Noname manuscript No.
(will be inserted by the editor)

A conjugate gradient like method for p-norm minimization in functional spaces

Claudio Estatico · Serge Gratton · Flavia Lenti · David Titley-Peloquin

Received: date / Accepted: date

Abstract We develop an iterative algorithm to recover the minimum $p-$norm solution of the functional linear equation $Ax = b$, where $A : \mathcal{X} \longrightarrow \mathcal{Y}$ is a continuous linear operator between the two Banach spaces $\mathcal{X} = L^p$, $1 < p < 2$, and $\mathcal{Y} = L^r$, $r > 1$, with $x \in \mathcal{X}$ and $b \in \mathcal{Y}$. The algorithm is conceived within the same framework of the Landweber method for functional linear equations in Banach spaces proposed by Schöpfer, Louis and Schuster (Inverse Probl., 22:311–329, 2006). Indeed, the algorithm is based on using, at the $n-$th iteration, a linear combination of the steepest current "descent functional" $A^* J\,(b - Ax_n)$ and the previous descent functional, where J denotes a duality map of the Banach space $\mathcal{Y}$. In this regard, the algorithm can be viewed as a generalization of the classical conjugate gradient method on the normal equations (CGNR) in Hilbert spaces.
We demonstrate that the proposed iterative algorithm converges strongly to the minimum $p-$norm solution of the functional linear equation $Ax = b$ and that it is also a regularization method, by applying the discrepancy principle as stopping rule.
According to the geometrical properties of L^p spaces, numerical experiments show that the method is fast, robust in terms of both restoration accuracy and stability, promotes sparsity and reduces the over-smoothness in reconstructing edges and abrupt intensity changes.

Keywords Iterative regularization · Banach spaces · Conjugate gradient method

C. Estatico
Dipartimento di Matematica, Università degli Studi di Genova, Italy
E-mail: estatico@dima.unige.it

S. Gratton
Université de Toulouse, INP-IRIT, France
E-mail: serge.gratton@enseeiht.fr

F. Lenti
Institut de Recherche en Informatique de Toulouse, France
E-mail: flavialenti@gmail.com

D. Titley-Peloquin
Department of Bioresource Engineering, McGill University, Canada
E-mail: david.titley-peloquin@mcgill.ca

Mathematics Subject Classification (2010) 65J20 · 47A52 · 65F10

1 Introduction

We discuss the problem of iteratively recovering a solution of the functional equation

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

where $A : \mathcal{X} \longrightarrow \mathcal{Y}$ is a linear and continuous operator between two functional spaces $\mathcal{X}$ and $\mathcal{Y}$, with $x \in \mathcal{X}$ and $b \in \mathcal{Y}$.
The problem we consider is ill-posed, since its solution is not unique or does not exist for arbitrary data or does not depend continuously on the data [2]. There is now an exhaustive and comprehensive literature about linear and nonlinear inverse ill-posed problems in Hilbert spaces. Over the last decade, there has been a growing interest in studying inverse ill-posed problems in Banach spaces, because many applicative problems are therein better modeled ([3], [8], [19], [20], [22], [21], [23], [25], to name just a few references).

In [20], Schöpfer, Louis and Schuster computed a regularized solution in Banach spaces by means of the following generalized Landweber, i.e. gradient-type, iterative scheme

$$x_{n+1} = j^{\mathcal{X}^*}\left(j^{\mathcal{X}}(x_n) - \alpha_n A^* j^{\mathcal{Y}}(Ax_n - b)\right)\,, \tag{2}$$

where $j^{\mathcal{X}}$, $j^{\mathcal{X}^*}$ and $j^{\mathcal{Y}}$ are duality mappings of the corresponding spaces $\mathcal{X}, \mathcal{X}^*, \mathcal{Y}$ [25], and $\alpha_n > 0$ is a appropriately chosen step size. The discrepancy principle is employed to obtain a suitable stopping index in the case of noisy data. Like in Hilbert spaces, the method turned out to have good regularizing properties but its convergence speed is generally very slow.

The conjugate gradient (CG) algorithm for linear systems is known to enjoy better convergence properties than methods based solely on gradient descent, but its original formulation by Hestenes and Stiefel applies only in Hilbert spaces. In this respect, the algorithm we develop in this paper can be considered as an extension to Banach space settings of the conventional CG method on the normal equations for linear systems in Hilbert spaces (also known as CGNR method). Indeed, on the basis of the theoretical framework of [20], we suggest using a different "descent functional" than the steepest one $A^* j^{\mathcal{Y}}(b - Ax_n)$ at each iterative step, which is defined by the same kind of linear combination as in the CGNR method. More specifically, in [20] the spaces $\mathcal{X}$ is assumed to be smooth and uniformly convex and $\mathcal{Y}$ can even be an arbitrary Banach space. Although our convergence proof closely follows the arguments of [20], in this paper we restrict the analysis to the case of $\mathcal{X} = L^p$ with $1 < p < 2$ and $\mathcal{Y} = L^r$ with $r > 1$. On the other hand, we argue that a generalization to other Banach spaces could be obtained without strong changes.

