[b]

[a]Department of Mathematics, Federal University of Bahia, Salvador, Brazil; [b]Department of Mathematics, Federal University of St. Catarina, Florianópolis, Brazil

ABSTRACT

This article is devoted to the study of *nonstationary Iterated Tikhonov* (nIT) type methods (Hanke M, Groetsch CW. Nonstationary iterated Tikhonov regularization. J Optim Theory Appl. 1998;98(1):37–53; Engl HW, Hanke M, Neubauer A. Regularization of inverse problems. Vol. 375, Mathematics and its Applications. Dordrecht: Kluwer Academic Publishers Group; 1996. MR 1408680) for obtaining stable approximations to linear ill-posed problems modelled by operators mapping between Banach spaces. Here we propose and analyse an *a posteriori* strategy for choosing the sequence of regularization parameters for the nIT method, aiming to obtain a predefined decay rate of the residual. Convergence analysis of the proposed nIT type method is provided (convergence, stability and semi-convergence results). Moreover, in order to test the method's efficiency, numerical experiments for three distinct applications are conducted: (i) a 1D convolution problem (smooth Tikhonov functional and Banach parameter-space); (ii) a 2D deblurring problem (nonsmooth Tikhonov functional and Hilbert parameter-space); (iii) a 2D elliptic inverse potential problem.

1. Introduction

In this article we investigate *nonstationary Iterated Tikhonov* (nIT) type methods [1,2] for obtaining stable approximations of linear ill-posed problems modelled by operators mapping between Banach spaces. The novelty of our approach consists in the introduction of an *a posteriori* strategy for choosing the sequence of regularization parameters (or, equivalently, the Lagrange multipliers) for the nIT iteration, which play a key role in the convergence speed of the nIT iteration.

This new *a posteriori* strategy aims to enforce a pre-defined decay of the residual in each iteration; it differs from the classical choice for the Lagrange multipliers (see, e.g. [2,3]), which is based on an *a priori* strategy (typically geometrical) and leads to an unknown decay rate of the residual.

The *inverse problem* we are interested in consists of determining an unknown quantity $x \in X$ from given data $y \in Y$, where X, Y are Banach spaces. We assume that data are

obtained by indirect measurements of the parameter, this process being described by the ill-posed operator equation

$$Ax = y, \tag{1}$$

where $A : X \to Y$ is a bounded linear operator, whose inverse $A^{-1} : R(A) \to X$ either does not exist, or is not continuous. In practical situations, one does not know the data y exactly; instead, only approximate measured data $y^\delta \in Y$ are available with

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

where $\delta > 0$ is the (known) noise level. A comprehensive study of linear ill-posed problems in Banach spaces can be found in the text book [4] (see, e.g. [1] for a corresponding theory in Hilbert spaces).

Iterated Tikhonov type methods are typically used for linear inverse problems. In the Hilbert space setting we refer the reader to [2] for linear operator equations, and also to [5] for the nonlinear case. In the Banach space setting the research is still ongoing. Some preliminary results can be found in [3] for linear operator equations; see [6] for the nonlinear case; see also [7]. In all references above, *a priori* strategies are used for choosing the Lagrange multipliers.

1.1. *Main results: presentation and interpretation*

The approach discussed in this manuscript is devoted to the Banach space setting, and consists in adopting an *a posteriori strategy* for the choice of the Lagrange multipliers. The strategy used here is inspired by the recent work [8], where the authors propose an endogenous strategy for the choice of the Lagrange multipliers in the nonstationary iterated Tikhonov method for solving (1), (2) when X and Y are Hilbert spaces. The penalty terms used in our Tikhonov functionals are the same as in [6] and consist of Bregman distances induced by (uniformly) convex functionals, e.g. the sum of the L^2-norm with the TV -seminorm. In our previous work [9], we implemented the method proposed in this paper and investigated its numerical performance (two different ill-posed problems were solved). Here, we extend our previous results by presenting a whole convergence analysis. Additionally, the corresponding algorithm is implemented for solving three benchmark problems and more details concerning the computation of the minimizers of the Tikhonov functional are provided; our numerical results are compared with the ones obtained using the nIT method using the classical geometrical choice of Lagrange multipliers [6].

In what follows, we briefly interpret the main results: The proposed method defines each Lagrange multiplier such that the residual of the corresponding next iterate lies in an interval which depends on both the noise level and the residual of the current iterate (see (23)). This fact has the following consequences: (1) it forces a geometrical decay of the residual (Proposition 4.1); (2) it guarantees the possibility of computing multipliers in agreement with the theoretical convergence results (Theorems 4.6, 4.8 and 4.9); (3) the computation of the multipliers demands less numerical effort than the classical strategy of computing the multipliers by solving an equation in each iteration; (4) the next iterate is not uniquely determined by the current one; instead, it is chosen within a set of successors of the current iterate (Definition 4.7). We also address the actual computation of the Lagrange multipliers.

Since each multiplier is implicitly defined by an inequality, we discuss a numerically effi-cient strategy for computing them, which is based on the decrease rate of the past residuals (Section 5.1).

1.2. Outline of the article

This manuscript is outlined as follows: In Section 2 a revision of relevant background mate-rial is presented. In Section 3 the new nIT method is introduced. Section 4 is devoted to the convergence analysis of the nIT method. In Section 5 possible implementations of our method are discussed; the evaluation of the Lagrange multipliers is addressed, as well as the issue of minimizing the Tikhonov functionals. Section 6 is devoted to numerical experiments, while Section 7 is dedicated to final remarks and conclusions.

2. Background material

For details on the material discussed in this section, we refer the reader to the textbooks [4,10].

Unless the contrary is explicitly stated, we always consider X a *real* Banach space. The *effective domain* of the convex functional $f \colon X \to \overline{\mathbb{R}} := (-\infty, \infty]$ is defined as

$$\mathrm{Dom}\,(f) := \left\{ x \in X : f(x) < \infty \right\}.$$

The set $\mathrm{Dom}(f)$ is always convex and we call f *proper* provided it is non-empty. We call f *uniformly convex* if there exists a continuous and strictly increasing function $\varphi \colon \mathbb{R}_0^+ \to \mathbb{R}_0^+$ with the property $\varphi(t) = 0$ implies $t = 0$, such that

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

for all $\lambda \in (0,1)$ and $x, y \in X$. Of course f uniformly convex implies f strictly convex, which in turn implies f convex. The functional f is *lower semi-continuous* (in short l.s.c.) if for any sequence $(x_k)_{k \in \mathbb{N}} \subset X$ satisfying $x_k \to x$, it holds

$$f(x) \le \lim_{k \to \infty} \inf f(x_k)\,.$$

It is called *weakly lower semi-continuous* (w.l.s.c.) if above property holds true with $x_k \to x$ replaced by $x_k \rightharpoonup x$. Obviously, every w.l.s.c functional is l.s.c. Further, any Banach space norm is w.l.s.c.

The *sub-differential* of a functional $f \colon X \to \overline{\mathbb{R}}$ is the point-to-set mapping $\partial f \colon X \to 2^{X^*}$ defined by

$$\partial f(x) := \left\{ x^* \in X^* : f(x) + \langle x^*, y - x \rangle \le f(y) \text{ for all } y \in X \right\}.$$

Any element in the set $\partial f(x)$ is called a *sub-gradient* of f at x. The effective domain of ∂f is the set

$$\mathrm{Dom}\,(\partial f) := \left\{ x \in X : \partial f(x) \ne \emptyset \right\}.$$

It is clear that the inclusion $\mathrm{Dom}(\partial f) \subset \mathrm{Dom}(f)$ holds whenever f is proper.

Sub-differentiable functionals and l.s.c. convex functionals are very close related concepts. In fact, a sub-differentiable functional f is convex and l.s.c. in any open convex set of $\mathrm{Dom}(f)$. On the other hand, a proper, convex and l.s.c. functional is always sub-differentiable on its effective domain.

The definition of sub-differential readily yields

$$0 \in \partial f\,(x) \Longleftrightarrow f\,(x) \leq f\,(y) \quad \text{for all } y \in X.$$

If $f, g\colon X \to \overline{\mathbb{R}}$ are convex functionals and there is a point $x \in \mathrm{Dom}(f) \cap \mathrm{Dom}(g)$ where f is continuous, then

$$\partial\,(f + g)\,(x) = \partial f\,(x) + \partial g\,(x) \quad \text{for all } x \in X. \tag{4}$$

Moreover, if Y is a real Banach space, $h\colon Y \to \overline{\mathbb{R}}$ is a convex functional, $b \in Y$, $A\colon X \to Y$ is a bounded linear operator and h is continuous at some point of the range of A, then

$$\partial\,(h\,(\cdot - b))\,(y) = (\partial h)\,(y - b) \quad \text{and } \partial\,(h \circ A)\,(x) = A^*\,(\partial h\,(Ax)),$$

for all $x \in X$ and $y \in Y$, where $A^*\colon Y^* \to X^*$ is the Banach-adjoint of A. Consequently,

$$\partial\,(h\,(A \cdot - b))\,(x) = A^*\,(\partial h)\,(Ax - b) \quad \text{for all } x \in X. \tag{5}$$

If a convex functional $f\colon X \to \overline{\mathbb{R}}$ is Gâteaux-differentiable at $x \in X$, then f has a unique sub-gradient at x, namely, the Gâteaux-derivative itself: $\partial f(x) = \{\nabla f(x)\}$.

The sub-differential of the convex functional

$$f\,(x) = \frac{1}{p}\,\|x\|^p, \quad p > 1, \tag{6}$$

is called the *duality mapping* and is denoted by J_p. It can be shown that for all $x \in X$,

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

Thus, the duality mapping has the inner-product-like properties:

$$\langle x^*, y \rangle \leq \|x\|^{p-1}\,\|y\| \quad \text{and} \quad \langle x^*, x \rangle = \|x\|^p,$$

for all $x^* \in J_p(x)$. By using the Riesz Representation Theorem, one can prove that $J_2(x) = x$ for all $x \in X$ whenever X is a Hilbert space.

Banach spaces are classified according with their geometrical characteristics. Many concepts concerning these characteristics are usually defined using the so called *modulus of convexity* and *modulus of smoothness*, but most of these definitions can be equivalently stated observing the properties of the functional f defined in (6).[1] This functional is convex and sub-differentiable in any Banach space X. If (6) is Gâteaux-differentiable in the whole space X, this Banach space is called *smooth*. In this case, $J_p(x) = \partial f(x) = \{\nabla f(x)\}$ and therefore, the duality mapping $J_p\colon X \to X^*$ is single-valued. If the functional f in (6)

is Fré chet-differentiable in X, this space is called *locally uniformly smooth* and it is called *uniformly smooth* provided f is uniformly Fréchet-differentiable in bounded sets. As a result, the duality mapping is continuous (resp. uniformly continuous in bounded sets) in locally uniformly smooth (resp. uniformly smooth) spaces. It is immediate that uniform smoothness of a Banach space implies local uniform smoothness, which in turn implies smoothness of this space. Moreover, none reciprocal is true. Similarly, a Banach space X is called *strictly convex* whenever (6) is a strictly convex functional. Moreover, X is called *uniformly convex* if the functional f in (6) is uniformly convex. It is clear that uniform convexity implies strict convexity. It is well-known that both uniformly smooth and uniformly convex Banach spaces are reflexive.

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

$$D_\xi f(x, y) := f(x) - f(y) - \langle \xi, x - y \rangle.$$

Obviously, $D_\xi f(y, y) = 0$ and, since $\xi \in \partial f(y)$, it additionally holds that $D_\xi f(x, y) \geq 0$. Moreover, it is straightforward proving the *Three Points Identity*:

$$D_{\xi_1} f(x_2, x_1) - D_{\xi_1} f(x_3, x_1) = D_{\xi_3} f(x_2, x_3) + \langle \xi_3 - \xi_1, x_2 - x_3 \rangle,$$

for all $x_2 \in X$, $x_1, x_3 \in \text{Dom}(\partial f)$, $\xi_1 \in \partial f(x_1)$ and $\xi_3 \in \partial f(x_3)$. Further, the functional $D_\xi f(\cdot, y)$ is strictly convex whenever f is strictly convex, and in this case, $D_\xi f(x, y) = 0$ iff $x = y$.

When f is the functional defined in (6) and X is a smooth Banach space, the Bregman distance has the special notation $\Delta_p(x, y)$, i.e.

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

Since J_2 is the identity operator in Hilbert spaces, a simple application of the polarization identity shows that $\Delta_2(x, y) = \frac{1}{2} \|x - y\|^2$ in these spaces.

