Iterative regularization with a general penalty term—theory and application to L^1 and TV regularization

2012 Inverse Problems 28 104010

(http://iopscience.iop.org/0266-5611/28/10/104010)

Inverse Problems **28** (2012) 104010 (19pp) doi:10.1088/0266-5611/28/10/104010

Iterative regularization with a general penalty term—theory and application to L^1 and TV regularization

Radu Ioan Boţ[1] **and Torsten Hein**[2]

[1] Faculty of Mathematics, Chemnitz University of Technology, 09107 Chemnitz, Germany
[2] Matheon, Technical University of Berlin, 10623 Berlin, Germany

E-mail: radu.bot@mathematik.tu-chemnitz.de and hein@math.tu-berlin.de

Received 18 July 2011, in final form 14 September 2011
Published 1 October 2012
Online at stacks.iop.org/IP/28/104010

Abstract

In this paper, we consider an iterative regularization scheme for linear ill-posed equations in Banach spaces. As opposed to other iterative approaches, we deal with a general penalty functional from Tikhonov regularization and take advantage of the properties of the regularized solutions which where supported by the choice of the specific penalty term. We present convergence and stability results for the presented algorithm. Additionally, we demonstrate how these theoretical results can be applied to L^1- and TV-regularization approaches and close the paper with a short numerical example.

(Some figures may appear in colour only in the online journal)

1. Introduction

Let $\mathcal{X}$ and $\mathcal{Y}$ denote real Banach spaces with topological dual spaces $\mathcal{X}^*$ and $\mathcal{Y}^*$, respectively. We consider the linear ill-posed operator equation

$$A\,x = y, \qquad x \in \mathcal{X}, \tag{1}$$

where $A : \mathcal{X} \longrightarrow \mathcal{Y}$ describes a linear continuous operator with a non-closed range $\mathcal{R}(A) := \{A\,x \in \mathcal{Y} : x \in \mathcal{X}\}$, i.e. $\mathcal{R}(A) \neq \overline{\mathcal{R}(A)}$. Additionally, we assume that only noisy data $y^\delta \in \mathcal{Y}$, with $\|y^\delta - y\| \leqslant \delta$, $\delta > 0$, and $y \in \mathcal{Y}$ are given. Consequently, we have to apply a regularization strategy.

Certainly the most popular stabilization approach is Tikhonov regularization. Motivated by its successful employment in various applications, the theory and the numerics of Tikhonov regularization with general residual and penalty terms have become fields of active research in the recent years; see, for example, [25, 6–9, 17, 14, 23] for some theoretical results as well as for some applications in image and sparse reconstruction. This variational approach represents nowadays a standard technique in the approximate determination of, in particular, non-smooth parameters and images. On the other hand, the use of Tikhonov regularization

0266-5611/12/104010+19$33.00

for identification problems has a major drawback: as opposed to control problems, the choice of the regularization parameter is crucial for the quality of the reconstructed solution. In order to apply a parameter choice strategy, the Tikhonov functional has to be minimized several times for different regularization parameters. In particular, very small regularization parameters have to be taken into account, leading to increasing numerical instabilities and costs. Therefore, iterative regularization methods seem to be a promising alternative: instead of solving several (non-quadratic and ill-conditioned) minimization problems exactly, we apply an iterative minimization process for the residual term and stop the algorithm whenever some stopping criterion is satisfied. Hence, numerically, only one minimization problem has to be solved inexactly, a fact which promises much less computational cost. However, the theoretical treatment of such processes in the context of inverse problems is much more difficult. As a consequence, the literature concerning iterative regularization methods is limited and it is mainly restricted to the case of quadratic penalty terms (Hilbert space norms and semi-norms), see [12, 13, 18]. Recently, some first iterative variants were developed in Banach spaces by taking norms as penalty functionals, see [27, 26, 19, 15, 16].

In this paper, an iterative regularization approach for solving (1) is investigated. In particular, motivated by [15], we deal for all $\delta \geqslant 0$ with the following iterative scheme: for a starting point $x_0^* \in \mathcal{X}^*$ we set $x_0^\delta := G(x_0^*)$ and iterate for $n \geqslant 0$:

$$\begin{aligned} \phi_n^* &:= A^\star J_p(Ax_n^\delta - y^\delta);\\ x_{n+1}^* &:= x_n^* - \mu_n \phi_n^*;\\ x_{n+1}^\delta &:= G(x_{n+1}^*). \end{aligned}$$

As usual for iterative regularization schemes, the process is terminated with an appropriate stopping criterion, which will be specified later on. Here, we use the following notation:

- $A^\star : \mathcal{Y}^* \longrightarrow \mathcal{X}^*$ denotes the adjoint operator of A, i.e.

$$\langle A^\star y^*, x\rangle = \langle y^*, Ax\rangle, \qquad \forall x \in \mathcal{X}, y^* \in \mathcal{Y}^*.$$

- For given $1 < p < +\infty$, the operator $J_p : \mathcal{Y} \longrightarrow \mathcal{Y}^*$ denotes the duality mapping with the gauge function $t \mapsto t^{p-1}$. Hence, when $\mathcal{Y}$ is additionally assumed to be smooth, ϕ_n^* is the Gâteaux gradient of the functional $x \mapsto \frac{1}{p}\|Ax - y^\delta\|^p$ at the element $x_n^\delta \in \mathcal{X}$ for all $n \geqslant 0$.
- $G : \mathcal{D}(G) \subseteq \mathcal{X}^* \longrightarrow \mathcal{X}$ describes an operator which transports $x_n^* \in \mathcal{X}^*$ back into the original space $\mathcal{X}$. Its proper choice and the investigation of its influence on the outcomes of the iterative regularization scheme represent the main purpose of this paper.
- In order to achieve a tolerable speed of convergence for the presented algorithm, a good choice of the step size $\mu_n > 0$ for $n \geqslant 0$ has to be taken into account.

Furthermore, for $\delta > 0$, let $N(\delta, y^\delta)$ denote the index where the iteration process is stopped, assuming that this happens. Then $x^\delta_{N(\delta,y^\delta)}$ is referred to as the regularized solution of (1). For $\delta = 0$, we omit writing the upper index for the sequence $\{x_n^0\}_{n\geqslant 0}$ and let $y^0 := y$. The main goal of this paper is to present a general framework for the employment of this approach concerning convergence and regularization. Nevertheless, we also suggest how to apply this method to some particular penalty functionals, beyond the ones considered in the classical Tikhonov regularization.

For some alternative approaches to iterative regularization methods for linear equations in Banach spaces, we refer to [1], where the iterative regularization is used for general convex problems in uniformly smooth and uniformly convex Banach spaces, and to [3], where the tools involved rely on operator calculus, while for complexity and effectiveness issues regarding convex optimization algorithms we refer the interested reader to [20].

The paper is organized as follows: sections 2 and 3 motivate and give analytical background for the specific choice of the operator G. This preliminary work is followed in section 4 by a detailed specification of the iterative scheme under consideration. In section 5, convergence and regularization properties of the algorithm are proved. An additional accelerated iterative scheme, obtained via an improved choice of the step size, is given in section 6. Finally, an application of the proposed method to regularization with L^1- and TV-penalty terms is given in section 7, along with a short numerical example.

2. Motivation—Tikhonov regularization

In order to get an idea about the choice of the operator G, we briefly consider Tikhonov regularization with a general penalty functional $P : \mathcal{X} \longrightarrow \mathbb{R} \cup \{+\infty\}$ assumed to be proper (i.e. its effective domain, $\operatorname{dom} P := \{x \in \mathcal{X} : P(x) < +\infty\}$, is supposed to be nonempty), convex and lower semicontinuous.

Then, for a given regularization parameter $\alpha > 0$, a regularized approximate solution x_α^δ of equation (1) is calculated as a minimizer of the Tikhonov functional:

$$T_\alpha^\delta : \mathcal{X} \longrightarrow \mathbb{R} \cup \{+\infty\}, \qquad T_\alpha^\delta(x) := \frac{1}{p}\|Ax - y^\delta\|^p + \alpha P(x).$$

Assume $\mathcal{Y}$ to be smooth and P to be Gâteaux differentiable on $\operatorname{core}(\operatorname{dom} P)$, the algebraic interior of $\operatorname{dom} P$, and suppose further $x_\alpha^\delta \in \operatorname{core}(\operatorname{dom} P)$. Writing down the necessary optimality condition, we consequently have

$$\nabla T_\alpha^\delta\big(x_\alpha^\delta\big) = A^\star J_p\big(Ax_\alpha^\delta - y^\delta\big) + \alpha \nabla P\big(x_\alpha^\delta\big) = 0$$

or, equivalently,

$$\nabla P\big(x_\alpha^\delta\big) = -\frac{1}{\alpha} A^\star J_p\big(Ax_\alpha^\delta - y^\delta\big).$$

The above considerations suggest for an iterative scheme the choice

$$G := (\nabla P)^{-1},$$

provided that the Gâteaux gradient of P is invertible. However, the assumption of differentiability of the penalty functional P seems to be too restrictive. In order to get an iterative approach applicable to not necessarily differentiable penalty functionals, we will make use of the notion of convex subdifferential. The convex subdifferential of P at $x \in \operatorname{dom} P$ is the set

$$\partial P(x) := \{x^* \in \mathcal{X}^* : P(\tilde{x}) - P(x) - \langle x^*, \tilde{x} - x\rangle \geqslant 0 \ \forall \tilde{x} \in \mathcal{X}\},$$

while for $x \notin \operatorname{dom} P$, $\partial P(x) := \emptyset$. Thus, $\partial P : \mathcal{X} \rightrightarrows \mathcal{X}^*$ represents a multi-valued operator having as a domain

$$\mathcal{D}(\partial P) := \{x \in \mathcal{X} : \partial P(x) \neq \emptyset\} \subseteq \operatorname{dom} P$$

and as a range

$$\mathcal{R}(\partial P) := \bigcup_{x \in \mathcal{X}} \partial P(x).$$

Its inverse operator $(\partial P)^{-1} : \mathcal{X}^* \rightrightarrows \mathcal{X}$ is the operator defined as

$$x \in (\partial P)^{-1}(x^*) \ \Leftrightarrow x^* \in \partial P(x).$$

Consequently, $\mathcal{D}((\partial P)^{-1}) = \mathcal{R}(\partial P)$ and $\mathcal{R}((\partial P)^{-1}) = \mathcal{D}(\partial P)$. Hence, for our iterative scheme we will choose

$$G : \mathcal{R}(\partial P) \subseteq \mathcal{X}^* \longrightarrow \mathcal{X}, \qquad G := (\partial P)^{-1}, \tag{2}$$

after we will preliminarily guarantee that $(\partial P)^{-1}$ is single-valued on its domain. Moreover, before proving convergence and stability results, we have to ensure that the sequences $\{x_n^\delta\}_{n\geqslant 0}$ and $\{x_n^*\}_{n\geqslant 0}$ are well defined. In particular, the following questions have to be taken into account.

