On Nonstationary Iterated Tikhonov Methods for Ill-Posed Equations in Banach Spaces

M. P. Machado, F. Margotti, and Antonio Leitão

Abstract In this article we propose a novel *nonstationary iterated Tikhonov* (nIT) type method for obtaining stable approximate solutions to ill-posed operator equations modeled by linear operators acting between Banach spaces. We propose a novel a posteriori strategy for choosing the sequence of regularization parameters (or, equivalently, the Lagrange multipliers) for the nIT iteration, aiming to obtain a fast decay of the residual.

Numerical experiments are presented for a 1D convolution problem (smooth Tikhonov functional and Banach parameter-space), and for a 2D deblurring problem (nonsmooth Tikhonov functional and Hilbert parameter-space).

1 Introduction

In this article we propose and (numerically) investigate a new *nonstationary Iterated Tikhonov* (nIT) type method [6, 9] for obtaining stable approximations of linear ill-posed problems modeled by operators mapping between Banach spaces.

The novelty of our approach consists in adopting an a posteriori strategy for the choice of the Lagrange multipliers, which aims to achieve a predefined decay of the residual in each iteration. This strategy differs from the classical choice for the Lagrange multipliers in [9, 10], which propose an a priori strategy, 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 the set of data $y \in Y$, where X, Y are Banach spaces. In practical situations, one does not know the data exactly; instead, only approximate

M. P. Machado
IMPA, Estrada Dona Castorina 110, 22460-320 Rio de Janeiro, Brazil
e-mail: majentcha@gmail.com

F. Margotti · A. Leitão (✉)
Department of Mathematics, Federal University of St. Catarina, P.O. Box 476, 88040-900 Florianópolis, Brazil
e-mail: fabiomarg@gmail.com; acgleitao@gmail.com

B. Hofmann et al. (eds.), *New Trends in Parameter Identification for Mathematical Models*, Trends in Mathematics, https://doi.org/10.1007/978-3-319-70824-9_10

measured data $y^\delta \in Y$ are available with

$$\|y^\delta - y\|_Y \le \delta\,, \tag{1}$$

where $\delta > 0$ is the (known) noise level. The available data y^δ are obtained by indirect measurements of the parameter x, this process being described by the ill-posed operator equation

$$A\,x = y^\delta\,, \tag{2}$$

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. For a comprehensive study of this type of problems, we refer the reader to the text book [13] and to the references therein.

Iterated Tikhonov type methods are typically used for linear inverse problems. Applications of this method for linear operator equations in Hilbert spaces can be found in [9] (see also [4] for the nonlinear case). In the Hilbert space setting, both a priori and a posteriori strategies for choosing the Lagrange multipliers have been extensively analyzed [6].

The research on the Banach space setting is still ongoing. Some preliminary results can be found in [10] for linear operator equations, and in [11] for nonlinear systems. In both references above, the authors consider a priori strategies for choosing the Lagrange multipliers.

The approach presented here is devoted to the Banach space setting, and consists in adopting an a posteriori strategy for the choice of the Lagrange multipliers. The penalty terms used in our Tikhonov functionals are the same as in [11] and consist of Bregman distances induced by (uniformly) convex functionals (e.g., the sum of the L^2-norm with the TV-seminorm).

This chapter is outlined as follows: In Sect. 2 a revision of relevant background material is presented. In Sect. 3 we introduce our nIT method. In Sect. 4 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 5 is devoted to numerical experiments, while in Sect. 6 we present some final remarks and conclusions.

2 Background Material

For details about the material discussed in this section, we refer the reader to the textbooks [3] and [13].