To give a numerical validation, we apply our algorithm to a standard image restoration problem. The exact solution presents discontinuities characterized by different dimensions and different intensities. Reconstructions obtained by using the proposed approach are compared with the ones obtained by CG in Hilbert

spaces, and by both steepest descent and Landweber methods in Hilbert and Banach spaces.
Since the proposed method is based on the minimization of p-norm (with $1 < p < 2$), it is expected to overcome the typical over-smoothing drawback effects of regularization in Hilbert spaces and to enhance the sparsity of the reconstructed solution. Our first numerical evidences confirm such a positive behavior in both convergence speed and accuracy of the restoration.

Due to the intrinsic non-linearity of the duality maps, we mention that, differing from original CG method in Hilbert spaces, the short term recursion formula for computing the descent functionals (which have the same role of the descent directions in Hilbert spaces) does not guarantee now their full mutual conjugacy. Thus, the proposed algorithm does not converge within n iterations for $n \times n$ linear systems in Banach spaces. A concomitant proposal by Herzog and Wollner [13] for the linear system (1) in a reflexive Banach space with $A : \mathcal{X} \longrightarrow \mathcal{X}^*$ self-adjoint and positive, preserves the conjugacy of the descent directions, so that convergence in a finite number of steps holds as in the finite dimensional Hilbertian case. Another positive fact is that, differing form the proposed algorithm, the method in [13] does not require to solve any one-dimensional minimization problem for the computation of the optimal step size at each iteration, since a closed formula holds as in the Hilbertian case. On the other hand, to keep the mutual conjugacy of any new descent direction, such an algorithm requires the storage of all the previous descent directions, which are all used for the implementation of a modified Arnoldi conjugation procedure as the iterations go on. Although finite convergence is not assured, the CG-like method proposed in the present paper does not require any other storage than the last descent (functional) direction and can be used in a more general setting, since the operator A is here not required to be self-adjoint. Moreover, differing form [13], the iterations are here explicitly computed in the dual space $\mathcal{X}^*$, as generally done in [20], [22] and [3].

The remainder of this paper is organized as follows. In Section 2, we give the necessary theoretical tools used in Section 3 and Section 4 to prove the strong convergence for noise-free and noisy data b, respectively. In Section 5, a practical implementation issue related to the step size computation is described and in Section 6 the numerical experiments are presented and discussed. Some conclusions are drawn in Section 7.

2 Preliminaries

The aim of this section is to briefly recall basic tools and classical notations usually used in the Banach space setting for the regularization of ill-posed problems. For details and proofs, we suggest the following monographs: [7], [17], [25].
Following the same notation as [20], throughout the paper, both $\mathcal{X}$ and $\mathcal{Y}$ are two real Banach spaces with dual spaces $\mathcal{X}^*$ and $\mathcal{Y}^*$ equipped with the corresponding operator norm . For $x^* \in \mathcal{X}^*$ and $x \in \mathcal{X}$ we denote by $\langle x^*, x\rangle_{\mathcal{X}^* \times \mathcal{X}}$ and $\langle x, x^*\rangle_{\mathcal{X} \times \mathcal{X}^*}$ the duality pairing defined as $\langle x^*, x\rangle_{\mathcal{X}^* \times \mathcal{X}} = \langle x, x^*\rangle_{\mathcal{X} \times \mathcal{X}^*} = x^*(x)$. In general, we omit subscripts indicating the space when this piece of information is implicitly clear. In particular, this will be done for any norm of vectors or oper-

ators throughout the paper.
The operator $A : \mathcal{X} \longrightarrow \mathcal{Y}$ is continuous and linear, and $A^* : \mathcal{Y}^* \longrightarrow \mathcal{X}^*$ denotes its adjoint operator of A, that is, the operator such that $\langle Ax, y\rangle = \langle x, A^*y\rangle,\ \forall x \in \mathcal{X}$ and $y \in \mathcal{Y}^*$. We have $\|A\| = \|A^*\|$ in each corresponding operator norm.
For two real numbers a and b, we write $a \vee b = \max\{a, b\}\,,\quad a \wedge b = \min\{a, b\}\,.$
Throughout the paper, for $p,\ r \in (1, +\infty)$, we usually denote by $q = p^*$ and $s = r^*$ their Hölder conjugates.

2.1 Duality mapping

The key point in the generalization of the regularization method in Banach spaces is the duality mapping [1], [14], [18]. A duality map is an appropriate function which associates an element of a Banach space $\mathcal{X}$ with an element of its dual $\mathcal{X}^*$, and it is useful when the Banach space $\mathcal{X}$ is not isometrically isomorphous to its dual $\mathcal{X}^*$. Formally, we have the following definition.

Definition 1 ***(Duality mapping)***
The (set-valued) mapping $J_p^{\mathcal{X}} : \mathcal{X} \longrightarrow 2^{\mathcal{X}^*}$ with $p \geq 1$ defined by

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

is called duality map of $\mathcal{X}$ with gauge function $t \longmapsto t^{p-1}$.
In general, by $j_p^{\mathcal{X}}(x) \in \mathcal{X}$ we will denote a single-valued selection of the subset $J_p^{\mathcal{X}}(x) \subset \mathcal{X}$. If $\mathcal{X} = l^r$ or $\mathcal{X} = L^r$, with $1 < r < +\infty$, the duality map is a single-valued function which will be denoted as $j_p^r(\cdot)$. If $r = p$, the apex will be usually omitted.