It is not difficult to prove (see e.g. [9]) that if $f : X \to \overline{\mathbb{R}}$ is uniformly convex, then

$$\varphi\left(\|x - y\|\right) \leq D_\xi f(x, y) \tag{7}$$

for all $x \in X$, $y \in \text{Dom}(\partial f)$ and $\xi \in \partial f(y)$, where φ is the function in (3). In particular, in a smooth and uniformly convex Banach space X, the above inequality reads $\varphi(\|x - y\|) \leq \Delta_p(x, y)$.

We say that a functional $f : X \to \overline{\mathbb{R}}$ has the *Kadec property* if for any sequence $(x_k)_{k \in \mathbb{N}} \subset X$, the weak convergence $x_k \rightharpoonup x$, together with $f(x_k) \to f(x) < \infty$, implies $x_k \to x$. It is not difficult to prove (see e.g. [6]) that any proper, w.l.s.c. and uniformly convex functional has the Kadec property. In particular, the norm in a uniformly convex Banach space has this property.

Concrete examples of Banach spaces of interest are the Lebesgue space $L^p(\Omega)$, the Sobolev space $W^{n,p}(\Omega)$, $n \in \mathbb{N}$, and the space of $p-$summable sequences $\ell^p(\mathbb{R})$. All

these Banach spaces are both uniformly convex and uniformly smooth provided that $1 < p < \infty$.

3. The nIT method

In this section, we present the nonstationary iterated Tikhonov (nIT) type method considered in this article, which aims to find stable approximate solutions to the inverse problem (1), (2). The method proposed here is in the spirit of the method in [6]. The distinguishing feature is the use of an *a posteriori* strategy for the choice of the Lagrange multipliers, as detailed below.

For a fixed $r > 1$ and a uniformly convex penalty term f, the nIT method defines sequences $(x_k^\delta)_{k \in \mathbb{N}}$ in X and $(\xi_k^\delta)_{k \in \mathbb{N}}$ in X^* iteratively by

$$x_k^\delta := \arg\min_{x \in X} \lambda_k^\delta r^{-1} \left\| Ax - y^\delta \right\|^r + D_{\xi_{k-1}^\delta} f\left(x, x_{k-1}^\delta\right),$$

$$\xi_k^\delta := \xi_{k-1}^\delta - \lambda_k^\delta A^* J_r(Ax_k^\delta - y^\delta),$$

where the Lagrange multiplier $\lambda_k^\delta > 0$ is to be determined using only information about A, δ, y^δ and x_{k-1}^δ.

Our strategy for choosing the Lagrange multipliers is inspired by the work [8], where the authors propose an endogenous strategy for the choice of the Lagrange multipliers in the nonsationary iterated Tikhonov method for solving (1), (2) when X and Y are Hilbert spaces. This method is based on successive orthogonal projection methods onto a family of shrinking, separating convex sets. Specifically, the iterative method in [8] obtains the new iterate projecting the current one onto a levelset of the residual function, whose level belongs to a range defined by the current residual and by the noise level. Moreover, the admissible Lagrange multipliers (in each iteration) shall be chosen in a non-degenerate interval.

Aiming to extend this framework to the Banach space setting, we are forced to introduce Bregman distance and *Bregman projections*. This is due to the well-known fact that in Banach spaces the *metric projection* onto a convex and closed set C, defined as $P_C(x) = arg\,min_{z \in C} \|z - x\|^2$, loses the decreasing distance property of the orthogonal projection in Hilbert spaces. In order to recover this property, one should minimize in Banach spaces the Bregman distance, instead of the norm-induced distance.

For the remaining of this article we adopt the following main assumptions:

(A.1) There exists an element $x^\star \in X$ such that $Ax^\star = y$, where $y \in R(A)$ is the exact data.
(A.2) f is a w.l.s.c. function.
(A.3) f is a uniformly convex function.
(A.4) X and Y are reflexive Banach spaces and Y is smooth.

Moreover, we denote by $\Omega_\mu^r \subset X$ the μ-levelset of the residual functional $\|Ax - y^\delta\|$, i.e.

$$\Omega_\mu^r := \left\{ x \in X : r^{-1} \|Ax - y^\delta\|^r \leq r^{-1} \mu^r \right\}.$$

Note that, since A is a continuous linear operator, it follows that Ω_μ^r is closed and convex. Now, given $\hat{x} \in \text{Dom}(\partial f)$ and $\xi \in \partial f(\hat{x})$, we define the *Bregman projection* of $\hat{x}$ onto Ω_μ^r,

as a solution of the minimization problem

$$\begin{cases} \min\ D_\xi f(x, \hat{x}) \\ \text{s.t. } r^{-1}\|Ax - y^\delta\|^r \le r^{-1}\mu^r. \end{cases} \tag{8}$$

It is worth noticing that a solution of the problem (8) depends on the sub-gradient ξ. Furthermore, since $D_\xi f(\cdot, \hat{x})$ is strictly convex (which follows from the uniformly convexity of f), problem (8) has at most one solution. The fact that the Bregman projection is well defined when $\mu > \delta$ (in this case we set $P^f_{\Omega^r_\mu}(\hat{x}) := \arg\min_{x \in \Omega^r_\mu} D_\xi f(x, \hat{x})$) is a consequence of the following lemma.

Lemma 3.1: *If $\mu > \delta$, then problem (8) has a solution.*

Proof: Assumption (A.1), together with equation (2) and the assumption that $\mu > \delta$, implies that the feasible set of problem (8), i.e. the set Ω^r_μ, is nonempty.

From Assumptions (A.2) and (A.3) it follows that $D_\xi f(\cdot, \hat{x})$ is proper, convex and l.s.c. Furthermore, relation (7) implies that $D_\xi f(\cdot, \hat{x})$ is a coercive function. Hence, the lemma follows using the reflexivity of X together with [11, Corollary 3.23].[2] ∎

Note that if $0 \le \mu' \le \mu$, then $\Omega^r_{\mu'} \subseteq \Omega^r_\mu$ and $A^{-1}(y) \subset \Omega^r_\mu$ for all $\mu \ge \delta$. Furthermore, with the available information of the solution set of (1), Ω^r_δ is the set of best possible approximate solution for this inverse problem. However, since problem (8) may be ill-posed when $\mu = \delta$, our best choice is to generate x^δ_k from $x^\delta_{k-1} \notin \Omega^r_\delta$ as a solution of problem (8), with $\hat{x} = x^\delta_{k-1}$ and $\mu = \mu_k$ such that we guarantee a reduction of the residual norm while preventing ill-posedness of (8).

For this purpose, we analyse in the sequel the minimization problem (8) by means of Lagrange multipliers. The Lagrangian function associated to problem (8) is

$$\mathcal{L}(x, \lambda) = \frac{\lambda}{r}(\|Ax - y^\delta\|^r - \mu^r) + D_\xi f(x, \hat{x}).$$

Note that, for each $\lambda > 0$, the function $\mathcal{L}(\cdot, \lambda): X \to \overline{\mathbb{R}}$ is l.s.c. and convex. For any $\lambda > 0$ define the functions

$$\pi(\hat{x}, \lambda) := \arg\min_{x \in X} \mathcal{L}(x, \lambda), \quad G_{\hat{x}}(\lambda) := \|A\pi(\hat{x}, \lambda) - y^\delta\|^r. \tag{9}$$

The next lemma provides a classical Lagrange multiplier result for problem (8), which will be useful for formulating the nIT method.

Lemma 3.2: *If $\|A\hat{x} - y^\delta\| > \mu > \delta$, then the following assertions are equivalent*

1. *x is a solution of (8);*
2. *there exists $\lambda^* > 0$ satisfying $x = \pi(\hat{x}, \lambda^*)$ and $G_{\hat{x}}(\lambda^*) = \mu^r$.*

Proof: It follows from (2), Assumption (A.1) and the hypothesis $\mu > \delta$ that $x^\star \in X$ satisfies

$$\|Ax^\star - y^\delta\|^r < \mu^r.$$

This inequality implies the Slater condition for problem (8). Thus, since A is continuous and $D_\xi f(\cdot, \hat{x})$ is l.s.c., we conclude that x is a solution of (8) if and only if there exists $\lambda \in$

$\mathbb{R}$ such that the point (x, λ) satisfies the *Karush–Kuhn–Tucker* (KKT) conditions for this minimization problem [12], namely

$$\lambda \geq 0, \quad G_{\hat{x}}(\lambda) \leq \mu^r, \quad \lambda(G_{\hat{x}}(\lambda) - \mu^r) = 0, \qquad 0 \in \partial_x \mathcal{L}(x, \lambda).$$

If we assume $\lambda = 0$ in the relations above, then the definition of the Lagrangian function, together with the strictly convexity of $D_\xi f(\cdot, \hat{x})$, implies that $\hat{x}$ is the unique minimizer of $\mathcal{L}(\cdot, 0)$. Moreover, since $\|A\hat{x} - y^\delta\| > \mu$, we conclude that the pair $(\hat{x}, 0)$ does not satisfy the KKT conditions. Consequently, we have $\lambda > 0$ and $G_{\hat{x}}(\lambda) - \mu^r = 0$. The lemma follows using the definition of $\pi(\hat{x}, \lambda)$. ∎

We are ready to present the nIT method for solving (1).

---

[1] choose an initial guess $x_0 \in X$ and $\xi_0 \in \partial f(x_0)$;

[2] choose $\eta \in (0, 1)$, $\tau > 1$ and set $k := 0$;

[3] while $\left(\|Ax_k^\delta - y^\delta\| > \tau\delta\right)$ do

  [3.1] $k := k + 1$;

  [3.2] compute λ_k^δ and x_k^δ such that

$$x_k^\delta = \arg\min_{x \in X} \lambda_k^\delta r^{-1} \|Ax - y^\delta\|^r + D_{\xi_{k-1}^\delta} f(x, x_{k-1}^\delta) \quad \text{and}$$

$$\delta^r < G_{x_{k-1}^\delta}(\lambda_k^\delta) \leq \left(\eta\delta + (1 - \eta)\|Ax_{k-1}^\delta - y^\delta\|\right)^r;$$

  [3.3] set $\xi_k^\delta := \xi_{k-1}^\delta - \lambda_k^\delta A^* J_r(Ax_k^\delta - y^\delta)$.

Algorithm 1: The iterative method.

Properties (4) and (5), together with the definition of the duality mapping J_r, imply that the point $x_k^\delta \in X$ minimizes the optimization problem in [3.2] if and only if

$$0 \in \lambda_k^\delta A^* J_r(Ax_k^\delta - y^\delta) + \partial f(x_k^\delta) - \xi_{k-1}^\delta. \tag{10}$$

Hence, since Y is a smooth Banach space, the duality mapping J_r is single valued and

$$\xi_{k-1}^\delta - \lambda_k^\delta A^* J_r(Ax_k^\delta - y^\delta) \in \partial f(x_k^\delta).$$

Consequently, ξ_k^δ in step 3.2 of Algorithm 1 is well defined and it is a sub-gradient of f at x_k^δ.

Notice that the stopping criteria in Algorithm 1 corresponds to the discrepancy principle, i.e. the iteration stops at step $k(\delta)$ defined by

$$k(\delta) := \min\{k \geq 1; \|Ax_j^\delta - y^\delta\| > \tau\delta, \ j = 0, \ldots, k - 1 \text{ and } \|Ax_k^\delta - y^\delta\| \leq \tau\delta\}. \tag{11}$$

Remark 3.3 (Novel properties of the proposed method):

— The strategy used here is inspired by the recent work [8], where the authors propose an endogenous strategy for the choice of the Lagrange multipliers in the nonstationary iterated Tikhonov method for solving (1), (2) when X and Y are Hilbert spaces.

— The penalty terms used in our Tikhonov functionals are the same as in [6] and consist of Bregman distances induced by (uniformly) convex functionals, e.g, the sum of the L^2-norm with the *TV* -seminorm.
— We present a whole convergence analysis for the proposed method, characterizing it as a regularization metod.

4. Convergence analysis

In this section, we analyse the convergence properties of Algorithm 1. We begin by presenting the following result that establishes an estimate for the decay of the residual $\|Ax_k^\delta - y^\delta\|$. It can be proved in much the same manner as [8, Proposition 4.1], and for the sake of brevity we omit the proof here.

Proposition 4.1: *Let* $(x_k^\delta)_{0 \leq k \leq k(\delta)}$ *be the (finite) sequence defined by the nIT method (Algorithm 1), with* $\delta \geq 0$ *and* $y^\delta \in Y$ *as in (2). Then,*

$$\left[\|Ax_k^\delta - y^\delta\| - \delta \right] \leq \eta \left[\|Ax_{k-1}^\delta - y^\delta\| - \delta \right] \leq \eta^k \left[\|Ax_0 - y^\delta\| - \delta \right], \quad k = 1, \ldots, k(\delta),$$

where $k(\delta) \in \mathbb{N}$ *is defined by (11).*

As a direct consequence of Proposition 4.1 we have that in the noisy data case, the discrepancy principle terminates the iteration after finitely many steps, i.e. $k(\delta) < \infty$. Furthermore, the corollary below gives an estimate for the stopping index $k(\delta)$.