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

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

The set $\mathrm{Dom}\,(f)$ is always convex and we call f *proper* provided $\mathrm{Dom}\,(f)$ is not empty. The functional f is called *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

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

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

$$f\,(x) \le \lim\inf_{k\to\infty} f\,(x_k)\,.$$

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

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

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

$$\mathrm{Dom}\,(\partial f) := \{x \in X :\;\; \partial f\,(x) \neq \emptyset\}\,.$$

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

Sub-differentiable and convex l.s.c. functionals are strongly connected to each other. 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

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

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

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

for all $x \in X$ and $y \in Y$, where $A^*: Y^* \to X^*$ is the Banach-adjoint of A. As a consequence,

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

If a convex functional $f: X \to \overline{\mathbb{R}}$ 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 $x \in X$,

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

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

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

for all $x^* \in J_p(x)$. In a Hilbert spaces X, by using the Riesz Representation Theorem, one can prove that $J_2(x) = x$ for all $x \in X$. Further, only in Hilbert spaces J_2 is a linear map.

Banach spaces are classified according with their geometrical characteristics. Many concepts concerning these characteristics are usually defined using the *modulus of convexity* and the *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 $J_p: X \to X^*$ is single-valued. If the functional f in (6) is Fréchet-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 implies local uniform smoothness, which in turn implies smoothness. Further, 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.

[1] Normally, the differentiability and convexity properties of this functional are independent of the particular choice of $p > 1$.

Assume f is proper. Then choosing elements $x, y \in X$ with $y \in \operatorname{Dom}(\partial f)$, 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 $D_\xi f(x,y) \geq 0$. 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 $x_2 \in X$, $x_1, x_3 \in \operatorname{Dom}(\partial f)$, $\xi_1 \in \partial f(x_1)$ 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.,

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

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.

If $f : X \to \overline{\mathbb{R}}$ is uniformly convex, then for all $y \in X$, $x \in \operatorname{Dom}(\partial f)$, $\xi \in \partial f(x)$ and $\lambda \in (0,1)$,

$$\begin{aligned} f(\lambda x + (1-\lambda) y) &\geq f(x) + \langle \xi, (\lambda x + (1-\lambda) y) - x \rangle \\ &= f(x) + (1-\lambda) \langle \xi, y - x \rangle , \end{aligned}$$

which in view of (3) implies

$$\langle \xi, y - x \rangle + \lambda \varphi(\|x-y\|) \leq f(y) - f(x) .$$

Now, letting $\lambda \to 1^-$, we obtain $\varphi(\|x-y\|) \leq D_\xi f(y,x)$. Analogously, the inequality

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

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

It is well-known that for $1 < p < \infty$, the Lebesgue space $L^p(\Omega)$, the Sobolev space $W^{n,p}(\Omega)$ and the space of p-summable sequences $\ell^p(\mathbb{R})$ are uniformly smooth and uniformly convex Banach spaces.

3 The Iterative Method

In this section we introduce the nonstationary iterated Tikhonov method to solve (2). The method we propose here is in the spirit of the method in [11], with the distinguish feature of using an endogenous strategy for the choice of the Lagrange multipliers λ_k^δ.

Specifically, fixing $r > 0$ and a uniformly convex penalty term f, the iterative method defines sequences (x_k^δ) in X and (ξ_k^δ) in X^* iteratively by

$$\begin{aligned} x_k^\delta &:= \arg\min_{x\in X} \frac{\lambda_k^\delta}{r} \left\|Ax - y^\delta\right\|^r + D_{\xi_{k-1}^\delta} f\left(x, x_{k-1}^\delta\right) \\ \xi_k^\delta &:= \xi_{k-1}^\delta - \lambda_k^\delta A^* J_r(Ax_k^\delta - y^\delta), \end{aligned}$$

where the multiplier λ_k^δ will be determined using only information about A, δ, y^δ and x_{k-1}^δ.

Our strategy for selecting the Lagrange multipliers is inspired in the recent work [1], where it was proposed an endogenous strategy for the choice of the Lagrange multiplier in the iterative method for solving (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 [1] 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. Further, the admissible Lagrange multipliers (in each iteration) shall belong to a non-degenerate interval.

With the view to extend this framework to Banach space setting we are forced to work with 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 $P_C(x) = \mathit{arg\,min}_{z\in C}\|z - x\|^2$, 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.

In what follows we assume the following conditions:

(A.1) There exist 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 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.