(1) How can one find an appropriate penalty functional P such that the operator G defined in (2) is single-valued on $\mathcal{R}(\partial P)$?
(2) Can we, in this case, always ensure that $x_n^* \in \mathcal{R}(\partial P)$ for all $n \geqslant 1$? Or, even more, under which conditions does $\mathcal{R}(\partial P) = \mathcal{X}^*$ hold?
(3) How to choose the step size μ_n for all $n \geqslant 0$?

The answers to these questions are given in the following sections.

3. Elements of convex analysis

Throughout the paper, we suppose the space $\mathcal{X}$ to be a reflexive Banach space and $\mathcal{X}^*$ its topological dual space. We denote by $w(\mathcal{X}, \mathcal{X}^*)$ (for short, w) the weak topology of $\mathcal{X}$ induced by $\mathcal{X}^*$ and by $w(\mathcal{X}^*, \mathcal{X})$ (for short, w^*) the weak* topology of $\mathcal{X}^*$ induced by $\mathcal{X}$. We also denote by $\langle x^*, x\rangle$ the value of the linear continuous functional $x^* \in \mathcal{X}^*$ at $x \in \mathcal{X}$. For a set $S \subseteq \mathcal{X}$, we denote by $\operatorname{int} S$ and $\overline{S}$ its interior and closure, respectively. The indicator function of S is defined as

$$\delta_S : \mathcal{X} \longrightarrow \mathbb{R} \cup \{+\infty\}, \qquad \delta_S(x) = \begin{cases} 0, & \text{if } x \in S, \\ +\infty, & \text{otherwise,} \end{cases}$$

while the convex subdifferential of δ_S,

$$N_S : \mathcal{X} \rightrightarrows \mathcal{X}^*, \qquad N_S(x) := \begin{cases} \{x^* \in \mathcal{X}^* : \langle x^*, \tilde{x} - x\rangle \leqslant 0 \ \forall \tilde{x} \in S\}, & \text{if} \quad x \in S, \\ \emptyset, & \text{otherwise,} \end{cases}$$

is called the normal cone of the set S. When S is a linear subspace, then for all $x \in S$,

$$N_S(x) = \{x^* \in \mathcal{X}^* : \langle x^*, \tilde{x}\rangle = 0 \ \forall \tilde{x} \in S\} = S^\perp,$$

the latter denoting the orthogonal space of S.

An important role in the following will be played by the notion of conjugate functional.

Definition 3.1. *The conjugate functional of* $P : \mathcal{X} \longrightarrow \mathbb{R} \cup \{+\infty\}$ *is* $P^* : \mathcal{X}^* \longrightarrow \mathbb{R} \cup \{\pm\infty\}$ *defined as*

$$P^*(x^*) := \sup_{x\in\mathcal{X}}\{\langle x^*, x\rangle - P(x)\}, \qquad x^* \in \mathcal{X}^*.$$

The conjugate of P is convex and weak* lower semicontinuous and in the case P is proper, convex and lower semicontinuous, P^* takes values in $\mathbb{R} \cup \{+\infty\}$, being proper. More than that, according to the theorem of Fenchel–Moreau (see, for instance, [30, theorem 2.3.3]), one has under these hypotheses that $P(x) = P^{**}(x)$ for all $x \in \mathcal{X}$, where

$$P^{**} : \mathcal{X} \longrightarrow \mathbb{R} \cup \{\pm\infty\}, \qquad P^{**}(x) := \sup_{x^*\in\mathcal{X}^*}\{\langle x^*, x\rangle - P^*(x^*)\}, \ x \in \mathcal{X},$$

represents the biconjugate functional of P. As an immediate consequence of the definition, the following holds.

Lemma 3.1. *For arbitrary* $x \in \mathcal{X}$ *and* $x^* \in \mathcal{X}^*$*, we have the so-called Young–Fenchel inequality, i.e.*

$$\langle x^*, x\rangle \leqslant P(x) + P^*(x^*).$$

Moreover, equality holds, i.e.

$$\langle x^*, x\rangle = P(x) + P^*(x^*)$$

if and only if $x^* \in \partial P(x)$.

The following result is of interest too (see [30, theorems 2.4.2 and 2.4.4]).

Proposition 3.1. *Let $P : \mathcal{X} \longrightarrow \mathbb{R} \cup \{+\infty\}$ be given.*

(i) It holds: $x^* \in \partial P(x) \;\Rightarrow\; x \in \partial P^*(x^*)$.
(ii) If P is proper, convex and lower semicontinuous, then

$$x^* \in \partial P(x) \;\Leftrightarrow\; x \in \partial P^*(x^*).$$

According to statement (ii) of the above result, whenever P is proper, convex and lower semicontinuous, one has that $(\partial P)^{-1} = \partial P^*$. Hence, an appropriate choice for P is a proper, convex and lower semicontinuous functional having as subdifferential of its conjugate a single-valued operator. This is obviously the case when P^* is Gâteaux differentiable, a property which is definitively fulfilled for the class of functionals which we introduce in the following [30, section 3.5].

Definition 3.2. *Let $s \geqslant 2$. The functional $P : \mathcal{X} \longrightarrow \mathbb{R} \cup \{+\infty\}$ is called s-convex if there exists a constant $G_s > 0$ such that for $\rho : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$, $\rho(t) := \frac{G_s}{s} t^s$, one has*

$$P\left((1-\lambda)x + \lambda\tilde{x}\right) + \lambda(1-\lambda)\rho(\|x-\tilde{x}\|) \leqslant (1-\lambda)P(x) + \lambda P(\tilde{x})$$

for all $x, \tilde{x} \in \mathcal{X}$ and all $\lambda \in (0, 1)$.

The following characterization of s-convex functionals is taken from [30, corollary 3.5.11].

Theorem 3.1. *Let $P : \mathcal{X} \longrightarrow \mathbb{R} \cup \{+\infty\}$ be a proper, convex and lower semicontinuous functional and $1 < s^* \leqslant 2 \leqslant s < +\infty$ with $(s^*)^{-1} + s^{-1} = 1$. Then the following statements are equivalent:*

(i) P is s-convex;
(ii) there exists $C_1 > 0$ such that for all $x \in \mathcal{D}(\partial P)$, $x^* \in \partial P(x)$ and all $\tilde{x} \in \mathcal{X}$, we have

$$P(\tilde{x}) - P(x) - \langle x^*, \tilde{x} - x\rangle \geqslant \frac{C_1}{s}\|\tilde{x} - x\|^s;$$

(iii) there exists $C_2 > 0$ such that for all $x \in \mathcal{D}(\partial P)$, $x^* \in \partial P(x)$ and all $\tilde{x}^* \in \mathcal{X}$, we have

$$P^*(\tilde{x}^*) - P^*(x^*) - \langle x, \tilde{x}^* - x^*\rangle \leqslant \frac{C_2^{1-s^*}}{s^*}\|\tilde{x}^* - x^*\|^{s^*}; \tag{3}$$

(iv) dom $P^* = \mathcal{X}^*$*, P^* is Fréchet differentiable on $\mathcal{X}^*$ and there exists $C_3 > 0$ such that*

$$\|\nabla P^*(\tilde{x}^*) - \nabla P^*(x^*)\| \leqslant C_3^{1-s^*}\|\tilde{x}^* - x^*\|^{s^*-1} \tag{4}$$

for all $x^*, \tilde{x}^* \in \mathcal{X}^*$.

Remark 3.1. If P is s-convex with $\rho : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$, $\rho(t) := \frac{G_s}{s} t^s$ in definition 3.2, where $G_s > 0$, then one can take in the previous result $C_1 = C_2 := G_s$ and $C_3 := \frac{2G_s}{s}$.

Example 3.1. Assume that $\mathcal{X}$ is an s-convex space for some $s \in [2, +\infty)$. Then $P : \mathcal{X} \longrightarrow \mathbb{R}$, $P(x) := \frac{1}{s}\|x\|^s$ is a proper, lower semicontinuous and s-convex functional and for all $x^* \in \mathcal{X}^*$ one has $P^*(x^*) = \frac{1}{s^*}\|x^*\|^{s^*}$, where $s^{-1} + (s^*)^{-1} = 1$. These types of penalty functionals P were considered in [15]. We also note that L^q-spaces, $1 < q < +\infty$, are s-convex with $s = \max\{q, 2\}$.

4. Choice of the step-size parameter and the algorithm

Before we present the algorithm in detail, we summarize the basic assumptions which we will consider in the subsequent analysis.