Corollary 4.2: *Let* $(x_k^\delta)_{0 \leq k \leq k(\delta)}$ *be the (finite) sequence defined by the nIT method (Algorithm 1), with* $\delta > 0$ *and* $y^\delta \in Y$ *as in (2). Then, the stopping index* $k(\delta)$, *defined in (11), satisfies*

$$k(\delta) \leq |\ln \eta|^{-1} \ln \left[\frac{\|Ax_0 - y^\delta\| - \delta}{(\tau - 1)\delta} \right] + 1.$$

In the next proposition we prove monotonicity of the sequence $(D_{\xi_k^\delta} f(x^\star, x_k^\delta))_{k \in \mathbb{N}}$ and we also estimate the gain $D_{\xi_{k-1}^\delta} f(x^\star, x_{k-1}^\delta) - D_{\xi_k^\delta} f(x^\star, x_k^\delta)$, where $x^\star \in X$ satisfies Assumption (A.1).

Proposition 4.3: *Let* $(x_k^\delta)_{0 \leq k \leq k(\delta)}$ *be the (finite) sequence defined by the nIT method (Algorithm 1), with* $\delta \geq 0$ *and* $y^\delta \in Y$ *as in (2). Then, for every* $x^\star$ *satisfying Assumption (A.1) it holds*

$$D_{\xi_k^\delta} f(x^\star, x_k^\delta) \leq D_{\xi_{k-1}^\delta} f(x^\star, x_{k-1}^\delta) - D_{\xi_{k-1}^\delta} f(x_k^\delta, x_{k-1}^\delta) - \lambda_k^\delta \left(1 - \frac{1}{\tau} \right) \|Ax_k^\delta - y^\delta\|^r, \quad (12)$$

for $k = 1, \ldots, k(\delta) - 1$.

Proof: Using the three points identity we have

$$D_{\xi_k^\delta} f(x^\star, x_k^\delta) - D_{\xi_{k-1}^\delta} f(x^\star, x_{k-1}^\delta) = -D_{\xi_{k-1}^\delta} f(x_k^\delta, x_{k-1}^\delta) + \langle \xi_k^\delta - \xi_{k-1}^\delta, x_k^\delta - x^\star \rangle$$
$$= -D_{\xi_{k-1}^\delta} f(x_k^\delta, x_{k-1}^\delta) - \lambda_k^\delta \langle A^* J_r(Ax_k^\delta - y^\delta), x_k^\delta - x^\star \rangle,$$

$$(13)$$

where the second equality above follows from the definition of ξ_k^δ. Simple manipulations of the second term on the right hand side above yield

$$\lambda_k^\delta \langle A^* J_r(Ax_k^\delta - y^\delta), x_k^\delta - x^\star \rangle = \lambda_k^\delta \langle J_r(Ax_k^\delta - y^\delta), Ax_k^\delta - y^\delta \rangle + \lambda_k^\delta \langle J_r(Ax_k^\delta - y^\delta), y^\delta - y \rangle$$
$$= \lambda_k^\delta \|Ax_k^\delta - y^\delta\|^r + \lambda_k^\delta \langle J_r(Ax_k^\delta - y^\delta), y^\delta - y \rangle.$$

$$(14)$$

Combining these two relations we obtain

$$D_{\xi_k^\delta} f(x^\star, x_k^\delta) - D_{\xi_{k-1}^\delta} f(x^\star, x_{k-1}^\delta) \le -D_{\xi_{k-1}^\delta} f(x_k^\delta, x_{k-1}^\delta) - \lambda_k^\delta \|Ax_k^\delta - y^\delta\|^r$$
$$+ \lambda_k^\delta \|Ax_k^\delta - y^\delta\|^{r-1} \|y^\delta - y\|$$
$$\le -D_{\xi_{k-1}^\delta} f(x_k^\delta, x_{k-1}^\delta) - \lambda_k^\delta \|Ax_k^\delta - y^\delta\|^r$$
$$+ \lambda_k^\delta \|Ax_k^\delta - y^\delta\|^{r-1} \delta,$$

where the last inequality follows from (2). Since $k \in \{1, \dots, k(\delta) - 1\}$, we have $\tau \delta \le \|Ax_k^\delta - y^\delta\|$. Thus,

$$D_{\xi_k^\delta} f(x^\star, x_k^\delta) - D_{\xi_{k-1}^\delta} f(x^\star, x_{k-1}^\delta) \le -D_{\xi_{k-1}^\delta} f(x_k^\delta, x_{k-1}^\delta) - \lambda_k^\delta \|Ax_k^\delta - y^\delta\|^r + \frac{\lambda_k^\delta}{\tau} \|Ax_k^\delta - y^\delta\|^r.$$

We deduce (12) from the above inequality. ∎

Corollary 4.4: *Let $(x_k)_{k \in \mathbb{N}}$ be the sequence defined by the nIT method (Algorithm 1) with $\delta = 0$ and $(\lambda_k)_{k \in \mathbb{N}}$ be the sequence of corresponding Lagrange multipliers. Then, for all $k = 1, 2, \dots$, and any $x^\star$ satisfying (A.1), we have*

$$D_{\xi_k} f(x^\star, x_k) = D_{\xi_{k-1}} f(x^\star, x_{k-1}) - D_{\xi_{k-1}} f(x_k, x_{k-1}) - \lambda_k \|Ax_k - y\|^r. \qquad (15)$$

Consequently, using a telescopic sum, we obtain

$$\sum_{k=1}^\infty \lambda_k \|Ax_k - y\|^r < \infty. \qquad (16)$$

Proof: In the exact data case, i.e. $\delta = 0$ and $y^\delta = y$, equality (14) becomes

$$\lambda_k \langle A^* J_r(Ax_k - y), x_k - x^\star \rangle = \lambda_k \|Ax_k - y\|^r.$$

Combining the formula above with (13), when $\delta = 0$, we deduce (15). The result in (16) follows directly from (15). ∎

We now fix a $x_0 \in \mathrm{Dom}\,(\partial f)$ and $\xi_0 \in \partial f(x_0)$, and study the existence and uniqueness of a vector $x^\dagger \in X$ with the property

$$D_{\xi_0} f \left(x^\dagger, x_0 \right) = \inf \left\{ D_{\xi_0} f \left(x, x_0 \right) : x \in \mathrm{Dom}(f) \text{ and } Ax = y \right\}. \tag{17}$$

Such an element $x^\dagger$ is called a x_0-*minimal-distance solution* and it is the equivalent of the x_0-minimal-norm solution of (1) in the Hilbert space setting [1].

Lemma 4.5: *There exists a unique element* $x^\dagger \in X$ *satisfying* (17).

Proof: Let $(x_k)_{k \in \mathbb{N}} \subset \mathrm{Dom}(f)$ be a sequence satisfying $Ax_k = y$ and

$$D_{\xi_0} f \left(x_k, x_0 \right) \to a := \inf \left\{ D_{\xi_0} f \left(x, x_0 \right) : x \in \mathrm{Dom}(f) \text{ and } Ax = y \right\},$$

then the sequence $(D_{\xi_0} f(x_k, x_0))_{k \in \mathbb{N}}$ is bounded, and because f is uniformly convex we have that $(x_k)_{k \in \mathbb{N}}$ is bounded as well. Since X is reflexive, there exist a vector $\bar{x} \in X$ and a subsequence $(x_{k_j})_{j \in \mathbb{N}}$ such that

$$x_{k_j} \rightharpoonup \bar{x}.$$

It follows that $A\bar{x} = y$ and because f is w.l.s.c. we have

$$D_{\xi_0} f \left(\bar{x}, x_0 \right) \leq \liminf_{j \to \infty} D_{\xi_0} f \left(x_{k_j}, x_0 \right) = \lim_{j \to \infty} D_{\xi_0} f \left(x_{k_j}, x_0 \right) = a,$$

which implies that $\bar{x}$ is a x_0-minimal-distance solution.

Suppose now that $x \neq z$ are two x_0-minimal-distance solutions. Since f is strictly convex, so is $D_{\xi_0} f(\cdot, x_0)$, and for any $\alpha \in (0, 1)$ we obtain

$$D_{\xi_0} f \left(\alpha x + (1 - \alpha) z, x_0 \right) < \alpha D_{\xi_0} f \left(x, x_0 \right) + (1 - \alpha) D_{\xi_0} f \left(z, x_0 \right) = D_{\xi_0} f \left(x, x_0 \right),$$

which contradicts the minimality of x. ∎

In the next theorem we prove strong convergence of the sequence generated by the nIT algorithm in the noise-free case to the solution $x^\dagger$.

Theorem 4.6: *Let* $(x_k)_{k \in \mathbb{N}}$ *be the sequence defined by the nIT method (Algorithm 1) with* $\delta = 0$ *and* $(\lambda_k)_{k \in \mathbb{N}}$ *the sequence of the corresponding Lagrange multipliers. Then,* $(x_k)_{k \in \mathbb{N}}$ *converges strongly to* $x^\dagger$.

Proof: We first prove that $(x_k)_{k \in \mathbb{N}}$ is a Cauchy sequence in X. Let $0 \leq l < m$ and x^* a solution of (1), using the three points identity we have

$$D_{\xi_l} f(x_m, x_l) - D_{\xi_l} f(x^*, x_l) = -D_{\xi_m} f(x^*, x_m) + \langle \xi_m - \xi_l, x_m - x^* \rangle. \tag{18}$$

We observe that

$$
\begin{aligned}
|\langle \xi_m - \xi_l, x_m - x^* \rangle| &= \left| \sum_{j=l+1}^{m} \langle \xi_j - \xi_{j-1}, x_m - x^* \rangle \right| \\
&= \left| \sum_{j=l+1}^{m} -\lambda_j \langle J_r(Ax_j - y), A(x_m - x^*) \rangle \right| \\
&\leq \sum_{j=l+1}^{m} \lambda_j \|Ax_j - y\|^{r-1} \|Ax_m - y\|,
\end{aligned}
\tag{19}
$$

where the second equality above follows form the definition of ξ_j and the inequality is a consequence of the properties of the duality mapping J_r. Proposition 4.1, with $\delta = 0$, implies that $\|Ax_m - y\| \leq \|Ax_j - y\|$ for all $j \leq m$. Therefore, we have

$$
|\langle \xi_m - \xi_l, x_m - x^* \rangle| \leq \sum_{j=l+1}^{m} \lambda_j \|Ax_j - y\|^r.
$$

Combining the inequality above with (18) we deduce that

$$
D_{\xi_l} f(x_m, x_l) \leq D_{\xi_l} f(x^*, x_l) - D_{\xi_m} f(x^*, x_m) + \sum_{j=l+1}^{m} \lambda_j \|Ax_j - y\|^r.
$$

From (15) it follows that the sequence $(D_{\xi_k} f(x^*, x_k))_{k \in \mathbb{N}}$ is monotonically decreasing. Thus, by inequality above and (16) we have $D_{\xi_l} f(x_m, x_l) \to 0$ as $l, m \to \infty$, and from the uniform convexity of f we obtain that $(x_k)_{k \in \mathbb{N}}$ is a Cauchy sequence in X. Therefore, there is $\bar{x} \in X$ such that $x_k \to \bar{x}$ as $k \to \infty$, and since Proposition 4.1 implies that $\|Ax_k - y\| \to 0$ as $k \to \infty$ and A is a continuous map, we conclude that $A\bar{x} = y$.