We define Ω_μ^r, the μ-levelset of the residual functional $\|Ax - y^\delta\|$, as

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

We observe that since A is a continuous linear operator it follows that Ω_μ^r is closed and convex.

Now, given $\hat{x} \in \mathrm{Dom}(\partial f)$ and $\xi \in \partial f(\hat{x})$, we can define the *Bregman projection* of $\hat{x}$ onto Ω_μ^r, as a solution of the minimization problem

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

It is clear that a solution of the above problem 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 projection is well defined when $\mu > \delta$, and in this case we can 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 1 *If $\mu > \delta$ then problem (8) has a solution.*

Proof Hypothesis (A.1), together with Eq. (1) and the assumption that $\mu > \delta$, imply that the feasible set of problem (8), i.e. the set Ω_μ^r, is nonempty.

By conditions (A.2) and (A.3) we have 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 that X is a reflexive space and applying [2, Corollary 3.23]. □

It is easy to see that if $0 \leq \mu' \leq \mu$ then $\Omega^r_{\mu'} \subseteq \Omega^r_\mu$, and $A^{-1}(y) \subset \Omega^r_\mu$ for all $\mu \geq \delta$. Furthermore, with the available information of the solution set of (2), 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_k^δ from $x_{k-1}^\delta \notin \Omega^r_\delta$ 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 now analyze the minimization problem (8) by means of Lagrange multipliers. The Lagrangian function associated with problem (8) is

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

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

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

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

Lemma 2 *Suppose that* $\|A\hat{x} - y^\delta\| > \mu > \delta$*, then the following assertions are equivalent*

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

Proof By (1), hypothesis (A.1) and the assumption $\mu > \delta$, we have that $x^\star \in X$ is such that

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

Inequality above implies the Slater condition for problem (8). Thus, using that A is continuous and $D_\xi f(\cdot, \hat{x})$ is l.s.c., we have 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, see [12].

The KKT conditions [12] for (8) are

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

If we suppose that $\lambda = 0$ in 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 $\mathscr{L}(\cdot, 0)$. Since $\|A\hat{x} - y^\delta\| > \mu$ we conclude that the pair $(\hat{x}, 0)$ does not satisfy the KKT conditions. Hence, we have $\lambda > 0$ and $G_{\hat{x}}(\lambda) - \mu^r = 0$. We conclude the lemma using the definition of $\pi(\hat{x}, \lambda)$. □

We are now ready to formulate the nIT method for solving (2).

Properties (4) and (5), together with the definition of the duality mapping, imply that the point $x_k^\delta \in X$ minimizes the *Tikhonov functional*

$$T_\lambda^\delta(x) := \frac{\lambda_k^\delta}{r} \left\|Ax - y^\delta\right\|^r + D_{\xi_{k-1}^\delta} f\left(x, x_{k-1}^\delta\right),$$

if and only if

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

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

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

Therefore, 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.

Algorithm 1 The iterative method

[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 $\lambda_k^\delta, x_k^\delta$ such that $x_k^\delta = \arg\min_{x\in X} \frac{\lambda_k^\delta}{r}\|Ax - y^\delta\|^r + D_{\xi_{k-1}^\delta} f(x, x_{k-1}^\delta)$,
and $\delta^r < G_{x_{k-1}^\delta}(\lambda_k^\delta) \le \left(\eta\delta + (1-\eta)\|Ax_{k-1}^\delta - y^\delta\|\right)^r$.
Set $\xi_k^\delta = \xi_{k-1}^\delta - \lambda_k^\delta A^* J_r(Ax_k^\delta - y^\delta)$.

4 Algorithms and Numerical Implementation

4.1 Determining the Lagrange Multipliers

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

$$T_\lambda(x) = \frac{\lambda}{r}\left\|Ax - y^\delta\right\|^r + D_\xi f(x, \hat{x})\,. \tag{11}$$

In order to determine the Lagrange multiplier in the iteration k, we need to calculate $\lambda_k > 0$ such that $G_{x_{k-1}}(\lambda_k) \in [a_k, b_k]$, where

$$a_k := \delta^r \quad \text{and} \quad b_k := (\eta\delta + (1-\eta)\|Ax_{k-1} - y^\delta\|)^r,$$

with $0 < \eta < 1$ pre-defined.