Example 1 Let us consider $x \in l^r(\mathbb{R}^n)$. For every $p \in (1, +\infty)$, the duality map j_p^r is given by

$$j_p^r(x) = \|x\|^{p-r}|x|^{r-1} sign(x)\,, \tag{3}$$

where $sign(\cdot)$ denotes the sign function and the product has to be considered as component-wise. In particular, for the Hilbert space l^2, according to the Riesz representation Theorem, j_2 is the identity operator, i.e. $j_2(x) = x$.

The meaning of any duality mapping is naturally related to the sub-gradient of the Banach norm. We first recall the following basic definition.

Definition 2 ***(Subgradient of convex functional)***
Let $f : \mathcal{X} \longrightarrow \mathbb{R} \cup \{+\infty\}$. Then, $x^* \in \mathcal{X}^*$ is a subgradient of f at $x \in \mathcal{X}$ if

$$f(y) \geq f(x) + \langle x^*, y - x\rangle,\ \forall y \in \mathcal{X}.$$

The set $\partial f(x) \subset \mathcal{X}^*$ of all subgradients of f at x is called subdifferential of f at x.

The following important theorem give us a simple and heuristic way to understand the meaning of any duality map.

Theorem 1 ***(Asplund Theorem)*** *[1] Let $\mathcal{X}$ be a Banach space and $p > 1$. Then*

$$J_p^{\mathcal{X}} = \partial\left(\frac{1}{p}\|\cdot\|^p\right). \tag{4}$$

The Asplund Theorem is a key step for an intuitive "geometrical" interpretation of any Banach space and its duality maps. In our context, it will be useful in the following sections for the computation of the subdifferential of the residual functional $\frac{1}{p}\|Ax - b\|^p$ by the chaining rule.

2.2 Geometry of Banach spaces and Bregman distance

Regarding the regularization theory in Banach spaces, the geometrical properties of the spaces, like convexity and smoothness, play a crucial role [7]. These properties, which can be viewed as an extension to Banach space of heuristic basic properties of Hilbert (or simply Euclidean) spaces, are strictly related to the duality maps and will be useful to prove the convergence of the proposed algorithm. The "degree" of convexity or smoothness of a Banach space is given by its modulus of convexity and its modulus of smoothness, and we refer to Section 2.1 of [20] for their formal definitions and for the subsequent characterization of p-convex, p-smooth, uniformly convex and uniformly smooth Banach spaces.

In [5], it has been argued that the Bregman distance is the correct measure for measuring the "quality" of the regularized solution in the case of convex regularization in Banach spaces. Moreover, due to the geometrical properties of Banach spaces, it is often more appropriate to exploit the Bregman distance between two vectors x and y instead of more conventional norm-distances like $\|x - y\|^p$ or $\|j_p^{\mathcal{X}}(x) - j_p^{\mathcal{X}}(y)\|^p$ to prove convergence of the algorithms [6], [10], [20].

Generally, the Bregman distance associated to a convex functional is defined as the difference between the functional and its linear approximation around x as follows [4].

Definition 3 ***(Bregman distance)***
Let $f : \mathcal{X} \longrightarrow \mathbb{R}$ be a convex and contiously-differentiable functional on a Banach space $\mathcal{X}$. Then the Bregman distance $\Delta(x, y)$ of f between $x \in \mathcal{X}$ and $y \in \mathcal{X}$ is defined as

$$\Delta(x, y) = f(x) - f(y) - f'(y)(x - y),$$

where $f'(y)$ is the first order Taylor expansion of f at y.

The role of a Bregman distance is similar to the role of any metric. However, any Bregman distance in general does not satisfy the triangle inequality nor symmetry. As shown by the following special case first example, the Bregman distance is a generalization of the square norm distance in Hilbertian contexts, when the basic square norm distance functional is considered.

Example 2 If $\mathcal{X}$ is an Hilbert space, then $\Delta_2(x, y) = \frac{1}{2}\|x-y\|^2$, for $f_2(x) = \frac{1}{2}\|x\|^2$.

Example 3 In a general Banach space $\mathcal{X}$, by considering the convex functional $f_p(x) = \frac{1}{p}\|x\|^p$, thanks to the Asplund Theorem (4) the associated Bregman distance Δ_p is

$$\Delta_p(x, y) = \frac{1}{p}\|x\|^p - \frac{1}{p}\|y\|^p - \langle j_p(y), x - y\rangle, \tag{5}$$

for any $x, y \in \mathcal{X}$, where j_p is a single-valued selection of $J_p^{\mathcal{X}}$.

In general $\Delta_p(x, y) \geq 0$, and $\Delta_p(x, y) = 0$ if and only if $j_p(y) \in J_p^{\mathcal{X}}(x)$ [25]. From that, for all $x,\ z\ \in \mathcal{X}$ we have $\frac{1}{p}\|z\|^p - \frac{1}{p}\|x\|^p - \langle j_p(x), z - x\rangle \geq 0$, which, by setting $y = -\,(z - x)$, yields

$$\frac{1}{p}\|x - y\|^p - \frac{1}{p}\|x\|^p + \langle j_p(x), y\rangle \geq 0. \tag{6}$$

In general, the Bregman distance in the primal space $\mathcal{X}$ and the Bregman distance in its dual space $\mathcal{X}^*$ are strictly linked, since

$$\Delta_p(x,y) = \Delta_q\big(j_p(x), j_p(y)\big) .$$

This duality, together with basic relationships about strong convergence in norm and convergence in Bregman distance (see [25], and [20] Theorem 2.12) and with the Xu-Roach inequality (see [26], and [20] Theorem 2.8), will have an important role in the convergence analysis of our algorithm.