(A1) $\mathcal{Y}$ is a smooth space.
(A2) $\mathcal{X}$ is a reflexive Banach space.
(A3) The functional $P : \mathcal{X} \longrightarrow \mathbb{R} \cup \{+\infty\}$ is proper, lower semicontinuous and s-convex (with $\rho : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$, $\rho(t) = \frac{G_s}{s} t^s$) for some exponent $2 \leqslant s < +\infty$.
(A4) There exists a solution $x^\dagger \in \operatorname{dom} P$ of equation (1), i.e. $A x^\dagger = y$ holds.

Due to theorem 3.1, one has that for all $x^* \in \mathcal{X}^*$, $(\partial P)^{-1}(x^*) = \partial P^*(x^*) = \{\nabla P^*(x^*)\}$ and thus the specific choice of $G := (\partial P)^{-1} = \nabla P^*$ from (2) provides a single-valued operator on its domain, which in this case is the whole space $\mathcal{X}^*$. This means that, in each iteration $n \geqslant 0$ of the regularization scheme, the element

$$x_{n+1}^\delta := \nabla P^*(x_n^* - \mu_n \phi_n^*)$$

is well defined for arbitrary choices $\mu_n \in \mathbb{R}$ and, according to proposition 3.1(ii), it holds

$$x_n^* - \mu_n \phi_n^* \in \partial P\big(x_{n+1}^\delta\big);$$

thus, $x_{n+1}^\delta \in \mathcal{D}(\partial P)$.

We now introduce Bregman distances, which have become a standard tool for the convergence analysis in Banach spaces.

Definition 4.1. *For given $x^* \in \mathcal{R}(\partial P)$, we define the Bregman distance $\Delta_{x^*}^P : \mathcal{X} \times (\partial P)^{-1}(x^*) \longrightarrow [0, +\infty]$ as being*

$$\Delta_{x^*}^P(\tilde{x}, x) := P(\tilde{x}) - P(x) - \langle x^*, \tilde{x} - x\rangle.$$

Using proposition 3.1(ii) one has for all $\tilde{x} \in \mathcal{X}$ and all $x \in (\partial P)^{-1}(x^*)$ that

$$\Delta_{x^*}^P(\tilde{x}, x) = P(\tilde{x}) + P^*(x^*) - \langle x^*, \tilde{x}\rangle.$$

Furthermore, for $\delta > 0$ and $n \geqslant 0$, let the nth iterate $x_n^\delta := \nabla P^*(x_n^*)$ be given. As given above, one has $x_n^* \in \partial P\big(x_n^\delta\big)$. We introduce the notation

$$\Delta_n := \Delta_{x_n^*}^P\big(x^\dagger, x_n^\delta\big) = P(x^\dagger) + P^*(x_n^*) - \langle x_n^*, x^\dagger\rangle \tag{5}$$

and, for $\mu > 0$,

$$\Delta_\mu := \Delta_{x_n^* - \mu\,\phi_n^*}^P(x^\dagger, \nabla P^*(x_n^* - \mu\,\phi_n^*)) = P(x^\dagger) + P^*(x_n^* - \mu\,\phi_n^*) - \langle x_n^* - \mu\,\phi_n^*, x^\dagger\rangle. \tag{6}$$

In order to determine a proper step size $\mu_n > 0$, we make the following evaluation:

$$\begin{aligned}\Delta_\mu - \Delta_n &= P^*(x_n^* - \mu\phi_n^*) - P^*(x_n^*) + \mu\langle\phi_n^*, x^\dagger - x_n^\delta + x_n^\delta\rangle\\ &= P^*(x_n^* - \mu\phi_n^*) - P^*(x_n^*) + \mu\big\langle J_p\big(Ax_n^\delta - y^\delta\big), y - y^\delta + y^\delta - Ax_n^\delta\big\rangle + \mu\langle\phi_n^*, x_n^\delta\rangle\\ &\leqslant P^*(x_n^* - \mu\phi_n^*) - P^*(x_n^*) - \mu\left(\|Ax_n^\delta - y^\delta\|^p - \|Ax_n^\delta - y^\delta\|^{p-1}\delta\right) + \mu\langle\phi_n^*, x_n^\delta\rangle.\end{aligned}$$

The term on the right-hand side of the above inequality can be seen as a function of μ. Hence, a natural choice for the step size μ_n would be to take it as the minimum of the function

$$f_n : \mathbb{R}_+ \longrightarrow \mathbb{R} \cup \{+\infty\}, \quad f_n(\mu) := P^*(x_n^* - \mu\,\phi_n^*) - \mu\,C_n^\delta + \mu\langle\phi_n^*, x_n^\delta\rangle, \tag{7}$$

in the case this exists, where

$$C_n^\delta := \big\|Ax_n^\delta - y^\delta\big\|^p - \|Ax_n^\delta - y^\delta\|^{p-1}\delta.$$

We refer the reader to section 6 for more details with respect to this idea.

On the other hand, we consider here a further estimate of $\Delta_\mu - \Delta_n$ by utilizing the s-convexity of P. More precisely, from theorem 3.1 we get

$$\begin{aligned}\Delta_\mu - \Delta_n &\leqslant P^*(x_n^* - \mu\phi_n^*) - P^*(x_n^*) - \mu C_n^\delta + \langle \mu\phi_n^*, x_n^\delta\rangle \\ &\leqslant -\mu C_n^\delta + \frac{G_s^{1-s^*}}{s^*}\|\phi_n^*\|^{s^*}\mu^{s^*}.\end{aligned}$$

We further assume that $C_n^\delta > 0$ and $\phi_n^* \neq 0$ and consider the following upper bound of $G_s^{1-s^*}\|\phi_n^*\|^{s^*}$:

$$\hat{C}_n^\delta := \max\left\{G_s^{1-s^*}\|\phi_n^*\|^{s^*}, C_n^\delta \bar{\mu}^{\frac{1}{1-s}}\|Ax_n^\delta - y^\delta\|^{\frac{p-s}{s-1}}\right\} > 0,$$

where $\bar{\mu} \in (0, +\infty]$ represents an *a priori* given upper bound for the step size. Hence, we get the following estimate:

$$\Delta_\mu - \Delta_n \leqslant -\mu C_n^\delta + \frac{\mu^{s^*}}{s^*}\hat{C}_n^\delta,$$

while the step size we choose will be the unique minimizer of the function

$$g_n : \mathbb{R}_+ \longrightarrow \mathbb{R}, \qquad g_n(\mu) := -\mu C_n^\delta + \frac{\mu^{s^*}}{s^*}\hat{C}_n^\delta.$$

This follows by an easy calculation and has the following formulation:

$$\begin{aligned}\mu_n = \left(\frac{C_n^\delta}{\hat{C}_n^\delta}\right)^{s-1} &= \min\left\{\left(\frac{C_n^\delta}{G_s^{1-s^*}\|\phi_n^*\|^{s^*}}\right)^{s-1}, \left(\bar{\mu}^{\frac{1}{s-1}}\|Ax_n^\delta - y^\delta\|^{\frac{s-p}{s-1}}\right)^{s-1}\right\} \\ &= \min\left\{\frac{(C_n^\delta)^{s-1}G_s}{\|\phi_n^*\|^s}, \bar{\mu}\|Ax_n^\delta - y^\delta\|^{s-p}\right\}.\end{aligned}$$

Hence, by denoting $\Delta_{n+1} := \Delta_{\mu_n}$, it holds

$$\Delta_{n+1} - \Delta_n = \Delta_{\mu_n} - \Delta_n \leqslant -\frac{1}{s}\frac{(C_n^\delta)^s}{(\hat{C}_n^\delta)^{s-1}} < 0. \tag{8}$$

Let us now present the algorithm under consideration in detail.

Algorithm 4.1.

(S0) *Initialization: choose the starting point $x_0^* \in \mathcal{X}^*$, $x_0 = x_0^\delta := \nabla P^*(x_0^*)$, an upper bound $\bar{\mu} \in (0, \infty]$ and a parameter $\tau > 1$. Set $n := 0$.*

(S1) *STOP: if for $\delta > 0$ the discrepancy criterion $\|A x_n^\delta - y^\delta\| \leqslant \tau\,\delta$ is fulfilled or we have $A x_n = y$ for $\delta = 0$.*

(S2) *Calculate*

$$\begin{aligned}\phi_n^* &:= A^\star J_p(Ax_n^\delta - y^\delta); \\ C_n^\delta &:= \|Ax_n^\delta - y^\delta\|^{p-1}(\|Ax_n^\delta - y^\delta\| - \delta); \\ \mu_n &:= \min\left\{\frac{(C_n^\delta)^{s-1}G_s}{\|\phi_n^*\|^s}, \bar{\mu}\|Ax_n^\delta - y^\delta\|^{s-p}\right\}.\end{aligned}$$

(S3) *Calculate the new iterate*

$$\begin{aligned}x_{n+1}^* &:= x_n^* - \mu_n\phi_n^*; \\ x_{n+1}^\delta &:= \nabla P^*(x_{n+1}^*).\end{aligned}$$

Set $n := n + 1$ and go to step (S1).

Remark 4.1. One can note that, if for $n \geqslant 0$ the stopping criterion $\|A x_n^\delta - y^\delta\| \leqslant \tau\,\delta$ for $\delta > 0$ and $A x_n = y$ for $\delta = 0$ is not fulfilled, then we have $C_n^\delta > 0$. Furthermore, we can choose τ arbitrarily close to 1. Moreover, it holds $\phi_n^* \neq 0$. Indeed, assuming the contrary, one would have that $x_n^\delta \in \operatorname{argmin} \frac{1}{p}\|A(\cdot) - y^\delta\|^p = \operatorname{argmin}\|A(\cdot) - y^\delta\|$. Thus,

$$\left\|Ax_n^\delta - y^\delta\right\| \leqslant \|Ax^\dagger - y^\delta\| = \|y - y^\delta\| \leqslant \delta,$$

which contradicts the fact that $\left\|A\,x_n^\delta - y^\delta\right\| > \tau\,\delta$. Consequently, algorithm 4.1 is well defined.