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

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

The idea behind the adaptive method is observing the behavior 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 the number c_{k-1} by a number $\sigma_1 \in (0,1)$ in order to reduce the speed of growing of the Lagrange multipliers 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 request the derivative of this functional, which makes it fit in a large range of applications.

Notice 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 (see the slope of the green curve on Fig. 3; bottom picture).

4.2 Minimization of the Tikhonov Functional

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

In order to apply the iterative method (Algorithm 1), a minimizer of the Tikhonov functional (11) needs to be calculated on each iteration. 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 $p \neq 2$ 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 Sect. 2. The resultant Tikhonov functional is

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

where g_{k-1} is the current iterate.[2] In this case, the optimality condition (10) reads:

$$F(\overline{g}) = \lambda A^* y^\delta + J_p(g_{k-1}), \tag{12}$$

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

[2]Here (2) is replaced by $Ag = y^\delta$.

In order to apply the Newton method to the nonlinear equation (12), one needs to evaluate the derivative of F, which (if it exists) is given by $F'(g) = \lambda A^* A + J_p'(g)$. Next, we prove that J_p is at least Gâteaux-differentiable in $L^p(\Omega)$, if $p \geq 2$. Further, we present an explicit expression for $J_p'(g)$, which will be used later in our numerical experiments.

The key for finding a formula for $J_p'(g)$ is observing the differentiability of the function $\gamma : \mathbb{R} \to \mathbb{R}$, $x \mapsto \frac{1}{p}|x|^p$. This function is differentiable in $\mathbb{R}$ whenever $p > 1$, and in this case,

$$\gamma'(x) = |x|^{p-1} \operatorname{sign}(x), \qquad \text{where } \operatorname{sign}(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{if } x = 0 \\ -1, & \text{if } x < 0 \end{cases}. \tag{13}$$

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

$$\gamma''(x) = (p-1)|x|^{p-2}. \tag{14}$$

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 $J_p(g) = \left(\frac{1}{p}\|g\|_{L^p}^p\right)'$ can be identified with (see [3])

$$J_p(g) = |g|^{p-1} \operatorname{sign}(g), \tag{15}$$

which looks very similar to γ' in (13), the bounded linear operator $J_p'(g) : L^p(\Omega) \to L^{p^*}(\Omega)$ is similar to γ'' in (14). Indeed, for any fixed $g \in L^p(\Omega)$, with $p \geq 2$, we have

$$\left\langle J_p'(g), h \right\rangle = \left\langle (p-1)|g|^{p-2}, h \right\rangle, \tag{16}$$

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

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 Eq. (14) holds true even for $1 < p < 2$ whenever $x \neq 0$. Using this fact, one can prove that (16) holds true for these values of p, for instance, if g does not change signal in Ω (i.e., $g > 0$ 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 $F'(g)$ in (12) if the function g has a small value in some point of the mesh, because the function in (14) 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 $\max\left\{(p-1)\,|g(x_i)|^{p-2}, 10^6\right\}$.

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 [5], which goes back to the works of Glowinski and Marrocco [8], and of Gabay and Mercier [7]. Specifically, this algorithm solves problems in the form:

$$\min_{(x,z)} \{\varphi(x) + \phi(z) \;:\; Mx + Bz = d\}, \tag{17}$$

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 (17) 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 (17). Indeed, ADMM consists of the iterations

$$\begin{aligned} x_{k+1} &= \arg\min_x \mathscr{L}_\rho(x, z_k, u_k) \\ z_{k+1} &= \arg\min_z \mathscr{L}_\rho(x_{k+1}, z, u_k) \\ u_{k+1} &= u_k + \rho(Mx_{k+1} + Bz_{k+1} - d), \end{aligned}$$

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

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

The convergence results for ADMM guarantee, under suitable assumptions, that the sequences (x_k), (z_k) and (u_k), 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 (17) and $u^\star$ is a solution of the dual problem associated with (17).