Now, we prove that $\bar{x} = x^\dagger$. We first observe that $\xi_k - \xi_0 \in \partial(D_{\xi_0} f(\cdot, x_0))(x_k)$, which yields

$$
\left\langle \xi_k - \xi_0, x^\dagger - x_k \right\rangle \leq D_{\xi_0} f\left(x^\dagger, x_0\right) - D_{\xi_0} f(x_k, x_0).
$$

Thus,

$$
D_{\xi_0} f(\bar{x}, x_0) \leq \liminf_{k \to \infty} D_{\xi_0} f(x_k, x_0) \leq D_{\xi_0} f\left(x^\dagger, x_0\right) + \liminf_{k \to \infty} \left\langle \xi_k - \xi_0, x_k - x^\dagger \right\rangle.
\tag{20}
$$

Next, we prove that

$$
\lim_{k \to \infty} \left\langle \xi_k - \xi_0, x_k - x^\dagger \right\rangle = 0,
\tag{21}
$$

which in view of (20) will ensure that $D_{\xi_0} f(\bar{x}, x_0) \leq D_{\xi_0} f(x^\dagger, x_0)$, proving that $\bar{x} = x^\dagger$. Indeed, using equation (19) with $x^* = x^\dagger$, for $m > l$, we have

$$
\left| \left\langle \xi_m - \xi_l, x_m - x^\dagger \right\rangle \right| = \left| \sum_{k=l+1}^{m} -\lambda_k \left\langle J_r\left(Ax_k - y\right), Ax_m - y \right\rangle \right| \leq \sum_{k=l+1}^{m} \lambda_k \left\| Ax_k - y \right\|^r \to 0
$$

as $l \to \infty$. Then, given $\epsilon > 0$ there exists $k_0 \in \mathbb{N}$ such that

$$k > k_0 \implies \left| \left\langle \xi_k - \xi_{k_0}, x_k - x^\dagger \right\rangle \right| < \frac{\epsilon}{2}.$$

Now,

$$\left| \left\langle \xi_{k_0} - \xi_0, x_k - x^\dagger \right\rangle \right| = \left| \sum_{n=1}^{k_0} -\lambda_n \left\langle J_r \left(A x_n - y \right), A x_k - y \right\rangle \right|$$

$$\leq \| A x_k - y \| \sum_{n=1}^{k_0} \lambda_n \| A x_n - y \|^{r-1}.$$

Since $\| A x_k - y \| \to 0$ as $k \to \infty$, we conclude that there exists a number $k_1 \geq k_0$ such that

$$k > k_1 \implies \left| \left\langle \xi_{k_0} - \xi_0, x_k - x^\dagger \right\rangle \right| < \frac{\epsilon}{2}.$$

Therefore, $k > k_1$ implies $\langle \xi_k - \xi_0, x_k - x^\dagger \rangle < \epsilon$, which proves (21) as we wanted. ∎

Our intention in the remainder of this section is to study the convergence properties of the family $(x_{k(\delta)}^\delta)_{\delta > 0}$ as the noise level δ approaches zero. To achieve this goal we first establish a stability result connecting x_k^δ to x_k. Observe that in general, x_{k+1}^δ is not uniquely defined from x_k^δ, which motivates the following definition.

Definition 4.7: Let $0 < \eta_0 \leq \eta_1 < 1$ be pre-fixed constants. $\widetilde{x} \in X$ is called a *successor* of x_k^δ if

(1) $k < k(\delta)$;
(2) There exists $0 \leq \widetilde{\lambda} < \infty$ such that $\widetilde{x} := \arg\min_{x \in X} T_{\widetilde{\lambda}}^\delta(x)$, where $T_{\widetilde{\lambda}}^\delta$ is the Tikhonov functional

$$T_{\widetilde{\lambda}}^\delta(x) := \widetilde{\lambda} r^{-1} \| A x - y^\delta \|^r + D_{\xi_k^\delta} f(x, x_k^\delta); \tag{22}$$

(3) The residual $\| A \widetilde{x} - y^\delta \|$ belongs to the interval

$$\left[(1 - \eta_0)\delta + \eta_0 \| A x_k^\delta - y^\delta \|, (1 - \eta_1)\delta + \eta_1 \| A x_k^\delta - y^\delta \| \right]. \tag{23}$$

In other words, there must exist a nonnegative Lagrange multiplier $\widetilde{\lambda}$, s.t. the minimizer $\widetilde{x}$ of the corresponding Tikhonov functional in (22) attains a residual $\| A \widetilde{x} - y^\delta \|$ in the interval (23) (which is defined by convex combinations of the noise level δ with the residual at the current iterate x_k^δ).

In the noisy-data case, as long as the discrepancy principle is not satisfied, the interval in (23) is a subset of

$$\left(\delta, (1 - \eta_1)\delta + \eta_1 \left\|Ax_k^\delta - y^\delta\right\|\right]$$ (24)

because

$$(1 - \eta_0)\delta + \eta_0 \left\|Ax_k^\delta - y^\delta\right\| \geq \delta + \eta_0(\tau - 1)\delta > \delta.$$

Therefore, if we consider a sequence generated by Algorithm 1 with the inequalities in [3.2] replaced by a more restrictive condition, as in item 3 of Definition 4.7, all the previous results still hold. Further, since $\|Ax_k^\delta - y^\delta\| > \tau\delta > \delta$ and $\eta_0 \leq \eta_1$, it is clear that interval (23) is non-empty.

We note that interval (23) becomes close to the interval (24) as $\eta_0 \to 0$, and therefore, the former interval is only a bit larger than the last one for $\eta_0 \approx 0$. In the noise-free case, (23) reduces to the non-empty interval

$$\left[\eta_0 \left\|Ax_k - y\right\|, \eta_1 \left\|Ax_k - y\right\|\right],$$ (25)

and according to Theorem 4.6, the sequence $(x_k)_{k\in\mathbb{N}}$ converges to $x^\dagger$ whenever x_{k+1} is a successor of x_k for all $k \in \mathbb{N}$. In this situation, we call $(x_k)_{k\in\mathbb{N}}$ a *noiseless sequence*.

Now, we study the behaviour of x_k^δ, for fixed k, as the noise level δ approaches zero. For the sake of the notation, we define

$$T_{\lambda_k}(x) := \frac{\lambda_k}{r}\left\|Ax - y\right\|^r + D_{\xi_k}f(x, x_k),$$

for $\delta = 0$ (compare it with (22)).

Theorem 4.8 (Stability): *Let $(\delta_j)_{j\in\mathbb{N}}$ be a positive-zero sequence and fix $\tau > 1$. Assume that the 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 $x_k^{\delta_j}$ for $k = 0, \ldots, k(\delta_j) - 1$. Further, assume that Y is a locally uniformly smooth Banach space. If x_0 is not a solution of (1), then there exists a noiseless sequence $(x_k)_{k\in\mathbb{N}}$ such that, for every fixed number $k \in \mathbb{N}$, there exists a subsequence $(\delta_{j_m})_{m\in\mathbb{N}}$ (depending on k) satisfying*

$$x_n^{\delta_{j_m}} \to x_n, \quad \xi_n^{\delta_{j_m}} \to \xi_n \quad \text{and} \quad f\left(x_n^{\delta_{j_m}}\right) \to f(x_n) \quad \text{as } m \to \infty, \quad \text{for } n = 0, \ldots, k.$$

Proof: Assume that x_0 is not a solution of $Ax = y$. We use an induction argument: since $x_0^\delta = x_0$ and $\xi_0^\delta = \xi_0$ for every $\delta \geq 0$, the result is clear for $k = 0$. The argument consists in successively choosing a subsequence of the current subsequence, and to avoid a notational overload, we denote a subsequence of (δ_j) still by (δ_j). Suppose the result holds true for some $k \in \mathbb{N}$. Then, there exists a subsequence $(\delta_j)_{j\in\mathbb{N}}$ satisfying

$$x_n^{\delta_j} \to x_n, \quad \xi_n^{\delta_j} \to \xi_n \quad \text{and} \quad f\left(x_n^{\delta_j}\right) \to f(x_n) \quad \text{as } j \to \infty, \quad \text{for } n = 0, \ldots, k,$$ (26)

where x_n is a successor of x_{n-1} for $n = 1, \ldots, k$. Because $x_{k+1}^{\delta_j}$ is a successor of $x_k^{\delta_j}$, it is true that $k < k(\delta_j)$ for all j. Due to the same reason, there exists a non-negative number $\lambda_k^{\delta_j}$ such

that

$$x_{k+1}^{\delta_j} = \arg\min_{x \in X} T_{\lambda_k^{\delta_j}}^{\delta_j}(x)$$

with the resulting residual $\|Ax_{k+1}^{\delta} - y^{\delta}\|$ lying in the interval (23). Our task now is proving that there exists a successor x_{k+1} of x_k and a subsequence (δ_j) of the current subsequence such that $x_{k+1}^{\delta_j} \to x_{k+1}$, $\xi_{k+1}^{\delta_j} \to \xi_{k+1}$ and $f(x_{k+1}^{\delta_j}) \to f(x_{k+1})$ as $j \to \infty$. Since the proof is relatively large, we divide it in 4 main steps: 1. we find a vector $\overline{x} \in X$ such that

$$x_{k+1}^{\delta_j} \rightharpoonup \overline{x} \tag{27}$$

for a specific subsequence $(\delta_j)_{j \in \mathbb{N}}$. 2. using the third item in Definition 4.7 we show that the sequence $(\lambda_k^{\delta_j})_{j \in \mathbb{N}}$ is bounded. Then, we choose a convergent subsequence and define

$$\lambda_k := \lim_{j \to \infty} \lambda_k^{\delta_j} < \infty, \tag{28}$$

as well as

$$x_{k+1} := \arg\min_{x \in X} T_{\lambda_k}(x). \tag{29}$$

3. we prove that $x_{k+1} = \overline{x}$ which guarantees that $x_{k+1}^{\delta_j} \rightharpoonup x_{k+1}$. 4. Finally, we prove that

$$f\left(x_{k+1}^{\delta_j}\right) \to f\left(x_{k+1}\right) \text{ as } j \to \infty, \tag{30}$$

which in view of (27) will guarantee that $x_{k+1}^{\delta_j} \to x_{k+1}$ as $j \to \infty$, since f has the Kadec property, see last paragraph in Section 2. The last result in turn, will prove that $\xi_{k+1}^{\delta_j} \to \xi_{k+1}$. Finally, we validate that x_{k+1} is a successor of x_k, completing the induction argument.

Step 1: From (12) and the uniform convexity of f follows that the sequence $(x_{k+1}^{\delta_j})_{j \in \mathbb{N}}$ is bounded. Thus, there exists a subsequence (δ_j) of the current subsequence, and a vector $\overline{x} \in X$ such that (27) holds true.

Step 2: We claim that, for each $k \in \mathbb{N}$ fixed, there exists a constant $\lambda_{\max,k} > 0$ such that

$$\lambda_k^{\delta_j} \leq \lambda_{\max,k} \quad \text{for all } j \in \mathbb{N}. \tag{31}$$

Indeed, assume the contrary. Then, there is a subsequence satisfying $\lambda_k^{\delta_j} \to \infty$ as $j \to \infty$. But in this case,

$$\liminf_{j \to \infty} \frac{1}{r} \left\| Ax_{k+1}^{\delta_j} - y^{\delta_j} \right\|^r \leq \liminf_{j \to \infty} \frac{T_{\lambda_k^{\delta_j}}^{\delta_j}\left(x_{k+1}^{\delta_j}\right)}{\lambda_k^{\delta_j}} \leq \liminf_{j \to \infty} \frac{T_{\lambda_k^{\delta_j}}^{\delta_j}\left(x^{\dagger}\right)}{\lambda_k^{\delta_j}}$$

$$= \lim_{j \to \infty} \left(\frac{1}{r} \left\| Ax^{\dagger} - y^{\delta_j} \right\|^r + \frac{D_{\xi_k^{\delta_j}} f\left(x^{\dagger}, x_k^{\delta_j}\right)}{\lambda_k^{\delta_j}} \right)$$

$$\leq \lim_{j \to \infty} \left(\frac{1}{r} \delta_j^r + \frac{D_{\xi_0} f\left(x^{\dagger}, x_0\right)}{\lambda_k^{\delta_j}} \right) = 0.$$

This, together with the lower bound of interval (23) implies that x_0 is a solution of $Ax = y$ because

$$0 \leq \eta_0^{k+1} \left\| Ax_0 - y \right\| = \eta_0^{k+1} \lim_{j \to \infty} \left(\left\| Ax_0^{\delta_j} - y^{\delta_j} \right\| - \delta_j \right)$$

$$\leq \liminf_{j \to \infty} \left(\left\| Ax_{k+1}^{\delta_j} - y^{\delta_j} \right\| - \delta_j \right) = 0, \tag{32}$$

and we have a contradiction. Thus (31) is true. We fix a convergent subsequence and define λ_k as in (28) and x_{k+1} as in (29).