Furthermore, for $\delta > 0$, we denote by $N(\delta, y^\delta)$ the index on which the iteration process stops, namely

$$\left\|Ax_{N(\delta,y^\delta)}^\delta - y^\delta\right\| \leqslant \tau\delta < \left\|Ax_n^\delta - y^\delta\right\| \quad \text{for} \quad 0 \leqslant n < N(\delta, y^\delta).$$

The existence of such an index, whenever $\delta > 0$, will be shown in the following.

One can also note that according to (8) whenever $0 < N(\delta, y^\delta)$, one has for all $0 \leqslant n < N(\delta, y^\delta)$ that

$$\Delta^P_{x^*_{n+1}}\left(x^\dagger, x^\delta_{n+1}\right) - \Delta^P_{x^*_n}\left(x^\dagger, x^\delta_n\right) = \Delta_{n+1} - \Delta_n < 0;$$

hence,

$$\Delta^P_{x^*_{n+1}}\left(x^\dagger, x^\delta_{n+1}\right) < \Delta^P_{x^*_n}\left(x^\dagger, x^\delta_n\right).$$

We want to emphasize that this result holds for the arbitrary solution $x^\dagger$ of equation (1).

The proof of the following preliminary result follows on the lines of the one given for [15, lemma 4.1].

Lemma 4.1. *Assume that (A1)–(A4) are fulfilled and that for $\delta > 0$ algorithm 4.1 stops with index $N(\delta, y^\delta) > 0$. Then, for all $0 \leqslant n < N(\delta, y^\delta)$, the following statements are true.*

(i) If $\delta > 0$, then

$$\mu_n \in \left[\min\left\{\frac{(1-\tau^{-1})^{s-1}G_s}{\|A\|^s}, \bar{\mu}\right\}, \bar{\mu}\right] \left\|Ax_n^\delta - y^\delta\right\|^{s-p}$$

and

$$\begin{aligned} -g_n(\mu_n) &\geqslant \frac{1-\tau^{-1}}{s}\mu_n\left\|Ax_n^\delta - y^\delta\right\|^p \\ &\geqslant \frac{1-\tau^{-1}}{s}\min\left\{\frac{(1-\tau^{-1})^{s-1}G_s}{\|A\|^s}, \bar{\mu}\right\}\left\|Ax_n^\delta - y^\delta\right\|^s. \end{aligned}$$

(ii) If $\delta = 0$, then

$$\mu_n \in \left[\min\left\{\frac{G_s}{\|A\|^s}, \bar{\mu}\right\}, \bar{\mu}\right] \|Ax_n - y\|^{s-p}$$

and

$$-g_n(\mu_n) = \frac{1}{s}\mu_n\|Ax_n - y\|^p \geqslant \frac{1}{s}\min\left\{\frac{G_s}{\|A\|^s}, \bar{\mu}\right\}\|Ax_n - y\|^s.$$

We apply these results for proving the following.

Lemma 4.2. *Assume that (A1)–(A4) are fulfilled, let $x_0^* \in \mathcal{X}^*$ be the starting point of algorithm 4.1 and let $\left\{x_n^\delta\right\}_{n\geqslant 0}$ be the sequence generated by it, for $\delta \geqslant 0$. The following assertions are true.*

(i) *For $\delta > 0$, the algorithm stops after a finite number $N(\delta, y^\delta)$ of iterations and there exists a constant $C > 0$ (not depending on δ) such that*

$$N(\delta, y^\delta) \leqslant C\delta^{-s}.$$

If $N := N(\delta, y^\delta) > 0$, then there exist constants $C_\tau, \tilde{C}_\tau > 0$ such that

$$\sum_{n=0}^{N-1} \mu_n \|Ax_n^\delta - y^\delta\|^p \leqslant C_\tau \Delta_{x_0^*}^P(x^\dagger, x_0^\delta) \qquad \text{and} \qquad \sum_{n=0}^{N-1} \|Ax_n^\delta - y^\delta\|^s \leqslant \tilde{C}_\tau \Delta_{x_0^*}^P(x^\dagger, x_0^\delta).$$

(ii) *For $\delta = 0$, denoting by $N := N(0, y)$ the index where algorithm 4.1 stops (the value $N = +\infty$ is here also allowed), if $N > 0$, there exist constants $C_0, \tilde{C}_0 > 0$ such that*

$$\sum_{n=0}^{N-1} \mu_n \|Ax_n - y\|^p \leqslant C_0 \Delta_{x_0^*}^P(x^\dagger, x_0) \qquad \text{and} \qquad \sum_{n=0}^{N-1} \|Ax_n - y\|^s \leqslant \tilde{C}_0 \Delta_{x_0^*}^P(x^\dagger, x_0).$$

Proof.

(i) Let $\delta > 0$. Assuming that algorithm 4.1 does not stop after a finite number of iterations, one has, for all $k > 0$,

$$\begin{aligned} \Delta_{x_0^*}^P(x^\dagger, x_0^\delta) &\geqslant \Delta_{x_0^*}^P(x^\dagger, x_0^\delta) - \Delta_{x_k^*}^P(x^\dagger, x_k^\delta) \\ &= \sum_{n=0}^{k-1} \left(\Delta_{x_n^*}^P(x^\dagger, x_n^\delta) - \Delta_{x_{n+1}^*}^P(x^\dagger, x_{n+1}^\delta)\right) \geqslant -\sum_{n=0}^{k-1} g_n(\mu_n). \end{aligned} \tag{9}$$

Using lemma 4.1(i), one further gets, for all $k > 0$,

$$\Delta_{x_0^*}^P(x^\dagger, x_0^\delta) \geqslant \frac{1-\tau^{-1}}{s} \min\left\{\frac{(1-\tau^{-1})^{s-1}G_s}{\|A\|^s}, \bar{\mu}\right\} k\tau^s\delta^s,$$

which leads to a contradiction. Hence, $N(\delta, y^\delta)$ exists; it is finite and, for

$$C := \frac{\Delta_{x_0^*}^P(x^\dagger, x_0^\delta)}{\frac{1-\tau^{-1}}{s} \min\left\{\frac{(1-\tau^{-1})^{s-1}G_s}{\|A\|^s}, \bar{\mu}\right\}\tau^s},$$

the inequality $N(\delta, y^\delta) \leqslant C\,\delta^{-s}$ is fulfilled. Assuming that $N = N(\delta, y^\delta) > 0$, from (9) and lemma 4.1(i), one also has

$$\Delta_{x_0^*}^P(x^\dagger, x_0^\delta) \geqslant -\sum_{n=0}^{N-1} g_n(\mu_n) \geqslant \frac{1-\tau^{-1}}{s} \sum_{n=0}^{N-1} \mu_n \|Ax_n^\delta - y^\delta\|^p$$

and

$$\begin{aligned} \Delta_{x_0^*}^P(x^\dagger, x_0^\delta) &\geqslant -\sum_{n=0}^{N-1} g_n(\mu_n) \\ &\geqslant \frac{1-\tau^{-1}}{s} \sum_{n=0}^{N-1} \min\left\{\frac{(1-\tau^{-1})^{s-1}G_s}{\|A\|^s}, \bar{\mu}\right\} \|Ax_n^\delta - y^\delta\|^s, \end{aligned}$$

which proves assertion (i).

(ii) Let $\delta = 0$. In analogy to (9), one has, for all $k > 0$,

$$\Delta^P_{x_0^*}(x^\dagger, x_0) \geqslant -\sum_{n=0}^{k-1} g_n(\mu_n)$$

and, via lemma 4.1(ii), we further have

$$\Delta^P_{x_0^*}(x^\dagger, x_0) \geqslant \frac{1}{s}\sum_{n=0}^{k-1} \mu_n \|Ax_n - y\|^p$$

and

$$\Delta^P_{x_0^*}(x^\dagger, x_0) \geqslant \frac{1}{s}\min\left\{\frac{G_s}{\|A\|^s}, \bar{\mu}\right\}\sum_{n=0}^{k-1} \mu_n \|Ax_n - y\|^s.$$

From here the conclusion follows if both a finite stopping index $N = N(0, y)$ exists and if the algorithm does not stop. □

Remark 4.2. Whenever in the previous result one has for $\delta \geqslant 0$ that $N(\delta, y^\delta) > 0$, it holds that $\Delta^P_{x_0^*}(x^\dagger, x_0^\delta) > 0$. Indeed, otherwise one would have that $x_0^* \in \partial P(x^\dagger) \Leftrightarrow x^\dagger = \nabla P^*(x_0^*) = x_0^\delta$. In this case, for $\delta > 0$, the discrepancy criterion $\|Ax_0^\delta - y^\delta\| \leqslant \tau\delta$ would be fulfilled, while for $\delta = 0$ it would hold $Ax_0 = y$. Hence, the algorithm would stop in both cases with $N(\delta, y^\delta) = 0$.

5. Convergence results

We discuss the convergence properties of the algorithm and start with the noiseless case $\delta = 0$. We omit giving the proof of the following result, as it follows in analogy to the one of theorem 5.1 in [15], by decisively using the s-convexity of the penalty functional P and the statements in lemma 4.1(ii).