3 CG and its convergence: the noise-free case

We restrict to the case $\mathcal{X} = L^p$ with $1 < p < 2$ and $\mathcal{Y} = L^r$, with $r > 1$. By implicitly considering the classical isomorphisms, in the following L^q and L^s denote the dual spaces of L^p and L^r, respectively.
Concerning our functional equation (1), we are interested in finding the minimum p-norm solution of (1), denoted hereinafter as $\bar{x}$, with exact data $b \in \mathcal{R}(A)$. In this respect, we first recall the following existence and characterization result.

Theorem 2 *[20] If $b \in \mathcal{R}(A)$, then there exists the minimum p-norm solution $\bar{x} \in L^p$ of (1) and $j_p(\bar{x}) \in \overline{\mathcal{R}(A^*)}$.*
Moreover, if $\tilde{x} \in L^p$ fulfils $j_p(\tilde{x}) \in \overline{\mathcal{R}(A^*)}$ and $\tilde{x} - \bar{x} \in \mathcal{N}(A)$, then $\tilde{x} = \bar{x}$.

To recover the minimum p-norm solution $\bar{x}$, we propose the following algorithm based on the well known conjugate gradient method in Hilbert spaces.

Algorithm 1 ***Conjugate Gradient based method in Banach spaces***

Choose a constant $C \in (0,1)$, and let

$$\gamma = \frac{Cr}{2^r - 1 + Cr} \tag{7}$$

be a relaxation parameter and d an arbitrary constant satisfying

$$0 < d \leq \left(1 - \frac{2^r - 1 + r}{r}\gamma\right)\frac{1}{\|A\|}.$$

Set $n = 0$, $x_0^* = 0$, $p_0^* = A^* j_r(b)$, $R_0 = \|b\|$, and

$$\alpha_0 = \arg \min_{\alpha \in \left[0, \frac{q^{p-1}}{\|A\|^p} R_0^{p-r}\right]} \|A j_q(x_0^* + \alpha p_0^*) - b\|^r. \tag{8}$$

While $R_n > 0$ **do**:
Update $n \leftarrow n + 1$.
Compute

$$\begin{cases} x_n^* = x_{n-1}^* + \alpha_{n-1} p_{n-1}^*, \\ x_n = j_q\left(x_n^*\right), \\ p_n^* = -A^* j_r\left(Ax_n - b\right) + \beta_n p_{n-1}^*, \end{cases} \tag{9}$$

where

$$\alpha_n = \arg \min_{\alpha \in [0, T_n]} \|A j_q(x_n^* + \alpha p_n^*) - b\|^r \tag{10}$$

and

$$\beta_n = \gamma \frac{R_n^r}{R_{n-1}^r}, \tag{11}$$

with

$$R_n = \|Ax_n - b\|,$$

$$T_n = \min\left\{ \frac{R_n^{2-r}\left(V_n - d\|A\|Q_n\right)}{G_q 2^{q-2}\|x_n^*\|^{q-2}(q-1)\|A\|^2 Q_n^2},\ \frac{\|x_n^*\|}{\|A\| R_n^{r-1} Q_n} \right\},$$

$$V_n = 1 - \frac{2^r - 1}{r}\gamma Q_{n-1},$$

$$Q_n = \frac{1-\gamma^{n+1}}{1-\gamma},$$

and G_q the constant value of the Banach space defined in [20].

End while

Before providing the convergence proof, which follows the idea developed in [20], some comments are useful. First, it is interesting to notice that the functional p_n^* is a linear combination of the current steepest descent functional $-A^* j_r\left(Ax_n - b\right)$ and the previous descent functional p_{n-1}^*. This is a key point in the definition of the CG method in Hilbert spaces, where the descent functionals are just descent directions.

On the other hand, there are two important facts to mention as main differences between (the proposed) CG in Banach and (the conventional) CG in Hilbert spaces. The first is that in Banach spaces the optimal step size α_n cannot be directly computed by means of an explicit formula, since the corresponding one-dimensional minimization problem (10) is no more quadratic. The second is related to the weight β_n of (11). Now β_n is different from the Fletcher Reeves formula for CG in Hilbert spaces. Indeed β_n now concerns the ratio of the norms of the two last residuals (instead of the norms of the last two gradients) and it requires the relaxation factor $\gamma < 1/2$ defined in (7), which does not appear in the original CG method. Hence, recalling the optimality condition on the step size (10), differing from the original CG method, β_n is always lower than 1/2. Although such a constraint is necessary only for our theoretical convergence proof, our numerical tests show that it helps the method to be more stable.

We prove the following main convergence result.