For minimizing the Tikhonov functional using ADMM we introduce an additional decision variable z such that problem

$$\min_{x \in X} T^\delta_{\lambda_k^\delta}(x)$$

is rewritten into the form of (17). The specific choice of the functions φ, ϕ and the operators M and B is problem dependent. For a concrete example, please see Sect. 5.2. This allows us to exploit the special form of the functional $T^\delta_{\lambda_k^\delta}$ and pose the problem in a more suitable manner to solve it numerically.

In our numerical simulations we stopped ADMM when $\|Mx_k\| + Bz_k - d$ was less than a prefixed tolerance.

5 Numerical Experiments

5.1 Deconvolution

The first application considered here is the deconvolution problem modeled by the linear integral operator

$$A\,x := \int_0^1 K(s,t)\,x(t)\,dt \;=\; y(s)\,,$$

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 [10]. 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 $1 \le r < \infty$.

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 $x^\star(48) = 2$, $x^\star(200) = 1.5$, $x^\star(270) = 1.75$ and $x^\star(i) = 0$, elsewhere.

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 $\| y - y^\delta \|_Y \le \delta$.

We follow [10] 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 are presented in Fig. 1.[3] The following methods are implemented:

- (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 Fig. 1 represent:

[Top] Iteration error in L^2-norm (left)[4]; residual in L^2-norm (right);
[Center] Number of linear systems/step (left); Lagrange multipliers (right);

[3] For simplicity, all legends in this figure refers to the space L^1; however, we used $p = 1.001$ in the computations.

[4] For the purpose of comparison, the iteration error is plotted in the in L^2-norm for both choices of the parameter space $X = L^2$ and $X = L^{1.001}$.