Theorem 5.1. *Assume that (A1)–(A4) are fulfilled and let $\delta = 0$. Then algorithm 4.1 stops either after a finite number $N := N(0, y)$ of iterations with x_N satisfying $Ax_N = y$ or the sequence $\{x_n\}_{n\geqslant 0}$ converges to a solution of (1).*

Next we give a characterization of the limit point of the sequence $\{x_n\}_{n\geqslant 0}$ generated by algorithm 4.1 when $\delta = 0$ in the case it does not stop after a finite number of iterations. In the following result, $\mathcal{N}(A) := \{x \in \mathcal{X} : Ax = 0\}$ denotes the kernel of the linear continuous operator A.

Theorem 5.2. *Assume that (A1)–(A4) are fulfilled, take $x_0^* \in \mathcal{X}^*$ and $x_0 := \nabla P^*(x_0^*) \in \mathcal{X}$.*

(i) The minimization problem

$$\inf \Delta^P_{x_0^*}(x, x_0) \quad \textit{subject to} \quad Ax = y \tag{10}$$

has a unique optimal solution $\bar{x}$ which fulfills, if $\mathrm{int}(\mathrm{dom}\, P) \cap \{x \in \mathcal{X} : Ax = y\} \neq \emptyset$,

$$x_0^* \in \partial P(\bar{x}) + \mathcal{N}(A)^\perp. \tag{11}$$

(ii) If, for $\delta = 0$, algorithm 4.1 having as a starting point $x_0^* \in \mathcal{X}^*$ does not stop after a finite number of iterations and the sequence $\{x_n\}_{n\geqslant 0}$ generated by it converges to an element belonging to $\mathrm{int}(\mathrm{dom}\, P)$, then this limit is nothing else than the unique optimal solution of (10).

Proof.

(i) Denote by $\gamma := \inf\{\Delta^P_{x_0^*}(x, x_0) : Ax = y\} \in [0, +\infty)$. Then for all $k \geqslant 1$ there exists $z_k \in \mathcal{X}$ such that $Az_k = y$ and

$$\gamma \leqslant \Delta^P_{x_0^*}(z_k, x_0) < \gamma + \frac{1}{k}.$$

By theorem 3.1 one has that $\frac{G_s}{s}\|z_k - x_0\|^s \leqslant \gamma + 1$ for all $k \geqslant 1$; thus, $\{z_k\}_{k\geqslant 1}$ is bounded. Then there exists a subsequence $\{z_{k_l}\}_{l\geqslant 1}$ which converges to $\bar{x} \in \mathcal{X}$ in the weak topology of $\mathcal{X}$ and, since $A^{-1}(\{y\}) := \{x \in \mathcal{X} : Ax = y\}$ is convex and (weakly) closed, it follows that $A\bar{x} = y$. Using the (weak) lower semicontinuity of P, it holds

$$\begin{aligned}\gamma &\geqslant \liminf_{l\to+\infty} \Delta^P_{x_0^*}(z_{k_l}, x_0)\\ &= \liminf_{l\to+\infty}(P(z_{k_l}) - P(x_0) - \langle x_0^*, z_{k_l} - x_0\rangle)\\ &\geqslant P(\bar{x}) - P(x_0) - \langle x_0^*, \bar{x} - x_0\rangle = \Delta^P_{x_0^*}(\bar{x}, x_0) \geqslant \gamma,\end{aligned}$$

which means that $\bar{x}$ is an optimal solution of (10). The uniqueness of $\bar{x}$ follows from the s-convexity of P. Thus,

$$0 \in \partial\big(\Delta^P_{x_0^*}(\cdot, x_0) + \delta_{A^{-1}(\{y\})}\big)(\bar{x}).$$

Since $\operatorname{int}(\operatorname{dom}\Delta^P_{x_0^*}(\cdot, x_0)) \cap A^{-1}(\{y\}) = \operatorname{int}(\operatorname{dom} P) \cap A^{-1}(\{y\}) \neq \emptyset$, by [5, theorem 7.5], one has, equivalently, that

$$0 \in \partial\Delta^P_{x_0^*}(\cdot, x_0)(\bar{x}) + N_{A^{-1}(\{y\})}(\bar{x}) = \partial P(\bar{x}) - x_0^* + N_{A^{-1}(\{y\})}(\bar{x}),$$

which is further equivalent to

$$x_0^* \in \partial P(\bar{x}) + N_{A^{-1}(\{y\})}(\bar{x}).$$

For the normal cone $N_{A^{-1}(\{y\})}(\bar{x})$, we have the following representation:

$$N_{A^{-1}(\{y\})}(\bar{x}) = N_{\bar{x}+\mathcal{N}(A)}(\bar{x}) = \{x^* \in \mathcal{X}^* : \langle x^*, z\rangle \leqslant 0,\ \ \forall z \in \mathcal{N}(A)\} = \mathcal{N}(A)^\perp$$

and in this way relation (11), namely

$$x_0^* \in \partial P(\bar{x}) + \mathcal{N}(A)^\perp$$

follows. We proved actually more, namely that $\bar{x} \in \operatorname{dom} P \cap A^{-1}(\{y\})$ is an optimal solution of (10) if and only if (11) holds.

(ii) Let $\tilde{x} \in \operatorname{int}(\operatorname{dom} P)$ such that $A\tilde{x} = y$ and $x_n \to \tilde{x}$ as $n \to +\infty$. According to algorithm 4.1, one has for all $n \geqslant 0$ that $x_n^* - x_0^* \in \mathcal{R}(A^*)$ and $x_n = \nabla P^*(x_n^*)$, which is equivalent to $x_n^* \in \partial P(x_n)$. Since $\tilde{x} \in \operatorname{int}(\operatorname{dom} P)$, one has that ∂P is locally bounded in $\tilde{x}$ (see [22]) and this means that $\{x_n^*\}_{n\geqslant 0}$ is bounded. Thus, there exists a subsequence $\{x_{n_l}^*\}_{l\geqslant 0}$, which converges to an element $\tilde{x}^* \in \mathcal{X}^*$ in the weak* topology of $\mathcal{X}^*$. As ∂P is norm-to-weak* upper semicontinuous at $\tilde{x}$ (see [22]), it holds $\tilde{x}^* \in \partial P(\tilde{x})$. Thus, $\tilde{x}^* - x_0^* \in \overline{\mathcal{R}(A^*)}^{w^*} = \mathcal{N}(A)^\perp$, which implies that $x_0^* \in \partial P(\tilde{x}) + \mathcal{N}(A)^\perp$. According to the proof of item (i), $\tilde{x}$ is the unique optimal solution of (10). □

In order to show that algorithm 4.1 describes in fact a regularization method we replace the smoothness assumption on $\mathcal{Y}$ by the following stronger one:

(A1′) The space $\mathcal{Y}$ is uniformly smooth.

Then we can prove the following.

Theorem 5.3. *Assume that (A1′), (A2)–(A4) are fulfilled and that, for $\delta = 0$, algorithm 4.1 does not stop after a finite number of iterations and the sequence $\{x_n\}_{n\geqslant 0}$ generated by it converges to $\bar{x} \in \mathrm{int}(\mathrm{dom}\,P)$. If $\{x_n^\delta\}_{n\geqslant 0}$ is the sequence generated by algorithm 4.1 for $\delta > 0$, then it holds $x^\delta_{N(\delta,y^\delta)} \to \bar{x}$ as $\delta \to 0$.*

Proof. We change the notation and write $x_n^{*^\delta}$ for the iterates in $\mathcal{X}^*$ when working with noisy data y^δ and x_n^* when working with exact data y. Since J_p is norm-to-norm uniformly continuous on bounded subsets of $\mathcal{Y}$, one can see that for all $n \geqslant 0$, as the step size μ_n depends continuously on δ (see (S2) in algorithm 4.1), $x_n^\delta \to x_n$ and $x_n^{*^\delta} \to x_n^*$ as $\delta \to 0$. By assumptions, one has that $N(\delta, y^\delta) \to +\infty$ as $\delta \to 0$. Let $n \geqslant 0$ be a fixed index. Then for all $\delta > 0$ such that $n < N(\delta, y^\delta)$, one has, by theorem 3.1, that

$$\begin{aligned}\frac{G_s}{s}\left\|\bar{x} - x^\delta_{N(\delta,y^\delta)}\right\|^s &\leqslant \Delta^P_{x^{*^\delta}_{N(\delta,y^\delta)}}\left(\bar{x}, x^\delta_{N(\delta,y^\delta)}\right)\\ &< \Delta^P_{x_n^{*^\delta}}\left(\bar{x}, x_n^\delta\right) = P(\bar{x}) - P\left(x_n^\delta\right) - \left\langle x_n^{*^\delta}, \bar{x} - x_n^\delta\right\rangle.\end{aligned}$$

Let $\delta \to 0$ and, so

$$\limsup_{\delta\to 0} \frac{G_s}{s}\left\|\bar{x} - x^\delta_{N(\delta,y^\delta)}\right\|^s \leqslant P(\bar{x}) - P(x_n) - \langle x_n^*, \bar{x} - x_n\rangle. \tag{12}$$

Thus, (12) holds for all $n \geqslant 0$. Furthermore, as $\bar{x} \in \mathrm{int}(\mathrm{dom}\,P)$ and ∂P is locally bounded and norm-to-weak* upper semicontinuous at $\bar{x}$, there exists a subsequence $\{x^*_{n_l}\}_{l\geqslant 0}$ converging to $\bar{x}^*$ in the weak* topology of $\mathcal{X}^*$ such that $\bar{x}^* \in \partial P(\bar{x})$. Thus, due to (12), for all $l \geqslant 0$,

$$\limsup_{\delta\to 0} \frac{G_s}{s}\left\|\bar{x} - x^\delta_{N(\delta,y^\delta)}\right\|^s \leqslant P(\bar{x}) - P(x_{n_l}) - \langle x^*_{n_l}, \bar{x} - x_{n_l}\rangle.$$