Theorem 3 *Let $\mathcal{X} = L^p$ with $1 < p < 2$ and $\mathcal{Y} = L^r$, with $r > 1$. The sequence of the iterations $\left(x_n\right)_n$ of Algorithm 1 defined in (9), either stops at or converges strongly to the minimum $p-$norm solution $\bar{x}$ of (1).*

Proof First of all, by simple induction on the definition of x_n^* and p_n^*, we have that any $x_n^* = j_p(x_n) \in \overline{\mathcal{R}(A^*)}$, since $x_0^* = 0 \in \overline{\mathcal{R}(A^*)}$ and $p_0^* = A^* j_r(b) \in \overline{\mathcal{R}(A^*)}$.
Then, if the algorithm at step n gives $R_n = 0$, by Theorem 2 we have that $x_n = \bar{x}$. Indeed, by $0 = R_n = \|Ax_n - b\| = \|A(x_n - \bar{x})\|$, we know that $x_n - \bar{x} \in \mathcal{N}(A)$, so that $x_n = \bar{x}$.
Otherwise we have that $R_n > 0$ for all $n > 0$. In this case, as similarly done in [20] for the Landweber method in Banach space, the proof of convergence will be structured in the following four steps:

1. the sequence of the Bregman distances $(\Delta_p(\bar{x}, x_n))_n$ obeys a recursive inequality which implies its convergence;
2. the sequence of the residuals $(R_n)_n$ is such that $\lim_{n\longrightarrow+\infty} R_n = 0$;
3. the sequence of the iterates $(x_n)_n$ has a Cauchy subsequence;
4. the sequence of the iterates $(x_n)_n$ converges strongly to $\bar{x}$.

Before going on, we itemize some notations used in the proof.

$$g_n^* = -A^* j_r(Ax_n - b),$$

$$\Delta_n = \Delta_p(\bar{x}, x_n),$$

$$c_{k,n} = \prod_{l=k+1}^{n} \beta_l = \gamma^{n-k} \frac{R_n^r}{R_k^r},$$

$$\tau_n = \frac{\alpha_n \|A\| R_n^{r-1} Q_n}{\|x_n^*\|},$$

$$W = G_q 2^{q-2}(q-1),$$

$$S_n = \frac{W\tau_n \|x_n^*\|^{q-1}}{R_n} \frac{Q_n}{V_n} \|A\|.$$

1. First step: the sequence of Bregman distances $(\Delta_n)_n$ is non-decreasing and convergent.
By (5), we can write

$$\begin{aligned}
\Delta_{n+1} &= \frac{1}{p}\|\bar{x}\|^p - \frac{1}{p}\|x_{n+1}\|^p + \langle x_{n+1}^*, \bar{x} - x_{n+1}\rangle = \frac{1}{q}\|x_{n+1}\|^p + \frac{1}{p}\|\bar{x}\|^p - \langle x_{n+1}^*, \bar{x}\rangle \\
&= \frac{1}{q}\|x_{n+1}^*\|^q + \frac{1}{p}\|\bar{x}\|^p - \langle x_{n+1}^*, \bar{x}\rangle \\
&= \frac{1}{q}\|x_n^* + \alpha_n p_n^*\|^q + \frac{1}{p}\|\bar{x}\|^p - \langle x_n^*, \bar{x}\rangle - \alpha_n\langle p_n^*, \bar{x}\rangle.
\end{aligned}$$

For $n = 0$, since $x_0 = 0$, then $R_0 = \|b\| > 0$, $p_0^* = A^* j_r(b)$ and $\Delta_0 = \frac{1}{p}\|\bar{x}\|^p$. Hence

$$\begin{aligned}
\Delta_1 &= \frac{1}{q}\|\alpha_0 p_0^*\|^q + \frac{1}{p}\|\bar{x}\|^p - \alpha_0\langle p_0^*, \bar{x}\rangle = \frac{1}{q}\alpha_0^q \|A^* j_r(b)\|^q + \Delta_0 - \alpha_0\langle j_r(b), A\bar{x}\rangle \\
&= \frac{1}{q}\alpha_0^q \|A^* j_r(b)\|^q + \Delta_0 - \alpha_0\langle j_r(b), b\rangle \le \frac{1}{q}\alpha_0^q \|A\|^q R_0^{(r-1)q} + \Delta_0 - \alpha_0 R_0^r
\end{aligned}$$

To obtain a reduction of the Bregman distance at first iteration, we consider the upper-bound for $\alpha_0 > 0$ defined in (8), since

$$\frac{1}{q}\|A\|^q \alpha_0^q R_0^{(r-1)q} - \alpha_0 R_0^r < 0 \Longleftrightarrow \alpha_0 < \frac{q^{p-1}}{\|A\|^p} R_0^{p-r} \Longrightarrow \Delta_1 < \Delta_0. \qquad (12)$$