Step 3: Observe that

$$\left| \langle \xi_k, \bar{x} - x_k \rangle - \left\langle \xi_k^{\delta_j}, x_{k+1}^{\delta_j} - x_k^{\delta_j} \right\rangle \right| \leq \left| \left\langle \xi_k, \bar{x} - x_{k+1}^{\delta_j} \right\rangle \right| + \|\xi_k\| \left\| x_k^{\delta_j} - x_k \right\|$$

$$+ \left\| \xi_k - \xi_k^{\delta_j} \right\| \left\| x_{k+1}^{\delta_j} - x_k^{\delta_j} \right\|.$$

Thus, from (27) and the induction hypothesis it follows that

$$\lim_{j \to \infty} \left\langle \xi_k^{\delta_j}, x_{k+1}^{\delta_j} - x_k^{\delta_j} \right\rangle = \langle \xi_k, \bar{x} - x_k \rangle.$$

Now, the weak lower semi-continuity of f, together with (27) and the induction hypothesis, implies that

$$D_{\xi_k} f(\bar{x}, x_k) = f(\bar{x}) - f(x_k) - \langle \xi_k, \bar{x} - x_k \rangle$$

$$\leq \liminf_{j \to \infty} f\left(x_{k+1}^{\delta_j} \right) - \lim_{j \to \infty} f\left(x_k^{\delta_j} \right) - \lim_{j \to \infty} \left\langle \xi_k^{\delta_j}, x_{k+1}^{\delta_j} - x_k^{\delta_j} \right\rangle$$

$$= \liminf_{j \to \infty} D_{\xi_k^{\delta_j}} f\left(x_{k+1}^{\delta_j}, x_k^{\delta_j} \right). \tag{33}$$

From (27) we have

$$Ax_{k+1}^{\delta_j} - y^{\delta_j} \rightharpoonup A\bar{x} - y \quad \text{as } j \to \infty, \tag{34}$$

which together with (28) and the lower semi-continuity of Banach space norms yields

$$T_{\lambda_k}(\bar{x}) = \frac{\lambda_k}{r} \left\| A\bar{x} - y \right\|^r + D_{\xi_k} f(\bar{x}, x_k)$$

$$\leq \liminf_{j \to \infty} \left(\frac{\lambda_k^{\delta_j}}{r} \left\| Ax_{k+1}^{\delta_j} - y^{\delta_j} \right\|^r + D_{\xi_k^{\delta_j}} f\left(x_{k+1}^{\delta_j}, x_k^{\delta_j} \right) \right)$$

$$= \liminf_{j \to \infty} T_{\lambda_k^{\delta_j}}^{\delta_j}\left(x_{k+1}^{\delta_j} \right) \leq \liminf_{j \to \infty} T_{\lambda_k^{\delta_j}}^{\delta_j}(x_{k+1}) = \lim_{j \to \infty} T_{\lambda_k^{\delta_j}}^{\delta_j}(x_{k+1}) = T_{\lambda_k}(x_{k+1}).$$

This proves that $\bar{x} = x_{k+1}$ because x_{k+1} is the unique minimizer of T_{λ_k}. Thus, $x_{k+1}^{\delta_j} \rightharpoonup x_{k+1}$ as $j \to \infty$. The above inequalities also ensure that $\liminf_{j \to \infty} T_{\lambda_k^{\delta_j}}^{\delta_j}(x_{k+1}^{\delta_j}) = T_{\lambda_k}(x_{k+1})$.

Taking a subsequence if necessary, we can assume that the following sequences converge:

$$a_j := D_{\xi_k^{\delta_j}} f\left(x_{k+1}^{\delta_j}, x_k^{\delta_j}\right), \quad a := \lim_{j \to \infty} a_j, \tag{35}$$

and

$$re_j := \left\|Ax_{k+1}^{\delta_j} - y^{\delta_j}\right\|^r, \quad re := \lim_{j \to \infty} re_j. \tag{36}$$

We can also assume for this subsequence that

$$\lim_{j \to \infty} T_{\lambda_k^{\delta_j}}^{\delta_j}\left(x_{k+1}^{\delta_j}\right) = T_{\lambda_k}\left(x_{k+1}\right). \tag{37}$$

Step 4: Define

$$c := D_{\xi_k} f\left(x_{k+1}, x_k\right)$$

and obseve that from (33), the inequality $c \le a$ holds. Thus, it suffices to prove that $a \le c$ to ensure that

$$\lim_{j \to \infty} D_{\xi_k^{\delta_j}} f\left(x_{k+1}^{\delta_j}, x_k^{\delta_j}\right) = D_{\xi_k} f\left(x_{k+1}, x_k\right),$$

which will prove (30). We assume that $a > c$ and derive a contradiction. From (28), (35), (36), (37), together with the definition of limit, it follows the existence of a number $N \in \mathbb{N}$ such that $j \ge N$ implies

$$T_{\lambda_k^{\delta_j}}^{\delta_j}\left(x_{k+1}^{\delta_j}\right) < T_{\lambda_k}\left(x_{k+1}\right) + \frac{a-c}{2}, \quad re \le re_j + \frac{a-c}{6\lambda_{\max,k}},$$

$$\lambda_k \le \lambda_k^{\delta_j} + \frac{a-c}{6re} \quad \text{and} \quad a \le a_j + \frac{a-c}{6}.$$

Therefore, for any $j \ge N$ it holds

$$T_{\lambda_k}\left(x_{k+1}\right) \le \lambda_k re + c = \left(\lambda_k - \lambda_k^{\delta_j}\right) re + \lambda_k^{\delta_j} re + a - (a-c)$$

$$\le \frac{a-c}{6re} re + \lambda_k^{\delta_j}\left(re_j + \frac{a-c}{6\lambda_{\max,k}}\right) + \left(a_j + \frac{a-c}{6}\right) - (a-c)$$

$$\le \lambda_k^{\delta_j} re_j + a_j - \frac{a-c}{2} = T_{\lambda_k^{\delta_j}}^{\delta_j}\left(x_{k+1}^{\delta_j}\right) - \frac{a-c}{2} < T_{\lambda_k}\left(x_{k+1}\right),$$

which leads to an obvious contradiction, proving that $a \le c$ as we wanted. Hence (30) holds, and since f has the Kadec property, it follows, in view of (27), that $x_{k+1}^{\delta_j} \to x_{k+1}$. This last result, together with the continuity of the duality mapping in locally uniformly smooth Banach spaces, implies that $\xi_{k+1}^{\delta_j} \to \xi_{k+1}$.

It only remains to prove that $\|Ax_{k+1} - y\|$ belongs to interval (25), which will guarantee that x_{k+1} is a successor of x_k. But this result follows from applying the limit $j \to \infty$ to the sequence $\|Ax_{k+1}^{\delta_j} - y^{\delta_j}\|$, which belongs to interval (23). ∎

Theorem 4.9 (Regularization): *Let $(\delta_j)_{j \in \mathbb{N}}$ be a positive-zero sequence and fix $\tau > 1$. Assume that the sequences $(x_k^{\delta_j})_{0 \le k \le k(\delta_j)}, j \in \mathbb{N}$, are fixed, where $x_{k+1}^{\delta_j}$ is a successor of $x_k^{\delta_j}$ for $k = 0, \ldots, k(\delta_j) - 1$. Further, assume that Y is a locally uniformly smooth Banach space. Then,*

$$\lim_{j \to \infty} x_{k(\delta_j)}^{\delta_j} = x^\dagger. \tag{38}$$

Proof: The result is clear if x_0 is a solution of $Ax = y$. Assume this is not the case. We first validate that the sequence $(k(\delta_j))_{j \in \mathbb{N}}$ has no convergent subsequences. Indeed, if for a subsequence $(\delta_{j_m})_{m \in \mathbb{N}}$ it is true that $k(\delta_{j_m}) \to n$ as $m \to \infty$, then since $k(\delta_{j_m}) \in \mathbb{N}$ for all $m \in \mathbb{N}$, we must have $k(\delta_{j_m}) = n \in \mathbb{N}$ for m large enough. From Theorem 4.8, the subsequence $(x_n^{\delta_{jm}})_{m \in \mathbb{N}}$ has itself a subsequence (which we still denote by $(x_n^{\delta_{jm}})_{m \in \mathbb{N}}$) which converges to x_n, this is,

$$\lim_{m \to \infty} x_{k(\delta_{jm})}^{\delta_{jm}} = \lim_{m \to \infty} x_n^{\delta_{jm}} = x_n.$$

But x_n is a solution of $Ax = y$ since

$$\begin{aligned}
\|y - Ax_n\| &= \lim_{m \to \infty} \left\| y - Ax_{k(\delta_{jm})}^{\delta_{jm}} \right\| \\
&\le \lim_{m \to \infty} \left(\left\| y - y^{\delta_{jm}} \right\| + \left\| y^{\delta_{jm}} - Ax_{k(\delta_{jm})}^{\delta_{jm}} \right\| \right) \\
&\le \lim_{m \to \infty} (\tau + 1)\, \delta_{j_m} = 0.
\end{aligned}$$

As in the proof of Theorem 4.8 (see (32)) we conclude that x_0 is a solution of $Ax = y$, contradicting our assumption. Therefore $k(\delta_j) \to \infty$ as $j \to \infty$.

We now prove that each subsequence of $(x_{k(\delta_j)}^{\delta_j})_{j \in \mathbb{N}}$ has itself a subsequence which converges to $x^\dagger$. This will prove (38). We observe that 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 which converges to $x^\dagger$.

Our first step is proving that, for every $\epsilon > 0$ fixed, there exists a subsequence (which we still denote by $(\delta_j)_{j \in \mathbb{N}}$) depending on ϵ, and a number $J = J(\epsilon)$, such that

$$j \ge J \implies D_{\xi_{k(\delta_j)}^{\delta_j}} f\left(x^\dagger, x_{k(\delta_j)}^{\delta_j} \right) < \epsilon. \tag{39}$$

In fact, fix $\epsilon > 0$ and let $(x_k)_{k \in \mathbb{N}}$ be the noiseless sequence constructed in Theorem 4.8. Since x_{k+1} is a successor of x_k for all $k \in \mathbb{N}$, the sequence $(x_k)_{k \in \mathbb{N}}$ converges to $x^\dagger$ (see Theorem 4.6). Then, there exists a natural number $N = N(\epsilon)$ such that

$$\left\| x_N - x^\dagger \right\| < \frac{1}{2}\sqrt{\frac{\epsilon}{2}} \quad \text{and} \quad D_{\xi_N} f\left(x^\dagger, x_N \right) < \frac{\epsilon}{2}.$$

From Theorem 4.8 there exists a subsequence (still denoted by $(\delta_j)_{j \in \mathbb{N}}$) depending on N, and a number $J_1 \in \mathbb{N}$, depending on ϵ, such that

$$j \ge J_1 \implies \left[\left\| \xi_N^{\delta_j} - \xi_N \right\| < \sqrt{\frac{\epsilon}{2}} \quad \text{and} \quad \left\| x_N^{\delta_j} - x_N \right\| < \frac{1}{2}\sqrt{\frac{\epsilon}{2}} \right].$$

Since $k(\delta_j) \to \infty$, there is a number $J_2 \in \mathbb{N}$ such that $k(\delta_j) \geq N$ for all $j \geq J_2$. It follows from (12) and the three points identity that for $j \geq J := \max\{J_1, J_2\}$,

$$D_{\xi_{k(\delta_j)}^{\delta_j}} f\left(x^\dagger, x_{k(\delta_j)}^{\delta_j}\right) \leq D_{\xi_N^{\delta_j}} f\left(x^\dagger, x_N^{\delta_j}\right)$$

$$\leq D_{\xi_N} f\left(x^\dagger, x_N\right) - D_{\xi_N} f\left(x_N^{\delta_j}, x_N\right) + \left\|\xi_N^{\delta_j} - \xi_N\right\| \left\|x_N^{\delta_j} - x^\dagger\right\|$$

$$\leq D_{\xi_N} f\left(x^\dagger, x_N\right) + \left\|\xi_N^{\delta_j} - \xi_N\right\| \left(\left\|x_N^{\delta_j} - x_N\right\| + \left\|x_N - x^\dagger\right\|\right) < \epsilon,$$

which proves (39).