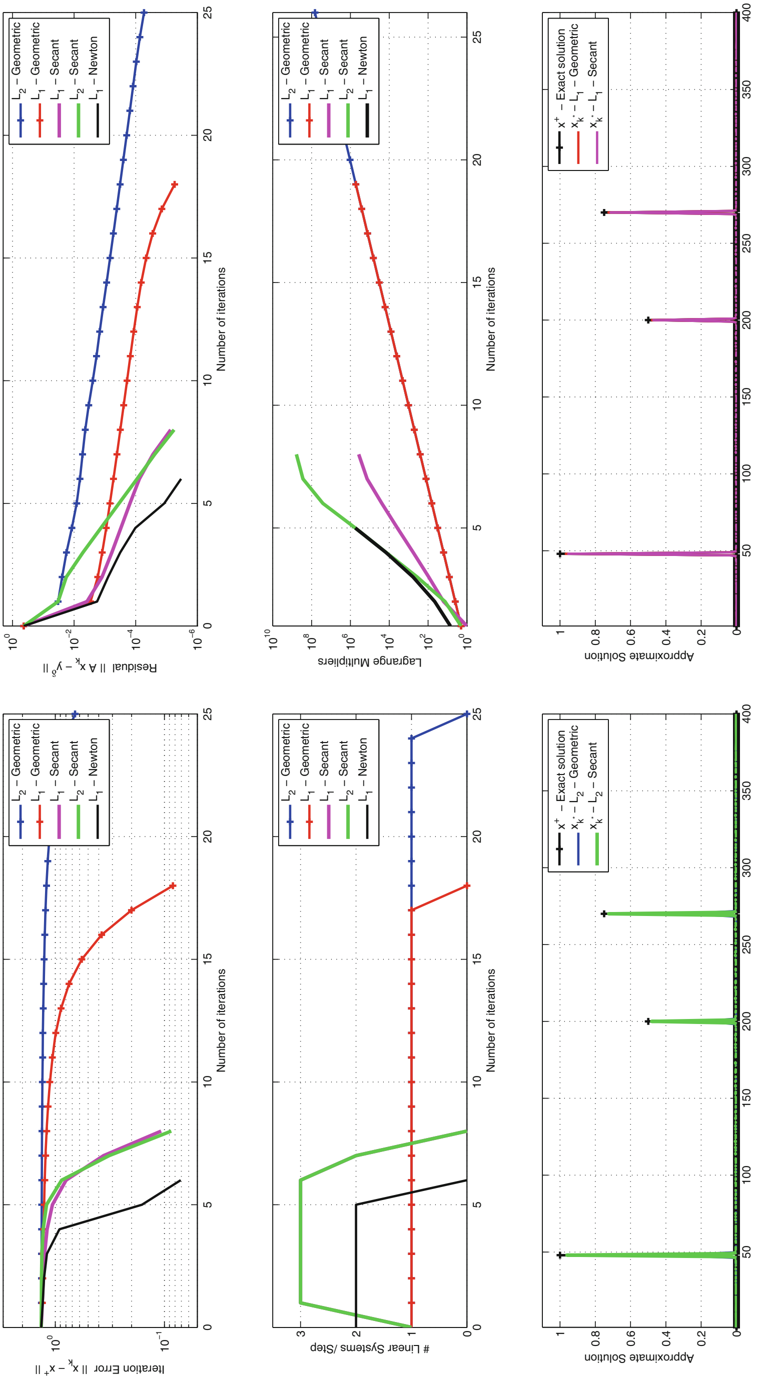

Fig. 1 Deconvolution problem: numerical experiments

[Bottom] Exact solution and reconstructions with L^2-penalization (left); exact solution and reconstructions with $L^{1.001}$-penalization (right).

5.2 Image Deblurring

The second application of the nIT method that we consider is the image deblurring problem. 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 (1) is available. The linear transformation A represents some blurring operator.

For our numerical simulations we consider the situation where the blur of the image is modeled by a space invariant point spread function (PSF). We use the 256×256 *Cameraman* test image, and y^δ is obtained adding artificial noise to the exact data $Ax = y$ (here A is the convolution operator corresponding to the PSF).

For this problem we implemented the nIT method with two different 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 Sect. 4. Specifically, if $f(x) = \frac{\mu}{2}\|x\|_2^2 + \|\nabla x\|_1$, then on each iteration we need to solve

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

To use ADMM we sate this problem into the form of problem (17) defining $z = \nabla x$, $\varphi(x) := \frac{\lambda_k^\delta}{2}\|Ax - y^\delta\|^2 + \frac{\mu}{2}\|x - x_{k-1}^\delta\|^2 - \left\langle \xi_{k-1}^\delta, x - x_{k-1}^\delta \right\rangle$, $\phi(z) = \|z\|_1 - \|\nabla x_{k-1}^\delta\|_1$, $M = -\nabla$, $B = I$ and $d = 0$.

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

Figure 2 shows the recovered images using the two penalization terms, and the different strategies we considered for choosing the Lagrange multipliers.

Figure 3 presents some numerical results. We implemented for this example the following methods:

- (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.