Now we consider $n > 0$. Thanks to the Xu-Roach inequality ([20], Theorem 2.8),

$$\begin{aligned}\Delta_{n+1} &= \frac{1}{q}\|x_n^* + \alpha_n p_n^*\|^q + \frac{1}{p}\|\bar{x}\|^p - \langle x_n^*, \bar{x}\rangle - \alpha_n\langle p_n^*, \bar{x}\rangle \\ &\leq \frac{1}{q}\Big[\|x_n^*\|^q + qG_q\int_0^1 \frac{(\|x_n^* + t\alpha_n p_n^*\| \vee \|x_n^*\|)^q}{t}\varrho_q\left(\frac{t\|\alpha_n p_n^*\|}{\|x_n^* + t\alpha_n p_n^*\| \vee \|x_n^*\|}\right)dt \\ &\quad -q\langle x_n, -\alpha_n p_n^*\rangle\Big] + \frac{1}{p}\|\bar{x}\|^p - \langle x_n^*, \bar{x}\rangle - \alpha_n\langle p_n^*, \bar{x}\rangle,\end{aligned}$$

that is

$$\begin{aligned}\Delta_{n+1} \leq \Delta_n &+ G_q\int_0^1 \frac{(\|x_n^* + t\alpha_n p_n^*\| \vee \|x_n^*\|)^q}{t}\varrho_q\left(\frac{t\|\alpha_n p_n^*\|}{\|x_n^* + t\alpha_n p_n^*\| \vee \|x_n^*\|}\right)dt \\ &+ \alpha_n\langle p_n^*, x_n - \bar{x}\rangle.\end{aligned} \tag{13}$$

We now study the last two addenda of the inequality (13).

A) Let us consider the term

$$\langle p_n^*, x_n - \bar{x}\rangle\,. \tag{14}$$

From

$$\begin{aligned}p_n^* &= g_n^* + \beta_n p_{n-1}^* = g_n^* + \sum_{k=0}^{n-1}\left[\prod_{l=k+1}^{n}\beta_l\right]g_k^* = g_n^* + \sum_{k=0}^{n-1}c_{k,n}g_k^* \\ &= -A^*\left[j_r(Ax_n - b) + \sum_{k=0}^{n-1}c_{k,n}j_r(Ax_k - b)\right],\end{aligned}$$

thanks to (6) we have

$$\begin{aligned}\langle p_n^*, x_n - \bar{x}\rangle &= -\langle j_r(Ax_n - b) + \sum_{k=0}^{n-1}c_{k,n}j_r(Ax_k - b), Ax_n - b\rangle \\ &= -R_n^r - \sum_{k=0}^{n-1}c_{k,n}\langle j_r(Ax_k - b), Ax_n - b\rangle \\ &\leq -R_n^r - \sum_{k=0}^{n-1}c_{k,n}\left[\frac{1}{r}\|Ax_k - b\|^r - \frac{1}{r}\|Ax_k - b - Ax_n + b\|^r\right] \\ &\leq -R_n^r - \sum_{k=0}^{n-1}c_{k,n}\left[\frac{1}{r}R_k^r - \frac{1}{r}\Big(\|Ax_k - b\| + \|Ax_n - b\|\Big)^r\right] \\ &= -R_n^r + \sum_{k=0}^{n-1}c_{k,n}\left[\frac{1}{r}\left(R_k + R_n\right)^r - \frac{1}{r}R_k^r\right] \\ &\leq -R_n^r + \frac{2^r - 1}{r}\sum_{k=0}^{n-1}c_{k,n}R_k^r\,,\end{aligned} \tag{15}$$

where the last inequality holds by construction of the optimal step size (10) which guarantees that the residuals R_k decrease.

B) Let us consider the term

$$G_q \int_0^1 \frac{(\|x_n^* + t\alpha_n p_n^*\| \vee \|x_n^*\|)^q}{t} \varrho_q \left(\frac{t\|\alpha_n p_n^*\|}{\|x_n^* + t\alpha_n p_n^*\| \vee \|x_n^*\|} \right) dt \,. \tag{16}$$

Since $0 \le t \le 1$ and $\|x_n^*\| \le \|x_n^* + t\alpha_n p_n^*\| \vee \|x_n^*\| \le \|x_n^*\| + \alpha_n\|p_n^*\|$, we have

$$p_n^* = -A^* \left[j_r(Ax_n - b) + \sum_{k=0}^{n-1} c_{k,n} j_r(Ax_k - b) \right],$$

so that

$$\begin{aligned}
\|p_n^*\| &= \left\| -A^* \left[j_r(Ax_n - b) + \sum_{k=0}^{n-1} c_{k,n} j_r(Ax_k - b) \right] \right\| \\
&\le \|A\| \left\| j_r(Ax_n - b) + \sum_{k=0}^{n-1} c_{k,n} j_r(Ax_k - b) \right\| \\
&\le \|A\| \left[\|j_r(Ax_n - b)\| + \sum_{k=0}^{n-1} c_{k,n} \|j_r(Ax_k - b)\| \right] \\
&= \|A\| \left[R_n^{r-1} + \sum_{k=0}^{n-1} \gamma^{n-k} \frac{R_n^r}{R_k^r} R_k^{r-1} \right] \le \|A\| R_n^{r-1} \left[1 + \sum_{k=0}^{n-1} \gamma^{n-k} \frac{R_n}{R_k} \right] \\
&\le \|A\| R_n^{r-1} \left[1 + \sum_{k=0}^{n-1} \gamma^{n-k} \right] = \|A\| R_n^{r-1} \left[\frac{1-\gamma^{n+1}}{1-\gamma} \right] = \|A\| R_n^{r-1} Q_n \,,
\end{aligned} \tag{17}$$

that is, $\|p_n^*\| \le \|A\| R_n^{r-1} Q_n$. This last inequality, together with condition on T_n of (10), gives

$$\alpha_n \|p_n^*\| \le \alpha_n \|A\| R_n^{r-1} Q_n \le \|x_n^*\| \,. \tag{18}$$