Choosing $\epsilon = 1$, we can find a subsequence $(\delta_j)_{j \in \mathbb{N}}$ and select a number $j_1 \in \mathbb{N}$ such that

$$D_{\xi_{k(\delta_{j_1})}^{\delta_{j_1}}} f\left(x^\dagger, x_{k(\delta_{j_1})}^{\delta_{j_1}}\right) < 1.$$

Since the current subsequence $(\delta_j)_{j \in \mathbb{N}}$ is also a positive-zero sequence, the above reasoning can be applied again in order to extract a subsequence of the current one satisfying (39) with $\epsilon = 1/2$. We choose a number $j_2 \geq j_1$ such that the inequality

$$D_{\xi_{k(\delta_{j_2})}^{\delta_{j_2}}} f\left(x^\dagger, x_{k(\delta_{j_2})}^{\delta_{j_2}}\right) < \frac{1}{2}$$

holds true. Using induction, it is therefore possible to construct a subsequence $(\delta_{j_m})_{m \in \mathbb{N}}$ with the property

$$D_{\xi_{k(\delta_{j_m})}^{\delta_{j_m}}} f\left(x^\dagger, x_{k(\delta_{j_m})}^{\delta_{j_m}}\right) < \frac{1}{m} \quad \text{for all } m \in \mathbb{N},$$

which implies that

$$\lim_{m \to \infty} D_{\xi_{k(\delta_{j_m})}^{\delta_{j_m}}} f\left(x^\dagger, x_{k(\delta_{j_m})}^{\delta_{j_m}}\right) = 0,$$

and since f is uniformly convex,

$$\lim_{m \to \infty} \left\|x_{k(\delta_{j_m})}^{\delta_{j_m}} - x^\dagger\right\| = 0.$$

∎

5. Algorithms and numerical implementation

5.1. Determining the Lagrange multipliers

As before, we consider the function $G_{\hat{x}}(\lambda) = \|Ax_\lambda - y^\delta\|^r$, where $x_\lambda = \pi(\lambda, \hat{x})$ represents the minimizer of the Tikhonov functional

$$T_\lambda^\delta(x) = \frac{1}{r}\lambda \|Ax - y^\delta\|^r + D_\xi f(x, \hat{x}).$$

In order to choose the Lagrange multiplier in the k-th iteration, our strategy consists in finding a $\lambda_k > 0$ such that $G_{x_{k-1}}(\lambda_k) \in [a_k, b_k]$, where

$$a_k := \left(\eta_0 \delta + (1 - \eta_0)\|Ax_{k-1} - y^\delta\|\right)^r \quad \text{and} \quad b_k := \left(\eta_1 \delta + (1 - \eta_1)\|Ax_{k-1} - y^\delta\|\right)^r.$$

Here $0 < \eta_0 \le \eta_1 < 1$ are pre-defined constants.

We have employed three different methods to compute λ_k: the well-known secant and Newton methods and a third strategy, here called *adaptive method*, which we now explain: fix $\sigma_1, \sigma_2 \in (0, 1)$, $c_1 > 1$ and some initial value $\lambda_0^\delta > 0$. In the k-th iteration, $k \ge 1$, we define $\lambda_k^\delta = c_k \lambda_{k-1}^\delta$, where

$$c_k = \begin{cases} c_{k-1}\sigma_1, & \text{if } G_{x_{k-2}}(\lambda_{k-1}^\delta) < a_{k-1} \\ c_{k-1}\sigma_2^{-1}, & \text{if } G_{x_{k-2}}(\lambda_{k-1}^\delta) > b_{k-1}, \\ c_{k-1}, & \text{otherwise} \end{cases} \quad \text{for } k \ge 2.$$

The idea behind the adaptive method is observing the behaviour of the residual in last iterations and trying to determine how much the Lagrange multiplier should be increased in the next iteration. For example, the residual $G_{x_{k-2}}(\lambda_{k-1}^\delta) = \|Ax_{k-1} - y^\delta\|^r$ lying on the left of the target interval $[a_{k-1}, b_{k-1}]$, means that λ_{k-1}^δ was too large. We thus multiply c_{k-1} by a number $\sigma_1 \in (0, 1)$ in order to reduce the rate of growth of the Lagrange multiplier λ_k^δ, trying to hit the target in the next iteration.

Although the Newton method is efficient, in the sense that it normally finds a good approximation for the Lagrange multiplier in very few steps, it has the drawback of demanding the differentiability of the Tikhonov functional, and therefore it cannot be applied in all situations.

Because it does not require the evaluation of derivatives, the secant method can be used even for a nonsmooth Tikhonov functional. A disadvantage of this method is the high computational effort required to perform it.

Among these three possibilities, the adaptive strategy is the cheapest one, since it only demands one minimization of the Tikhonov functional per iteration. Further, this simple strategy does not require the derivative of this functional, which makes it fit in a large range of applications.

Note that this third strategy may generate a λ_k^δ such that $G_{x_{k-1}}(\lambda_k^\delta) \notin [a_k, b_k]$ in some iterative steps. This is the reason for correcting the factors c_k in each iteration. In our numerical experiments, the condition $G_{x_{k-1}}(\lambda_k^\delta) \in [a_k, b_k]$ was satisfied in almost all steps.

5.2. *Minimization of the Tikhonov functional*

In our numerical experiments, we are interested in solving the inverse problem (1), where $A : L^p(\Omega) \to L^2(\Omega)$, with $1 < p < \infty$, is linear and bounded, noisy data y^δ are available, and the noise level $\delta > 0$ is known.

In order to implement the nIT method (Algorithm 1), a minimizer of the Tikhonov functional (22) needs to be calculated in each iteration step. Minimizing this functional can be itself a very challenging task. We have used two algorithms for achieving this goal in our numerical experiments: 1. The Newton method was used for minimizing this functional in the case $p \ne 2$ and with a smooth function f, which induces the Bregman distance in the

penalization term. 2. The so called ADMM method was employed in order to minimize the Tikhonov functional for the case $p = 2$ (Hilbert space) and a nonsmooth functional f. In the following, we explain the details.

First we consider the Newton method. Define the Bregman distance induced by the norm-functional $f(g) := \frac{1}{p}\|g\|_{L^p}^p$, $1 < p < \infty$, which leads to the smooth penalization term $D_\xi f(g, h) = \Delta_p(g, h)$, see Section 2. The resulting Tikhonov functional reads

$$T_\lambda(g) = \tfrac{1}{2}\lambda\|Ag - y^\delta\|^2 + \Delta_p(g, g_{k-1}),$$

where g_{k-1} is the current iterate.[3] In this case, the optimality condition (10) reads:

$$F(\overline{g}) = \lambda A^* y^\delta + J_p(g_{k-1}), \tag{40}$$

where $\overline{g} \in L^p(\Omega)$ is the minimizer of $T_\lambda(g)$ and $F(g) := \lambda A^* Ag + J_p(g)$.

In order to apply the Newton method to the nonlinear equation (40), one needs to evaluate the derivative of F, which (whenever exists) is given by $F'(g) = \lambda A^* A + J_p'(g)$. Next we present an explicit expression for the Gâteaux-derivative $J_p'(g)$ (the derivation of this expression is postponed to Appendix A, where the Gâteaux-differentiability of J_p in $L^p(\Omega)$, for $p \geq 2$, is investigated). Given $g \in L^p(\Omega)$, with $p \geq 2$, it holds

$$\big(J_p'(g)\big)(h) = \langle (p-1)|g|^{p-2}, h \rangle, \quad \forall h \in L^p(\Omega), \tag{41}$$

where the linear operator $(p-1)|g|^{p-2} : h \mapsto (p-1)|g(\cdot)|^{p-2}h(\cdot)$ is to be understood in pointwise sense. In the discretized setting, $J_p'(g)$ is a diagonal matrix whose $i-$th element on its diagonal is $(p-1)|g(x_i)|^{p-2}$, with x_i being the $i-$th point of the chosen mesh.

In our numerical simulations, we consider the situation where the sought solution is sparse and, therefore, the case $p \approx 1$ is of our interest. We stress the fact that (A2) (see Appendix A) holds true even for $1 < p < 2$ whenever $x \neq 0$. Using this fact, one can prove that (41) holds for these values of p, e.g. if g does not change signal in Ω (i.e, either $g > 0$ in Ω, or $g < 0$ in Ω) and the direction h is a bounded function in this set. However, these strong hypotheses are very difficult to check, and even if they are satisfied, we still expect having stability problems for inverting the matrix $F'(g)$ if the function g attains a small value in some point of the mesh, because the function in (A2) satisfies $\gamma''(x) \to \infty$ as $x \to 0$. In order to avoid this kind of problem in our numerical experiments, we have replaced the $i-$th element on the diagonal of the matrix $J_p'(g)$ by $\min\{(p-1)|g(x_i)|^{p-2}, 10^6\}$.

The second method that we used in our experiments was the well-known *Alternating Direction Method of Multipliers* (ADMM), which has been implemented to minimize the Tikhonov functional associated with the inverse problem $Ax = y^\delta$, where $X = Y = \mathbb{R}^n$, $A : \mathbb{R}^n \to \mathbb{R}^n$, and $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a nonsmooth function.

ADMM is an optimization scheme for solving linearly constrained programming problems with decomposable structure [13], which goes back to the works of Glowinski and Marrocco [14], and of Gabay and Mercier [15]. Specifically, this algorithm solves problems in the form:

$$\min_{(x,z)\in\mathbb{R}^n\times\mathbb{R}^m} \{\varphi(x) + \phi(z) : Mx + Bz = d\}, \tag{42}$$

where $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $\phi : \mathbb{R}^m \to \overline{\mathbb{R}}$ are convex proper l.s.c. functions, $M : \mathbb{R}^n \to \mathbb{R}^l$ and $B : \mathbb{R}^m \to \mathbb{R}^l$ are linear operators, and $d \in \mathbb{R}^l$.

ADMM solves the coupled problem (42) performing a sequences of steps that decouple functions φ and ϕ, making it possible to exploit the individual structure of these functions. It can be interpreted in terms of alternating minimization, with respect to x and z, of the augmented Lagrangian function associated with problem (42). Indeed, ADMM consists of the iterations

$$x_{k+1} = \arg\min_{x \in \mathbb{R}^n} \mathcal{L}_\rho(x, z_k, u_k)$$

$$z_{k+1} = \arg\min_{z \in \mathbb{R}^M} \mathcal{L}_\rho(x_{k+1}, z, u_k)$$

$$u_{k+1} = u_k + \rho(Mx_{k+1} + Bz_{k+1} - d),$$

where $\rho > 0$ and $\mathcal{L}_\rho$ is the augmented Lagrangian function

$$\mathcal{L}_\rho(x, z, u) := \varphi(x) + \phi(z) + \langle u, Mx + Bz - d\rangle + \tfrac{1}{2}\rho\|Mx + Bz - d\|_2^2.$$

The convergence results for ADMM guarantee, under suitable assumptions, that the sequences $(x_k)_{k\in\mathbb{N}}$, $(z_k)_{k\in\mathbb{N}}$ and $(u_k)_{k\in\mathbb{N}}$ generated by the method are such that $Mx_k + Bz_k - d \to 0$, $\varphi(x_k) + \phi(z_k) \to s^\star$ and $u_k \to u^\star$, where $s^\star$ is the optimal value of problem (42) and $u^\star$ is a solution of the dual problem associated with (42).

For minimizing the Tikhonov functional using ADMM we introduce an additional decision variable z such that problem of minimizing $T_{\lambda_k^\delta}^\delta(x)$ for $x \in X$ is rewritten into the form of (42).

The specific choice of the functions φ, ϕ and the operators M and B is problem dependent (for a concrete example see, e.g. Section 6.2). This allows us to exploit the special form of the functional $T_{\lambda_k^\delta}^\delta$, and also to pose the minimization problem in a more suitable form, in order to be solved numerically.

In all numerical simulations presented in Section 6, the ADMM method is stopped when $\|Mx_k + Bz_k - d\|$ becomes smaller than a predefined threshold.

6. Numerical experiments

6.1. Deconvolution

In what follows we consider the application of the nIT method to the deconvolution problem modelled by the linear integral operator

$$Ax := \int_0^1 K(s, t)x(t)\,\mathrm{d}t = y(s),$$

where the kernel K is the continuous function defined by

$$K(s, t) = \begin{cases} 49s(1 - t), & s \le t \\ 49t(1 - s), & s > t \end{cases}.$$

This benchmark problem is considered in [3]. There, it is observed that $A : L^p[0, 1] \to C[0, 1]$ is continuous and bounded for $1 \le p \le \infty$. Thus, $A : L^p[0, 1] \to L^r[0, 1]$ is compact for $1 \le r < \infty$.
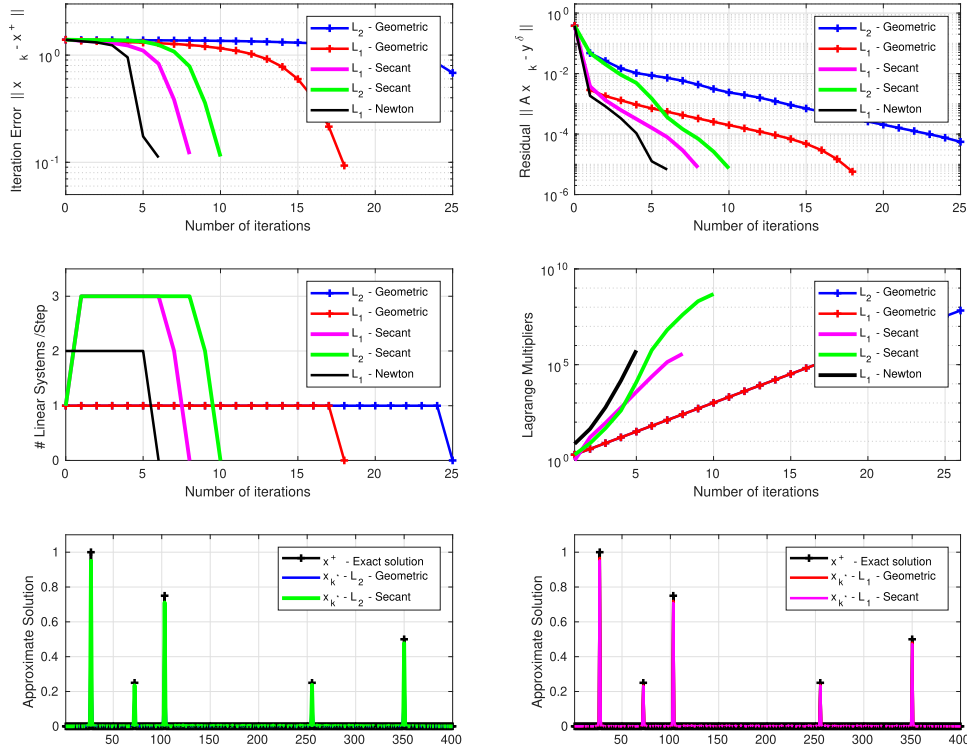
Figure 1. Deconvolution problem: Numerical experiments.