We let l converge to $+\infty$ which leads to

$$\limsup_{\delta\to 0} \frac{G_s}{s}\left\|\bar{x} - x^\delta_{N(\delta,y^\delta)}\right\|^s \leqslant 0.$$

Consequently, $x^\delta_{N(\delta,y^\delta)} \to \bar{x}$ as $\delta \to 0$. This concludes the proof. □

Example 5.1. Assume that (A1) is fulfilled, $\mathcal{X}$ is an s-convex space for some $1 < s < +\infty$ and (1) has a solution. Then $P : \mathcal{X} \longrightarrow \mathbb{R}$, $P(x) := \frac{1}{s}\|x - x_\sharp\|^s$ for $x_\sharp \in \mathcal{X}$ an *a priori* guess fulfills (A2). For all $x^* \in \mathcal{X}^*$ it holds $\nabla P^*(x^*) = x_\sharp + J^{\mathcal{X}^*}_{s^*}(x^*)$, where $J^{\mathcal{X}^*}_{s^*} : \mathcal{X}^* \longrightarrow \mathcal{X}$ denotes the corresponding duality mapping with the gauge function $t \mapsto t^{s^*-1}$ and $s^{-1} + (s^*)^{-1} = 1$. We set $x_0^* := 0$. Then $x_0 = x_\sharp$ and

$$\Delta^P_{x_0^*}(x, x_0) = P(x) - P(x_0) - \langle x_0^*, x - x_0\rangle = P(x) = \frac{1}{s}\|x - x_\sharp\|^s.$$

Hence, for $\delta = 0$, for this choice of the penalty functional the sequence $\{x_n\}_{n\geqslant 0}$ converges to the $x_\sharp$-minimum-norm solution of equation (1), provided the algorithm does not stop after a finite number of iterations.

6. On an accelerated approach

In this section, we shortly discuss an accelerated version of algorithm 4.1, for which the choice of the step size is done by minimizing on a certain interval the function $f_n : \mathbb{R}_+ \longrightarrow \mathbb{R}$,

$$f_n(\mu) := P^*(x_n^* - \mu\phi_n^*) - \mu C_n^\delta + \mu\langle x_n^*, x_n^\delta\rangle,$$

already introduced in (7). This gives the rise to the following algorithm.

Algorithm 6.1.

(S0) Initialization: choose the starting point $x_0^* \in \mathcal{X}^*$, $x_0 = x_0^\delta := \nabla P^*(x_0^*)$, *an upper bound* $\bar{\mu} \in (0, +\infty)$ *and a parameter* $\tau > 1$. *Set* $n := 0$.

(S1) STOP: when $\delta > 0$, *if the discrepancy criterion* $\|A\,x_n^\delta - y^\delta\| \leqslant \tau\,\delta$ *is fulfilled or, when* $\delta = 0$, *if* $A\,x_n = y$.

(S2) Calculate

$$\begin{aligned} \phi_n^* &:= A^\star J_p(Ax_n^\delta - y^\delta);\\ C_n^\delta &:= \|Ax_n^\delta - y^\delta\|^{p-1}\left(\|Ax_n^\delta - y^\delta\| - \delta\right);\\ \bar{\mu}_n &:= \bar{\mu}\|Ax_n^\delta - y^\delta\|^{s-p}. \end{aligned}$$

(S3) If $f_n'(\bar{\mu}_n) < 0$, *set* $\mu_n := \bar{\mu}_n$. *Otherwise, take* μ_n *as being the greatest* $\mu \in (0, \bar{\mu}_n]$ *such that* $f_n'(\mu) = 0$.

(S4) Calculate the new iterate

$$\begin{aligned} x_{n+1}^* &:= x_n^* - \mu_n\phi_n^*;\\ x_{n+1}^\delta &:= \nabla P^*(x_{n+1}^*). \end{aligned}$$

Set $n := n+1$ *and go to step (S1).*

Remark 6.1. Assuming that (A1)–(A4) are fulfilled, according to remark 4.1, if for $n \geqslant 0$ algorithm 6.1 does not stop, then $C_n^\delta > 0$ and $\phi_n^* \neq 0$. For all $\mu \in \mathbb{R}$, it holds

$$f_n'(\mu) = -\langle \phi_n^*, \nabla P^*(x_n^* - \mu\phi_n^*) - x_n^\delta\rangle - C_n^\delta$$

and, so $f_n'(0) = -C_n^\delta < 0$. Due to the fact that ∇P^* is Lipschitz continuous, f_n' is continuous and one can easily see that f_n' is increasing on $[0, +\infty)$. Consequently, in (S3), μ_n is taken as a minimizer of f_n on $[0, \bar{\mu}_n]$. It is worthwhile to note that, when $f_n'(\bar{\mu}_n) \geqslant 0$, the function can have more than one minimum on this interval.

By denoting with $\tilde{\mu}_n$ the minimizer of g_n on $[0, +\infty)$, which is in fact the step size considered in algorithm 4.1, and noting that $\tilde{\mu}_n \in (0, \bar{\mu}_n]$, one has

$$\begin{aligned} \Delta_{x_{n+1}^*}^P\left(x^\dagger, x_{n+1}^\delta\right) - \Delta_{x_n^*}^P\left(x^\dagger, x_n^\delta\right) &\leqslant f_n(\mu_n) - P^*(x_n^*)\\ &\leqslant f_n(\tilde{\mu}_n) - P^*(x_n^*) \leqslant g_n(\tilde{\mu}_n). \end{aligned}$$

Thus, according to lemma 4.2, when $\delta > 0$ the algorithm stops after a finite number of iterations $N(\delta, y^\delta)$, which fulfills $N(\delta, y^\delta) \leqslant C\,\delta^{-s}$ for a positive constant $C > 0$, while in the case $N(\delta, y^\delta) > 0$, there exists a constant $\tilde{C}_\tau$ such that

$$\sum_{n=0}^{N(\delta,y^\delta)} \left\|Ax_n^\delta - y^\delta\right\|^s \leqslant \tilde{C}_\tau\,\Delta_{x_0^*}^P\left(x^\dagger, x_0^\delta\right).$$

When $\delta = 0$, denoting by $N := N(0, y)$ the index where algorithm 6.1 stops (the value $N = +\infty$ is here also allowed), in the case $N > 0$, there exists a constant $\tilde{C}_0 > 0$ such that

$$\sum_{n=0}^{N-1} \|Ax_n - y\|^s \leqslant \tilde{C}_0\Delta_{x_0^*}^P(x^\dagger, x_0).$$

Due to this fact, theorems 5.1 and 5.2 remain valid for algorithm 6.1, too. Unfortunately, we are not aware if this applies also for theorem 5.3, as the continuous dependence of the step size μ_n considered in algorithm 6.1 on δ is at this moment an open question.

7. Applications and numerical results

Taking a closer look at algorithm 4.1, one can see that for $\delta \geqslant 0$ the determination in step (S3) of x_{n+1}^δ via

$$x_{n+1}^\delta := \nabla P^*(x_{n+1}^*), \tag{13}$$

for $n \geqslant 0$, implies knowledge of the conjugate functional P^* and of its Gâteaux gradient ∇P^*. Alternatively, one can try to calculate x_{n+1}^δ as follows. One has

$$\begin{aligned} x_{n+1}^\delta = \nabla P^*(x_{n+1}^*) &\Leftrightarrow x_{n+1}^* \in \partial P\big(x_{n+1}^\delta\big) \\ &\Leftrightarrow 0 \in \partial(P - \langle x_{n+1}^*, \cdot\rangle)\big(x_{n+1}^\delta\big) \\ &\Leftrightarrow x_{n+1}^\delta = \arg\min\{P(x) - \langle x_{n+1}^*, x\rangle\}. \end{aligned}$$

Thus, x_{n+1}^δ can be determined as the unique minimizer of the functional

$$x \mapsto P(x) - \langle x_{n+1}^*, x\rangle = P(x) - \langle x_n^*, x\rangle + \mu_n\langle \phi_n^*, x\rangle.$$

Remark 7.1. Assume $\delta = 0$. By considering the finite-dimensional setting $\mathcal{X} = \mathbb{R}^m$ and $\mathcal{Y} = \mathbb{R}^k$ with $m > k$, and constant step size $\mu_n \equiv 1$, the determination of x_{n+1} as the unique minimizer of

$$x \mapsto P(x) - \langle x_{n+1}^*, x\rangle + \frac{1}{2\alpha}\|x - x_n\|^2$$

for $\alpha > 0$ (see for instance [29] and the references therein) gives rise to the so-called *linearized Bregman method* for solving the constraint minimization problem:

$$\inf P(x) \quad \text{subject to} \quad Ax = y.$$

For a more involved version of this, we refer to [28], where an additional control of the step size μ_n was applied.

We consider next two examples which are of interest in the field of application of regularization approaches.

7.1. Sparse reconstruction

For $\Omega \subset \mathbb{R}^d$ a bounded domain and $\mathcal{X} := L^2(\Omega)$ one can consider as penalty functional $P_\beta : L^2(\Omega) \longrightarrow \mathbb{R}$,

$$P_\beta(x) := \|x\|_{L^1(\Omega)} + \frac{1}{2\beta}\|x\|_{L^2(\Omega)}^2, \tag{14}$$

where $\beta > 0$. Obviously, P_β is 2-convex with $G_2 = \beta^{-1}$. As given above, for $x^* \in L^2(\Omega)$ one has that

$$\begin{aligned} \nabla P_\beta^*(x^*) &= \arg\min_{x\in L^2(\Omega)}\left\{\|x\|_{L^1(\Omega)} + \frac{1}{2\beta}\|x\|_{L^2(\Omega)}^2 - \langle x^*, x\rangle\right\} \\ &= \arg\min_{x\in L^2(\Omega)}\left\{\beta\|x\|_{L^1(\Omega)} + \frac{1}{2}\|x - \beta x^*\|_{L^2(\Omega)}^2\right\} \\ &= \begin{cases} \beta(x^*(t) - 1), & \text{if} \quad x^*(t) > 1 \\ 0, & \text{if} \quad |x^*(t)| \leqslant 1 \qquad \text{a.e. on} \quad \Omega. \\ \beta(x^*(t) + 1), & \text{if} \quad x^*(t) < -1 \end{cases} \end{aligned}$$

The operator ∇P_β^* is a version of the so-called *soft-threshold (shrinkage) operator*, which has been applied in several fields for sparse reconstruction.