By means of the latter, we have that

$$\begin{aligned}
&\int_0^1 \frac{(\|x_n^* + t\alpha_n p_n^*\| \vee \|x_n^*\|)^q}{t} \varrho_q \left(\frac{t\|\alpha_n p_n^*\|}{\|x_n^* + t\alpha_n p_n^*\| \vee \|x_n^*\|} \right) dt \\
&\quad \le \int_0^1 \frac{(\|x_n^*\| + \alpha_n\|p_n^*\|)^q}{t} \varrho_q \left(\frac{t\|\alpha_n p_n^*\|}{\|x_n^*\|} \right) dt \\
&\quad \le \int_0^1 \frac{(\|x_n^*\| + \alpha_n\|p_n^*\|)^q}{t} \varrho_q \left(\frac{t\alpha_n \|A\| R_n^{r-1} Q_n}{\|x_n^*\|} \right) dt \\
&\quad \le \left(\|x_n^*\| + \alpha_n \|p_n^*\| \right)^q \int_0^1 \frac{1}{t} \varrho_q (t\tau_n) \, dt = \left(\|x_n^*\| + \alpha_n \|p_n^*\| \right)^q \int_0^{\tau_n} \frac{1}{t_1} \varrho_q (t_1) \, dt_1 \\
&\quad \le \left(\|x_n^*\| + \alpha_n \|p_n^*\| \right)^q \int_0^1 \frac{1}{t} \varrho_q (t\tau_n) \, dt = \left(\|x_n^*\| + \alpha_n \|p_n^*\| \right)^q \int_0^{\tau_n} \frac{1}{t_1} \varrho_q (t_1) \, dt_1 \\
&\quad \le \left(\|x_n^*\| + \alpha_n \|p_n^*\| \right)^q \int_0^{\tau_n} \frac{1}{t_1} \frac{q-1}{2} t_1^2 dt_1 = \left(\|x_n^*\| + \alpha_n \|p_n^*\| \right)^q \frac{q-1}{2^2} \tau_n^2 \\
&\quad \le (2\|x_n^*\|)^q \frac{(q-1)}{2^2} \tau_n^2 = 2^{q-2} \|x_n^*\|^{q-2} (q-1) \alpha_n^2 \|A\|^2 R_n^{2(r-1)} Q_n^2.
\end{aligned} \tag{19}$$

Now, reconsidering (13), we can write

$$\begin{aligned}\Delta_{n+1} &\le \Delta_n + G_q \int_0^1 \frac{(\|x_n^* + t\alpha_n p_n^*\| \vee \|x_n^*\|)^q}{t} \varrho_q \left(\frac{t\|\alpha_n p_n^*\|}{\|x_n^* + t\alpha_n p_n^*\| \vee \|x_n^*\|} \right) dt \\ &\quad + \alpha_n \langle p_n^*, x_n - \bar{x} \rangle \\ &\le \Delta_n + G_q \left[2^{q-2} \|x_n^*\|^{q-2} (q-1) \alpha_n^2 \|A\|^2 R_n^{2(r-1)} Q_n^2 \right] \\ &\quad + \alpha_n \left[-R_n^r + \frac{2^r - 1}{r} \sum_{k=0}^{n-1} c_{k,n} R_k^r \right].\end{aligned}$$

We have that $(\Delta_n)_n$ is a non-increasing sequence, that is, $\Delta_{n+1} \le \Delta_n$, if

$$\alpha_n \left(\alpha_n \left[G_q 2^{q-2} \|x_n^*\|^{q-2} (q-1) \|A\|^2 R_n^{2(r-1)} Q_n^2 \right] - R_n^r + \frac{2^r - 1}{r} \sum_{k=0}^{n-1} c_{k,n} R_k^r \right) \le 0. \tag{20}$$

The latter holds if

$$R_n^r - \frac{2^r - 1}{r} \sum_{k=0}^{n-1} c_{k,n} R_k^r \ge 0 \tag{21}$$

and

$$0 \le \alpha_n \le \min \left\{ \frac{R_n^r - \frac{2^r-1}{r} \sum_{k=0}^{n-1} c_{k,n} R_k^r}{G_q 2^{q-2} \|x_n^*\|^{q-2} (q-1) \|A\|^2 R_n^{2(r-1)} Q_n^2}, \frac{\|x_n^*\|}{\|A\| R_n^{r-1} Q_n} \right\}, \tag{22}$$

$\forall n = 1, 2, 3, \ldots$, where condition (18) has been considered.
First we show that the relaxation parameter $\gamma < \frac{1}{2}$ of (7) allows (21) to be always true. Indeed we have

$$\sum_{k=0}^{n-1} c_{k,n} R_k^r = \sum_{k=0}^{n-1} \gamma^{n-k} \frac{R_n^r}{R_k^r} R_k^r = \sum_{k=0}^{n-1} \gamma^{n-k} R_n^r = \frac{\gamma - \gamma^{n+1}}{1-\gamma} R_n^r \le \frac{\gamma}{1-\gamma} R_n^r ,$$

so that

$$R_n^r - \frac{2^r - 1}{r} \sum_{k=0}^{n-1} c_{k,n} R_k^r \ge \left(1 - \frac{2^r - 1}{r} \frac{\gamma}{1-\gamma} \right) R_n^r = (1 - C) R_n^r > 0 . \tag{23}$$