In our experiment, A is replaced by the discrete operator A_d, where the above integral is computed using a quadrature formula (trapezoidal rule) over an uniform partition of the interval $[0, 1]$ with 400 nodes.

The exact solution of the discrete problem is the vector $x^\star \in \mathbb{R}^{400}$ with $x^\star(27) = 2$, $x^\star(72) = 1.25$, $x^\star(103) = 1.75$, $x^\star(255) = 1.25$, $x^\star(350) = 1.5$ and $x^\star(i) = 0$, elsewhere.[4]

We compute $y = A_d x^\star$, the exact data, and add random Gaussian noise to $y \in \mathbb{R}^{400}$ to get the noisy data y^δ satisfying $\|y - y^\delta\|_Y \leq \delta$.

We follow [3] in the experimental setting and choose $\delta = 0.0005$, $\tau = 1.001$ (discrepancy principle), and $Y = L^2$. For the parameter space, two distinct choices are considered, namely $X = L^{1.001}$ and $X = L^2$.

Numerical results for the deconvolution problem are presented in Figure 1 (for simplicity, all legends in this figure refere to the space L^1; however, $p = 1.001$ is used in the computations). The following methods are depicted:

− (BLUE) L^2-penalization, Geometric sequence;
− (GREEN) L^2-penalization, Secant method;
− (RED) $L^{1.001}$-penalization, Geometric sequence;
− (PINK) $L^{1.001}$-penalization, Secant method;
− (BLACK) $L^{1.001}$-penalization, Newton method.

The six pictures in Figure 1 represent:

[TOP] Iteration error in L^2-norm (left);[5] residual in L^2-norm (right);
[CENTER] Number of linear systems/step (left); Lagrange multipliers (right);
[BOTTOM] exact solution and reconstructions with L^2-penalization (left); exact solution and reconstructions with $L^{1.001}$-penalization (right).

6.2. *Image deblurring*

In the sequel an application of the nIT method to an image deblurring problem is considered. This is a finite dimensional problem with spaces $X = \mathbb{R}^n \times \mathbb{R}^n$ and $Y = \mathbb{R}^n \times \mathbb{R}^n$. The vector $x \in X$ represents the pixel values of the original image to be restored, and $y \in Y$ contains the pixel values of the observed blurred image. In practice, only noisy blurred data $y^\delta \in Y$ satisfying (2) is available. The linear transformation A represents some blurring operator.

In the numerical simulations we consider the situation where the blur of the image is modelled by a space invariant point spread function (PSF). The exact solution is the 512×512 *Barbara* image (see Figure 2), and y^δ is obtained adding artificial noise to the exact data $y = Ax$ (here A is the covolution operator corresponding to the PSF).

For this problem the nIT method is implemented with two distinct penalization terms, namely $f(x) = \|x\|_2^2$ (L^2 penalization) and $f(x) = \frac{\mu}{2}\|x\|_2^2 + TV(x)$ ($L^2 + TV$ penalization). Here $\mu > 0$ is a regularization parameter and $TV(x) = \|\nabla x\|_1$ is the *total variation* norm of x, where $\nabla : \mathbb{R}^n \times \mathbb{R}^n \to (\mathbb{R}^n \times \mathbb{R}^n) \times (\mathbb{R}^n \times \mathbb{R}^n)$ is the *discrete gradient* operator. We minimize the Tikhonov functional associated with the $L^2 + TV$ penalization term using the ADMM described in Section 5.

In our experiments the values $\mu = 10^{-4}$, $\delta = 0.00001$ and $\tau = 1.5$ are used. Moreover, $x_0 = y^\delta$ and $\xi_0 = \nabla^*(sign(\nabla x_0))$ are chosen as initial guesses.

In Figure 2, the exact solution, the convolution kernel, and the noisy data are shown. The reconstructed images are shown in Figure 3, while the numerical results are presented in Figure 4. The following methods were implemented:

- (BLUE) L^2-penalization, Geometric sequence;
- (RED) $L^2 + TV$-penalization, Geometric sequence;
- (PINK) $L^2 + TV$-penalization, Secant method;
- (GREEN) $L^2 + TV$-penalization, Adaptive method.

The four pictures in Figure 4 represent:

[TOP] Iteration error $\|x^\star - x_k^\delta\|$;
[CENTER TOP] Residual $\|Ax_k^\delta - y^\delta\|$;
[CENTER BOTTOM] Number of linear systems solved in each step;
[BOTTOM] Lagrange multiplier λ_k. and $\tau = 1.5$. Moreover, the initial guesses $x_0 = y^\delta$

Remark 6.1: The Tikhonov functional associated with the $L^2 + TV$ penalization term is minimized using the ADMM in Section 5. Note that, if $f(x) = \frac{\mu}{2}\|x\|_2^2 + \|\nabla x\|_1$ then one

Figure 2. Image deblurring problem: (LEFT) Point Spread Function; (CENTER) Exact image; (RIGHT) Blurred image.

Figure 3. Image deblurring problem: Reconstructions (TOP LEFT) L^2–Geometric; (TOP RIGHT) $L^2 + TV$–Geometric; (BOTTOM LEFT) $L^2 + TV$–Secant; (BOTTOM RIGHT) $L^2 + TV$–Adaptive.

needs to solve

$$\min_{x \in X} \tfrac{1}{2}\lambda_k \left\| Ax - y^\delta \right\|^2 + \tfrac{\mu}{2}\|x - x_{k-1}^\delta\|^2 + \|\nabla x\|_1 - \|\nabla x_{k-1}^\delta\|_1 - \left\langle \xi_{k-1}^\delta, x - x_{k-1}^\delta \right\rangle$$

in each iteration. In order to use ADMM, we sate this problem into the form of (42) by defining $z = \nabla x$, $\varphi(x) := \frac{\lambda_k}{2}\|Ax - y^\delta\|^2 + \frac{\mu}{2}\|x - x_{k-1}^\delta\|^2 - \langle \xi_{k-1}^\delta, x - x_{k-1}^\delta \rangle$, $\phi(z) = \|z\|_1 - \|\nabla x_{k-1}^\delta\|_1$, $M = -\nabla$, $B = I$ and $d = 0$.
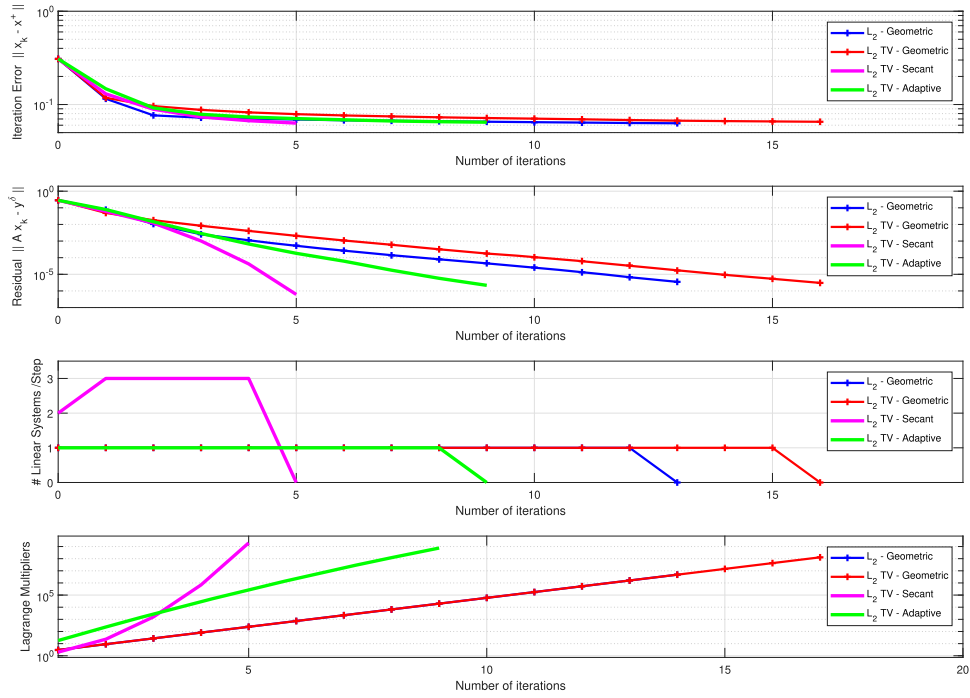
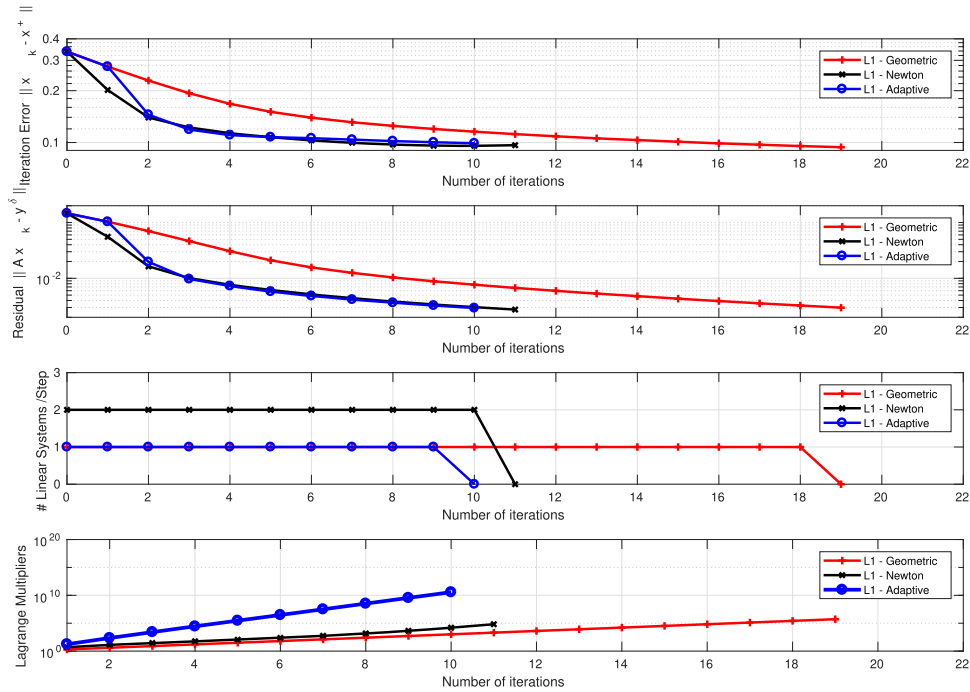**Figure 4.** Image deblurring problem: Numerical experiments.

Figure 5. Inverse Potential problem: Numerical experiments.

6.3. Inverse potential problem

The third application considered in this section, the Inverse Potentinal Problem (IPP), consists of recovering a coefficient function $x : \Omega \to \mathcal{R}$, from measurements of the Cauchy data of its corresponding potential on the boundary of the domain $\Omega = (0, 1) \times (0, 1)$.

The direct problem is modelled by the linear operator $A : L^2(\Omega) \ni x \mapsto w_\nu|_{\partial\Omega} \in L^2(\partial\Omega)$, where $w \in H^1(\Omega)$ solves the elliptic boundary value problem

$$\Delta w = x, \text{ in } \Omega; \quad w = 0, \text{ at } \partial\Omega. \tag{43}$$

Since $x \in L^2(\Omega)$, the Dirichlet boundary value problem in (43) admits a unique solution (known as potential) $w \in H^2(\Omega) \cap H^1_0(\Omega)$ [16].

The inverse problem we are concerned with, consists in determining the piecewise constant source function x from measurements of the Neumann trace of w at $\partial\Omega$. Using the above notation, the IPP can be written in the abbreviated form (1), with data $y = w_\nu|_{\partial\Omega}$.