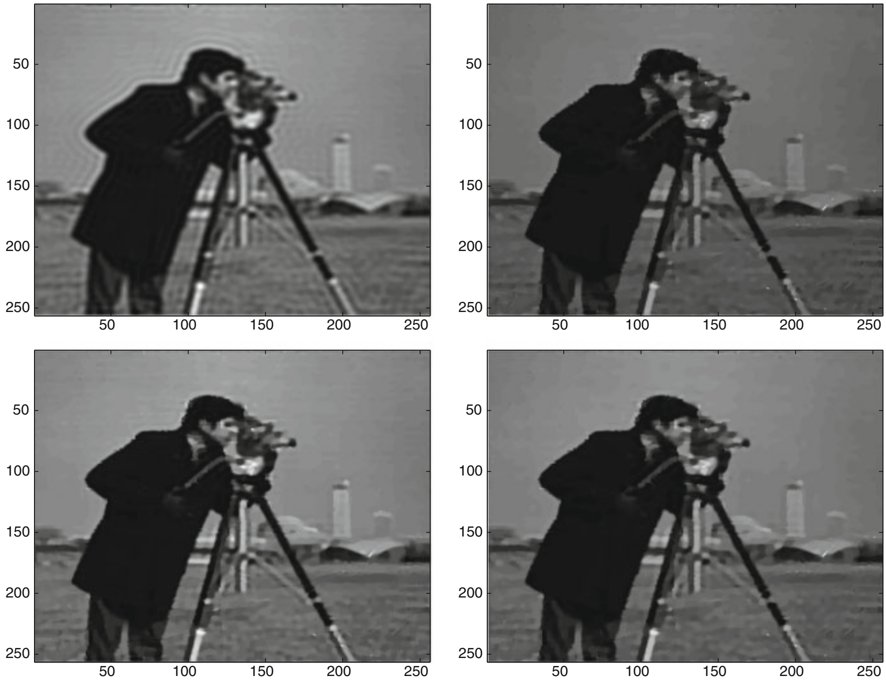

Fig. 2 Image deblurring problem: (top left) Geometric sequence, L^2 penalization; (top right) Geometric sequence, L^2 + TV penalization; (bottom left) Secant method, L^2 + TV penalization; (bottom right) Adaptive method, L^2 + TV penalization

The four pictures in Fig. 3 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^δ.

6 Conclusions

In this chapter we propose a novel nonstationary iterated Tikhonov (nIT) type method for obtaining stable approximate solutions to ill-posed operator equations modeled by linear operators acting between Banach spaces.

The novelty of our approach consists in defining strategies for choosing a sequence of regularization parameters (Lagrange multipliers) for the nIT method.

The Lagrange multipliers are chosen (a posteriori) in order to enforce a fast decay of the residual functional (see Algorithm 1 and Sect. 4.1). The computation of these multipliers is performed by means of three distinct methods: (1) a secant