Remark 7.2. Assuming additionally that $\mathcal{Y}$ is a Hilbert space, via

$$x_{n+1}^{\delta} := \nabla P_{\beta}^{*}\left(\frac{1}{\beta}\big(x_n^{\delta} - A^{*}\big(Ax_n^{\delta} - y^{\delta}\big)\big)\right),$$

one introduces the so-called *iterative soft-threshold algorithm* (see [11]), which is widely used in sparse reconstruction for minimizing the Tikhonov functional:

$$T_{\beta}^{\delta}(x) := \tfrac{1}{2}\|Ax - y^{\delta}\|^2 + \beta\|x\|_{L^1(\Omega)}.$$

This corresponds to step (S3) in algorithm 4.1, by identifying x_n^{*} with x_n^{δ} and by taking as step size $\mu_n \equiv 1$, for $n \geqslant 0$. The sequence $\{x_n^{\delta}\}_{n\geqslant 0}$ converges to a minimizer x_{α}^{δ} of T_{α}^{δ}, even if the constant step size provides slow convergence for this algorithm.

The above remark points out the following: instead of minimizing a Tikhonov functional several times for different regularization parameters $\alpha > 0$, we suggest here an iterative regularization scheme with almost the same numerical amount in each iteration step, which promises faster convergence because of the step size control and for which only one incomplete minimization is applied. This observation emphasizes the chances of saving numerical costs by applying the presented iterative regularization approach.

7.2. TV regularization

For $\Omega \subset \mathbb{R}^d$ a bounded domain we denote by $TV : L^2(\Omega) \longrightarrow \mathbb{R} \cup \{+\infty\}$ the extension of the total variation from $BV(\Omega)$ (see [2]) to $\mathcal{X} := L^2(\Omega)$, by defining it as being equal to $+\infty$ for $x \in L^2(\Omega)\backslash BV(\Omega)$. For $\beta > 0$, the penalty functional $P_{\beta} : L^2(\Omega) \longrightarrow \mathbb{R} \cup \{+\infty\}$,

$$P_{\beta}(x) := TV(x) + \frac{1}{2\beta}\|x\|_{L^2(\Omega)}^2, \tag{15}$$

fits into the framework considered in this paper, being proper, lower semicontinuous and 2-convex with $G_2 = \beta^{-1}$. As opposed to the previous example, ∇P_{β}^{*} here is not explicitly known. Nevertheless, as given above, one can determine x_{n+1}^{δ}, for $\delta \geqslant 0$ and $n \geqslant 0$ as being

$$x_{n+1}^{\delta} = \arg\min\left\{TV(x) + \frac{1}{2\beta}\|x\|_{L^2(\Omega)}^2 - \langle x_{n+1}^{*}, x\rangle\right\},$$

which is again equivalent to

$$x_{n+1}^{\delta} = \arg\min\left\{\beta TV(x) + \tfrac{1}{2}\|x - \beta x_{n+1}^{*}\|_{L^2(\Omega)}^2\right\}. \tag{16}$$

This is the well-known ROF model (see [24]) in image denoising, while for solving this minimization problem there exists a various number of algorithms, like, for example, the projected gradient method of [10] and its acceleration FPG [4]. At first glance, it seems not to be very attractive to apply the minimization (16) in each iteration step. However, first of all, one can note that the operator A does not occur in this minimization problem, which means that the numerical effort for solving it is not that high. On the other hand, even modern algorithms such as ISTA (see [11]) and its acceleration FISTA (see [4]) for determining a minimizer of the Tikhonov functional

$$T_{\beta}^{\delta}(x) := \beta TV(x) + \tfrac{1}{2}\|Ax - y^{\delta}\|^2,$$

for $\beta > 0$, apply a solution of the ROF model (16) in each iteration step.

Table 1. Reconstruction errors for the sample function $x_1^\dagger$.

	$\beta = 1$		$\beta = 100$		$\beta = 10\,000$	
δ_{rel}	$N(\delta, y^\delta)$	$\frac{\|x_N^\delta - x_1^\dagger\|}{\|x_1^\dagger\|}$	$N(\delta, y^\delta)$	$\frac{\|x_N^\delta - x_1^\dagger\|}{\|x_1^\dagger\|}$	$N(\delta, y^\delta)$	$\frac{\|x_N^\delta - x_1^\dagger\|}{\|x_1^\dagger\|}$
0.01	759	0.2115	1528	0.1901	134 352	0.4391
10^{-3}	8203	0.0596	12 314	0.0581	260 545	0.0639
10^{-4}	20 217	0.0070	30 803	0.0069	239 080	0.0071

7.3. Numerical results

We shortly recall the situation. Motivated by the above considerations, we set $\mathcal{X} = \mathcal{Y} = L^2(0, 1)$ and deal with the linear benchmark operator of integration, e.g., $A : \mathcal{X} \longrightarrow \mathcal{Y}$ is given as

$$[Ax](t) := \int_0^t x(\tau)\, \mathrm{d}\tau, \qquad t \in [0, 1].$$

We set $p = 2$ and apply an equidistant discretization with $K = 1000$ subintervals. Let $\varphi_j = \chi_{(t_{j-1}, t_j)}$, $1 \leqslant j \leqslant K$, with $t_j := j/K$, $0 \leqslant j \leqslant K$, describe the piecewise constant ansatz functions. Then we approximate

$$x(t) \approx \sum_{j=1}^K x_j \varphi_j(t) \quad \text{and} \quad y(t) \approx \sum_{j=1}^K y_j \varphi_j(t), \quad t \in [0, 1].$$

For the discretization of the data $y \in \mathcal{Y}$, we can choose the functional values of $y \in \mathcal{Y}$ at the right-end points of the K subintervals, i.e. we set $y_j := y(t_j)$, $1 \leqslant j \leqslant K$. In order to simulate noisy data we perturb the exact data with random Gaussian noise for different relative noise levels $\delta_{\mathrm{rel}} = 10^{-4} \cdots 10^{-2}$.

We consider the sample functions

$$x_1^\dagger(t) := \begin{cases} 5, & t \in [0.25, 0.27], \\ -3, & t \in [0.4, 0.45], \\ 4, & t \in [0.7, 0.73], \\ 0, & \text{else}. \end{cases} \quad \text{and} \quad x_2^\dagger(t) := \begin{cases} 3, & t \in [0.15, 0.3], \\ -5, & t \in [0.55, 0.75], \\ 0, & \text{else}. \end{cases}$$

In particular, $x_i^\dagger$, $i = 1, 2$, are chosen such that no discretization error occurs. For the discrepancy criterion, we set $\tau := 1.2$, and $x_0^* \equiv 0$ is taken as the starting point (hence, we get $x_0 = x_0^\delta = 0$ for both situations considered here). The number of iterations was limited by $n_{\max} = 10^6$.

For the approximate determination of $x_1^\dagger$, we apply the penalty P_β from (14) with different choices for the parameter β. The needed iteration numbers $N(\delta, y^\delta)$ as well as the relative error of the regularized solutions can be found in table 1. In particular, for $\beta = 10\,000$ the iteration number is much higher than in the other two cases. This fact is devoted to a phenomenon called *stagnation*: even if $x_{n+1}^* \neq x_n^*$ in each iteration, because of the structure of the shrinkage operator, it might happen that $x_{n+1}^\delta = x_n^\delta$. To avoid such effects a technique called *kicking* (see [21]) can be applied, which is not done here. In figure 1, we see the reconstruction of $x_1^\dagger$ on the interval $[0.22, 0.3]$ for $\delta_{\mathrm{rel}} = 10^{-2}$ and the different values for the parameter β. Here, the influence of the choice of β can be described as follows: the larger β, the sharper the zero part of the function $x_1^\dagger$ to be reconstructed, the price to be paid for it being the larger oscillations on the non-zero part. This is a well-known effect of the L^1-regularization.

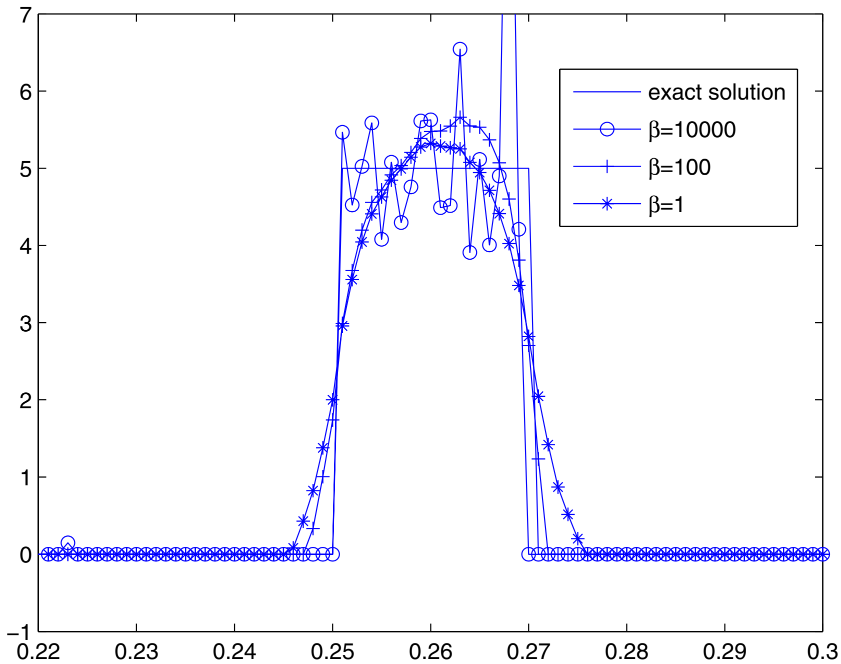

Figure 1. Exact versus regularized solution of $x_1^\dagger$ on the interval [0.22, 0.3].