In our implementations we follow [8] in the experimental setup: we set $\Omega = (0, 1) \times (0, 1)$ and assume that $x^\star \in H^1(\Omega)$ is a function with sharp gradients (see Figure 6).

The boundary value problem (43) is solved for w using $x = x^\star$, and the exact data $y = w_\nu|_{\partial\Omega}$ for the inverse problem is computed. The noisy data y^δ for the inverse problem is obtained by adding to y a normally distributed noise with zero mean, in order to achieve a prescribed relative noise level.
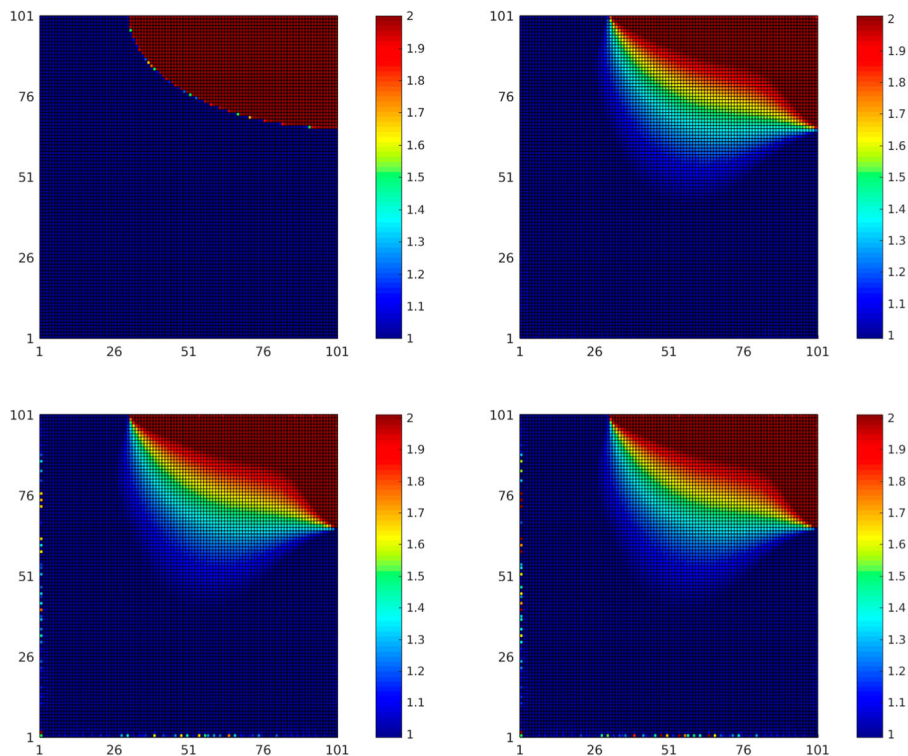

Figure 6. Inverse Potential problem: (TOP LEFT) Exact solution $x^\star$. Reconstructions (TOP RIGHT) L^1–Geometric; (BOTTOM LEFT) L^1–Newton; (BOTTOM RIGHT) L^1–Adaptive.

In our numerical implementations we set $\tau = 5$ (discrepancy principle constant), $\delta = 0.5\%$ (relative noise level), and the initial guess $x_0 \equiv 1.5$ (a constant function in Ω). For the parameter space we choose $X = L^p(\Omega)$ with $p = 1.5$, while the data space is $Y = L^2(\partial\Omega)$.

Numerical results for the Inverse potential problem are presented in Figure 5. The following methods are depicted:

- — (RED) L^p-penalization, Geometric sequence;
- — (BLACK) L^p-penalization, Newton method;
- — (BLUE) L^p-penalization, Adaptive method;

The four pictures in Figure 5 represent: [*TOP*] Iteration error in L^p-norm; [*CENTER TOP*] Residual in L^2-norm; [*CENTER BOTTOM*] Number of linear systems/step; [*BOTTOM*] Lagrange multipliers.

The corresponding reconstruction results are shown in Figure 6: [*TOP LEFT*] Exact Solution; [*TOP RIGHT*] geometric method; [*BOTTOM LEFT*] Newton method; [*BOTTOM LEFT*] Adaptive method.

7. Conclusions

In this manuscript we investigate a novel strategy for choosing the regularization parameters in the nonstationary iterated Tikhonov (nIT) method for solving ill-posed operator equations modelled by linear operators acting between Banach spaces. The novelty of our approach consists in defining strategies for choosing a sequence of regularization parameters (i.e. Lagrange multipliers) for the nIT method.

A preliminary (numerical) investigation of this method was conducted in [9]. In the present manuscript we derive a complete convergence analysis, proving convergence, stability and semi-convergence results (see Section 4). In Sections 6.1 and 6.2 we revisit two numerical applications discussed in [9]. Moreover, in Section 6.3 we investigate a classical benchmark problem in the inverse problems literature, namely the 2D elliptic Inverse Potential Problem.

The Lagrange multipliers are chosen (a posteriori) in order to enforce a fast decay of the residual functional (see Algorithm 1 and Section 4). The computation of these multipliers is performed by means of three distinct methods: (1) a secant type method; (2) a Newton type method; (3) an adaptive method using a geometric sequence with non-constant growth rate, where the rate is updated after each step.

The computation of the iterative step of the nIT method requires the minimization of a Tikhonov Functional (see Section 4). This task is solved here using two distinct methods: (1) in the case of smooth penalization and Banach parameter-spaces the optimality condition (related to the Tikhonov functional) leads to a nonlinear equation, which is solved using a Newton type method; (2) in the case of nonsmooth penalization and Hilbert parameter-space, the ADMM method is used for minimizing the Tikhonov functional.

Notes

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

2. This Lemma guarantees that, given a reflexive Banach space E, and a nonempty closed convex set $A \subset E$, then any convex l.s.c proper function $\phi : A \to (-\infty, \infty]$ achieves its minimum on A.
3. Here (1) is replaced by $Ag = y^{\delta}$.
4. Notice that we are dealing with a discrete inverse problem, and discretization errors associated to the continuous model are not being considered.
5. For the purpose of comparison, the iteration error is ploted in the in L^2-norm for both choices of the parameter space $X = L^2$ and $X = L^{1.001}$.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

References

[1] Engl HW, Hanke M, Neubauer A. Regularization of inverse problems. Dordrecht: Kluwer Academic Publishers Group; 1996. (Mathematics and its applications; 375). MR 1408680.
[2] Hanke M, Groetsch CW. Nonstationary iterated Tikhonov regularization. J Optim Theory Appl. 1998;98(1):37–53.
[3] Jin Q, Stals L. Nonstationary iterated Tikhonov regularization for ill-posed problems in Banach spaces. Inverse Probl. 2012;28(3):104011.
[4] Schuster T, Kaltenbacher B, Hofmann B, et al. Regularization methods in Banach spaces. Berlin: de Gruyter; 2012.
[5] De Cezaro A, Baumeister J, Leitão A. Modified iterated Tikhonov methods for solving systems of nonlinear ill-posed equations. Inverse Probl Imaging. 2011;5(1):1–17.
[6] Jin Q, Zhong M. Nonstationary iterated Tikhonov regularization in Banach spaces with uniformly convex penalty terms. Numer Math. 2014;127(3):485–513. MR 3216817
[7] Lorenz D, Schoepfer F, Wenger S. The linearized bregman method via split feasibility problems: analysis and generalizations. SIAM J Imaging Sci. 2014;7(2):1237–1262.
[8] Boiger R, Leitão A, Svaiter BF. Range-relaxed criteria for choosing the Lagrange multipliers in nonstationary iterated Tikhonov method. IMA J Numer Anal. 2019;39(1). DOI:10.1093/imanum/dry066
[9] Machado MP, Margotti F, Leitão A. On nonstationary iterated Tikhonov methods for ill-posed equations in Banach spaces. In: Hofmann B, Leitão A, Zubelli JP, editors. New trends in parameter identification for mathematical models. Cham: Springer International Publishing; 2018. p. 175–193.
[10] Cioranescu I. Geometry of Banach spaces, duality mappings and nonlinear problems. Dordrecht: Kluwer Academic Publishers Group; 1990. (Mathematics and its applications; 62). MR 1079061.
[11] Brezis H. Functional analysis, Sobolev spaces and partial differential equations. New York: Springer; 2011. (Universitext). MR 2759829.
[12] Rockafellar RT Conjugate duality and optimization, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1974, Lectures given at the Johns Hopkins University, Baltimore, Md., June, 1973, Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 16. MR 0373611.
[13] Eckstein J, Bertsekas DP. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math Program. 1992;55(3):293–318. MR 1168183.

[14] Glowinski R, Marrocco A. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires. Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge Anal. Numér.. 1975;9(R-2):41–76. MR 0388811.

[15] Gabay D, Mercier B. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput Math Appl. 1976;2(1):17–40.

[16] Evans LC. Partial differential equations. Providence, RI: American Mathematical Society; 1998. (Graduate studies in mathematics; 19).

Appendix

In this appendix we discuss how to compute the Gâteaux derivative of J_p. Given $g \in L^p(\Omega)$, the key idea for deriving a formula for $J_p'(g)$ is to observe the differentiability of the function $\gamma : \mathbb{R} \ni x \mapsto p^{-1}|x|^p \in \mathbb{R}$. This function is differentiable in $\mathbb{R}$ whenever $p > 1$ and, in this case it holds

$$\gamma'(x) = |x|^{p-1}\text{sign}(x), \quad \text{where sign}(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{if } x = 0 \\ -1, & \text{if } x < 0 \end{cases}. \tag{A1}$$

Furthermore, γ is twice differentiable in $\mathbb{R}$ if $p \geq 2$, with derivative given by

$$\gamma''(x) = (p-1)|x|^{p-2}. \tag{A2}$$

This formula still holds true for $1 < p < 2$, but only in $\mathbb{R}\setminus\{0\}$. In this case, $\gamma''(0)$ does not exist and $\gamma''(x)$ grows to infinity as x approaches to zero.

Since $J_p(g) = (\frac{1}{p}\|g\|_{L^p}^p)'$ can be identified with

$$J_p(g) = |g|^{p-1}\text{sign}(g) \tag{A3}$$

(see [10]), which looks very similar to γ' in (A1), the bounded linear operator $J_p'(g) : L^p(\Omega) \to L^{p^*}(\Omega)$ should be in some sense similar to γ'' in (A2). We then fix $g \in L^p(\Omega)$ with $p \geq 2$ and try to prove (41), i.e.

$$(J_p'(g))(h) = \langle (p-1)|g|^{p-2}, h \rangle, \quad \forall h \in L^p(\Omega),$$

where the linear operator $(p-1)|g|^{p-2} : h \mapsto (p-1)|g(\cdot)|^{p-2}h(\cdot)$ is to be understood in pointwise sense. This ensures that J_p is Gâteaux-differentiable in $L^p(\Omega)$ and its derivative J_p' can be identified with $(p-1)|\cdot|^{p-2}$.

Note that, given $h \in L^p(\Omega)$, equality (41) holds iff $\lim_{t \to 0} f_t = (p-1)|g(\cdot)|^{p-2}h(\cdot)$, where $f_t : \Omega \to \mathbb{R}$ is defined by $f_t(x) := t^{-1}[J_p(g+th) - J_p(g)](x)$. This is equivalent to prove that

$$\lim_{t \to 0} \left\| f_t - (p-1)|g(\cdot)|^{p-2}h(\cdot) \right\|_{L^{p^*}(\Omega)}^{p^*} = 0.$$

In view of (A3) and (A2), it follows that for each $x \in \Omega$ fixed we have

$$\begin{aligned} \lim_{t \to 0} f_t(x) &= \lim_{t \to 0} t^{-1}\left[|g(x)+th(x)|^{p-1}\text{sign}(g(x)+th(x)) - |g(x)|^{p-1}\text{sign}(g(x)) \right] \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\left[|g(x)+th(x)|^{p-1}\text{sign}(g(x)+th(x)) \right]_{t=0} \\ &= \left[(p-1)|g(x)+th(x)|^{p-2}h(x) \right]_{t=0} = (p-1)|g(x)|^{p-2}h(x). \end{aligned}$$

Thus, making use of Lebesgue's Dominated Convergence Theorem, we conclude that

$$\lim_{t\to 0}\left\|f_t - (p-1)|g(\cdot)|^{p-2}h(\cdot)\right\|^{p^*}_{L^{p^*}(\Omega)} = \lim_{t\to 0}\int_\Omega \left|f_t(x) - (p-1)|g(x)|^{p-2}h(x)\right|^{p^*} dx$$

$$= \int_\Omega \left|\lim_{t\to 0} f_t(x) - (p-1)|g(x)|^{p-2}h(x)\right|^{p^*} dx = 0.$$

which proves (41).