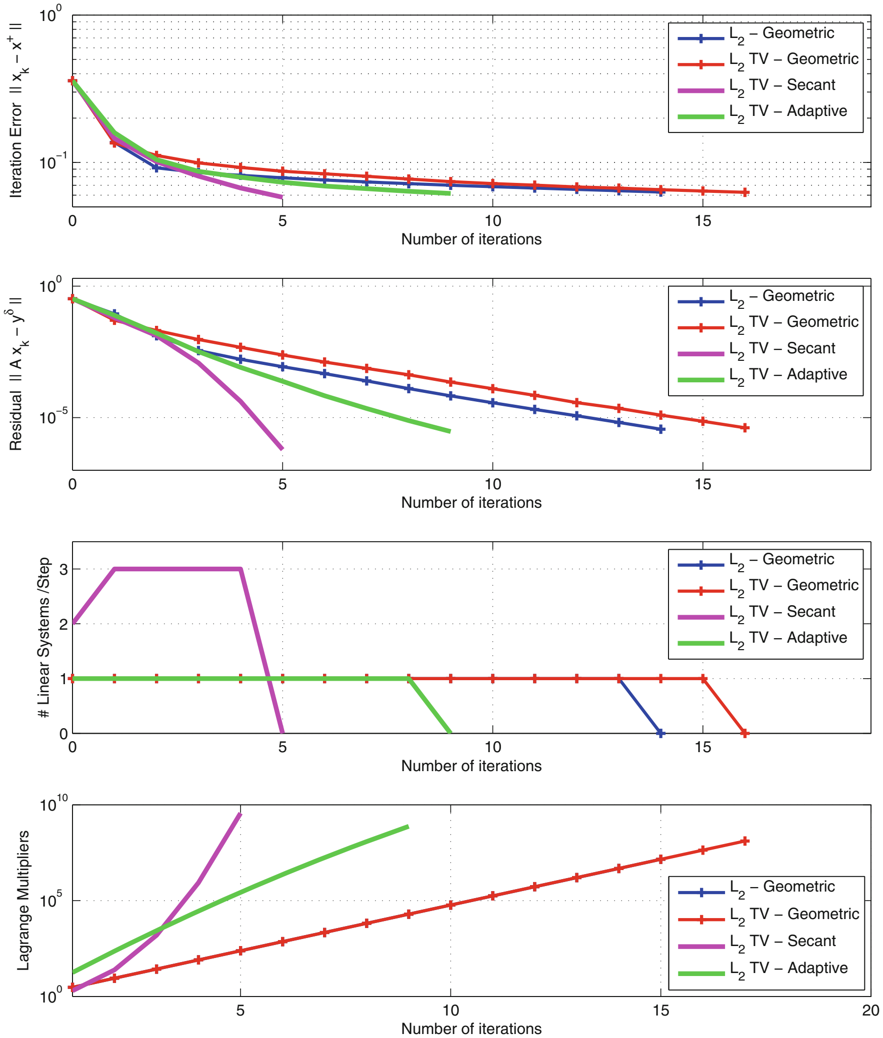

Fig. 3 Image deblurring problem: numerical experiments

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 type Functional (see Sect. 4.2). 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.

What concerns the Deconvolution problem in Sect. 5.1[5]:

– The secant and the Newton methods produce a sequence of multipliers with faster growth, when compared to the geometric (a priori) choice of multipliers.
– The fact above is observed in both parameter spaces L^2 and $L^{1.001}$.
– The secant and the Newton methods converge within fewer iterations than the geometric choice of multipliers.
– The numerical effort required by the secant type method is similar to the one required by the geometric choice of multipliers.
– The Newton method requires the smallest amount of computational effort.
– As expected, the sparse solution $x^\star$ is better approximated by the methods operating in the $L^{1.001}$ parameter-space.

What concerns the Deblurring problem in Sect. 5.2[6]:

– The secant and the adaptive methods produce a sequence of multipliers with faster growth, when compared to the geometric (a priori) choice of multipliers.
– The secant and the adaptive methods converge within fewer iterations.
– The numerical effort required by the secant type method is similar to the one required by the geometric choice of multipliers.
– The adaptive method requires the smallest amount of computational effort.
– The first reconstructed image (L^2 penalization) differs from the other three reconstructions ($L^2 + TV$ penalization), which produce images with sharper edges and better defined contours.

Acknowledgements F. Margotti is supported by CAPES and IMPA. A. Leitão acknowledges support from CNPq (grant 309767/2013-0), and from the AvH Foundation.

References

1. R. Boiger, A. Leitão, B.F. Svaiter, Endogenous strategies for choosing the Lagrange multipliers in nonstationary iterated Tikhonov method for ill-posed problems (2016, submitted)
2. H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext (Springer, New York, 2011)
3. I. Cioranescu, *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*. Mathematics and Its Applications, vol. 62 (Kluwer Academic, Dordrecht, 1990)
4. A. De Cezaro, J. Baumeister, A. Leitão, Modified iterated Tikhonov methods for solving systems of nonlinear ill-posed equations. Inverse Prob. Imaging **5**(1), 1–17 (2011)
5. J. Eckstein, D.P. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. Ser. A **55**(3), 293–318 (1992)
6. H.W. Engl, M Hanke, A Neubauer, *Regularization of Inverse Problems*. Mathematics and Its Applications, vol. 375 (Kluwer Academic, Dordrecht, 1996)

[5]In this situation we have smooth penalization terms and Banach parameter-spaces.

[6]In this situation we have nonsmooth penalization terms and Hilbert parameter-spaces.

7. D. Gabay, B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. **2**(1), 17–40 (1976)
8. R. Glowinski, A. Marrocco, 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. **9**(R-2), 41–76 (1975)
9. M. Hanke, C.W. Groetsch, Nonstationary iterated Tikhonov regularization. J. Optim. Theory Appl. **98**(1), 37–53 (1998)
10. Q. Jin, L. Stals, Nonstationary iterated Tikhonov regularization for ill-posed problems in Banach spaces. Inv. Probl. **28**(3), 104011 (2012)
11. Q. Jin, M. Zhong, Nonstationary iterated Tikhonov regularization in Banach spaces with uniformly convex penalty terms. Numer. Math. **127**(3), 485–513 (2014)
12. R.T. Rockafellar, *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, vol. 16
13. T. Schuster, B. Kaltenbacher, B. Hofmann, K.S. Kazimierski, *Regularization Methods in Banach Spaces* (de Gruyter, Berlin, 2012)