Table 2. Reconstruction errors for the sample function $x_2^\dagger$.

	$\beta = 1$		$\beta = 100$		$\beta = 10\,000$	
δ_{rel}	$N(\delta, y^\delta)$	$\frac{\|x_N^\delta - x_2^\dagger\|}{\|x_2^\dagger\|}$	$N(\delta, y^\delta)$	$\frac{\|x_N^\delta - x_2^\dagger\|}{\|x_2^\dagger\|}$	$N(\delta, y^\delta)$	$\frac{\|x_N^\delta - x_2^\dagger\|}{\|x_2^\dagger\|}$
0.01	132	0.1668	213	0.1623	3613	0.1223
10^{-3}	1380	0.0725	1116	0.0714	10 164	0.0559
10^{-4}	19 327	0.0175	10 608	0.0161	25 246	0.0139

We now turn to the second sample function $x_2^\dagger$ and apply the penalty functional P_β from (15). Here, for solving the ROF model, the FPG algorithm [4] is applied. Additionally, in order to save numerical costs, we store the final primal and dual variables inside the FPG algorithm and use them as (good) initial guess in the next iteration step for solving the new ROF model. The numerical results for different noise levels δ_{rel} and different β are presented in table 2. Based on the specific structure of $x_2^\dagger$, one can note an increased quality of the reconstructed solutions with growing β, combined with higher costs for solving the ROF models in the first iteration steps (this is because x_n^* is multiplied by β and hence it becomes larger when β is increased). An illustration of this observation is given in figure 2. Here, the reconstruction of $x_2^\dagger$ on the intervals [0.25, 0.35] and [0.5, 0.6] for $\delta_{\mathrm{rel}} = 10^{-2}$ depending on β is shown. As we can see, the identification of the jumps is sharper the larger we choose β.

Summarizing these numerical results, we observe that our iterative regularization method for specific penalty terms points out the same properties of a solution of equation (1) as when we apply a Tikhonov regularization strategy with the same penalty functional. Hence, because

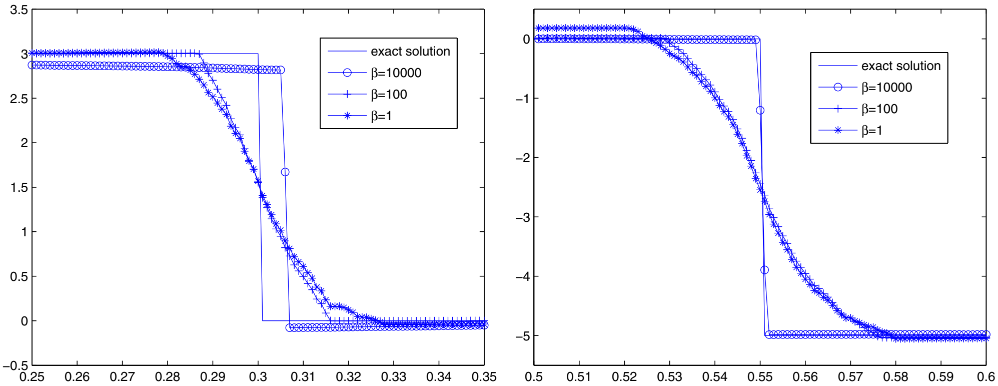

Figure 2. Exact versus regularized solution of $x_2^\dagger$ on [0.25, 0.35] (left) and on [0.5, 0.6] (right).

of the expected less numerical costs, the application of such iterative approaches is quite promising from the numerical point of view.

8. Summary

Motivated by the chances of reducing numerical costs, we presented an iterative regularization approach which can be considered as an alternative to Tikhonov regularization with s-convex penalty terms. Convergence and regularization properties were shown, as well as some applications in image and sparse reconstruction were provided. Since the presented algorithm is closely related to well-established methods for minimizing non-smooth Tikhonov functionals, we understand our presentation also as a motivation for considering the following question: Whenever an algorithm minimizes a (non-smooth) Tikhonov functional, does this approach (with possible small modifications) have the potential of being itself an iterative regularization scheme? The answer to this question seems to be of high interest for further numerical applications.

Acknowledgments

The authors are grateful to two anonymous reviewers for remarks and suggestions which improved the quality of the paper.

References

[1] Alber Y and Ryazantseva I 2006 *Nonlinear Ill-Posed Problems of Monotone Type* (Dordrecht: Springer)
[2] Ambrosio L, Fusco N and Pallara D 2000 *Functions of Bounded Variation and Free Discontinuity Problems* (New York: Oxford University Press)
[3] Bakushinsky A B and Kokurin M Yu 2004 *Iterative Methods for Approximate Solution of Inverse Problems (Mathematics and Its Applications)* (Dordrecht: Springer) p 577
[4] Beck A and Teboulle M 2009 A fast iterative shrinkage-thresholding algorithm for linear inverse problems *SIAM J. Imaging Sci.* **2** 183–202
[5] Boţ R I 2010 *Conjugate Duality in Convex Optimization* (Berlin: Springer)
[6] Boţ R I and Hofmann B 2010 An extension of the variational inequality approach for obtaining convergence rates in regularization of nonlinear ill-posed problems *J. Integral Eqns Appl.* **22** 369–92

[7] Bonesky T, Kazimierski K S, Maass P, Schöpfer F and Schuster T 2008 Minimization of Tikhonov functionals in Banach spaces *Abstr. Appl. Anal.* **2008** 192679
[8] Bredies K and Lorenz D A 2008 Iterated hard shrinkage for minimization problems with sparsity constraints *SIAM J. Sci. Comput.* **30** 657–83
[9] Burger M and Osher M 2004 Convergence rates of convex variational regularization *Inverse Probl.* **20** 1411–21
[10] Chambolle A 2004 An algorithm for total variation minimization and applications *J. Math. Imaging Vis.* **20** 89–97
[11] Daubechies I, Defrise M and De Mol C 2004 An iterative thresholding algorithm for linear inverse problems with a sparsity constraint *Commun. Pure Appl. Math.* **57** 1413–57
[12] Hanke M 1992 Regularization with differential operators: an iterative approach *Numer. Funct. Anal. Optim.* **13** 523–40
[13] Hanke M, Neubauer A and Scherzer O 1995 A convergence analysis of the Landweber iteration for nonlinear ill-posed problems *Numer. Math.* **72** 21–7
[14] Hein T and Hofmann B 2009 Approximate source conditions for nonlinear ill-posed problems—chances and limitations *Inverse Probl.* **25** 035003
[15] Hein T and Kazimierski K S 2010 Accelerated Landweber iteration in Banach spaces *Inverse Probl.* **26** 055002
[16] Hein T and Kazimierski K S 2010 Modified Landweber iteration in Banach spaces—convergence and convergence rates *Numer. Funct. Anal. Optim.* **31** 1158–84
[17] Hofmann B, Kaltenbacher B, Pöschl C and Scherzer O 2007 A convergence rates result in Banach spaces with non-smooth operators *Inverse Probl.* **23** 987–1010
[18] Kaltenbacher B, Neubauer A and Scherzer O 2008 *Iterative Regularization Methods for Nonlinear Ill-Posed Problems* (Berlin: de Gruyter)
[19] Kaltenbacher B, Schöpfer F and Schuster T 2009 Iterative methods for nonlinear ill-posed problems in Banach spaces: convergence and application to parameter identification problems *Inverse Probl.* **25** 065003
[20] Nemirivskii A S and Judin D B 1979 *Complexity of Problems and Effectiveness of Methods of Optimization* (Moscow: Nauka) (in Russian)
[21] Osher S, Mao Y, Dong B and Yin W 2010 Fast linearized Bregman iteration for compressive sensing and sparse denoising *Commun. Math. Sci.* **8** 93–111
[22] Phelps R R 1993 *Convex Functions, Monotone Operators and Differentiability* (Berlin: Springer)
[23] Pöschl C 2008 Tikhonov regularization with general residual term *PhD thesis* University of Innsbruck
[24] Rudin L, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica* D **60** 259–68
[25] Scherzer O, Grasmair M, Grossauer H, Haltmeier M and Lenzen F 2009 *Variational Methods in Imaging* (New York: Springer)
[26] Schöpfer F and Schuster T 2009 Fast regularizing sequential subspace optimization in Banach spaces *Inverse Probl.* **25** 015013
[27] Schöpfer F, Schuster T and Louis A K 2008 Metric and Bregman projections onto affine subspaces and their computation via sequential subspace optimization methods *J. Inverse Ill-Posed Probl.* **16** 479–506
[28] Yin W 2010 Analysis and generalizations of the linearized Bregman method *SIAM J. Imaging Sci.* **3** 856–77
[29] Yin W, Osher S, Goldfarb D and Darbon J 2008 Bregman iterative algorithms for l_1-minimization and applications to compressed sensing *SIAM J. Imaging Sci.* **1** 143–68
[30] Zălinescu C 2002 *Convex Analysis in General Vector Spaces* (New Jersey: World Scientific)