By means of these conditions on γ and α_n , we can finally write

$$0 \le \Delta_{n+1} \le \Delta_n + W \|x_n^*\|^q \tau_n^2 - \frac{\|x_n^*\| \tau_n}{\|A\| Q_n} V_n R_n \le \Delta_n , \tag{24}$$

where τ_n, W and V_n have been previously properly defined. According to the aim of the first step of this proof, this last inequality shows that the sequence $(\Delta_n)_n$ is both non-increasing and bounded by below, so that it is convergent.

Remark 1 From (15) and (23), we obtain that $-\langle p_n^*, x_n - \bar{x} \rangle > 0$, which shows that p_n^* is a descent functional at the current point x_n, as expected by any gradient-type iteration scheme.

2. Second step: the sequence of residuals $(R_n)_n$ vanishes, that is, $\lim_{n\longrightarrow+\infty} R_n = 0$.
On the ground of [20] Theorem 2.12 (b), since by first step we know that $(\Delta_n)_n$ is bounded, then the sequence $(x_n)_n$ is bounded. In addition, since $\|x_n^*\| = \|j_p(x_n)\| = \|x_n\|^{p-1}$ and $R_n = \|Ax_n - b\|$, then both the sequences $(x_n^*)_n$ and $(R_n)_n$ are bounded too.

We rewrite (24) as follows

$$\Delta_{n+1} \le \Delta_n - \frac{1-S_n}{\|A\|}\frac{V_n}{Q_n}\tau_n\|x_n^*\|R_n \le \Delta_n$$

that is

$$\Delta_n - \Delta_{n+1} \ge \frac{1-S_n}{\|A\|}\frac{V_n}{Q_n}\tau_n\|x_n^*\|R_n \ge 0\,.$$

For all $n \in \mathbf{N}$ we have

$$0 \le \sum_{k=0}^{n}\frac{1-S_k}{\|A\|}\frac{V_k}{Q_k}\tau_k\|x_k^*\|R_k \le \sum_{k=0}^{n}(\Delta_k - \Delta_{k+1}) = \Delta_0 - \Delta_{n+1} \le \Delta_0$$

which yields that

$$\sum_{k=0}^{+\infty}\frac{1-S_k}{\|A\|}\frac{V_k}{Q_k}\tau_k\|x_k^*\|R_k < +\infty, \tag{25}$$

that is, the numerical series is convergent.

Since $S_k \ge 0$ by construction, in order to first ensure that the limit inferior of $(R_n)_n$ is zero we impose the following condition

$$0 \le S_k \le 1 - \frac{d\|A\|Q_k}{V_k} \quad \text{so that} \quad \frac{1-S_k}{\|A\|}\frac{V_k}{Q_k} \ge d > 0, \tag{26}$$

where d is a positive constant independent from n. This guarantees that the product of the first two factors of any element of the series (25) stays uniformly bounded away from zero.

The latter inequality leads to the following constructive condition on the constant d

$$1 - \frac{d\|A\|Q_k}{V_k} \ge 0$$

which is satisfied when d is chosen such that

$$0 < d \le \left(1 - \frac{2^r - 1 + r}{r}\gamma\right)\frac{1}{\|A\|} \tag{27}$$

since

$$0 < 1 - \frac{2^r - 1 + r}{r}\gamma \le \frac{V_k}{Q_k}\,. \tag{28}$$

On this ground, we can write

$$1 - \frac{d\|A\|Q_k}{V_k} \ge S_k = \frac{W\tau_k\|x_k^*\|^{q-1}}{R_k}\frac{Q_k}{V_k}\|A\|\,,$$

which is equivalent to

$$\tau_k = \frac{\alpha_k \|A\|_q R_k^{r-1} Q_k}{\|x_k^*\|} \leq \frac{(V_k - d\|A\|Q_k)\, R_k}{W\|x_k^*\|^{q-1} Q_k \|A\|},$$

that is,

$$\alpha_k \leq \frac{R_k^{2-r}\,(V_k - d\|A\|Q_k)}{W\|x_k^*\|^{q-2} Q_k^2 \|A\|^2}, \tag{29}$$

as requested by the the upper bound T_n of (10). From (25), we finally have

$$d\sum_{k=0}^{+\infty} \tau_k \|x_k^*\| R_k \leq \sum_{k=0}^{n} \frac{1-S_k}{\|A\|} \frac{V_k}{Q_k} \tau_k \|x_k^*\| R_k < +\infty. \tag{30}$$

Suppose now that $\liminf_{n\longrightarrow+\infty} R_n > 0$. Then there exists $n_0 \in \mathbf{N}$ and $\varepsilon > 0$ such that $R_n \geq \varepsilon\ \forall n \geq n_0$. In this case, from (30), we would have

$$\varepsilon d \sum_{k=n_0}^{\infty} \tau_k \|x_k^*\| \leq d \sum_{k=n_0}^{+\infty} \tau_k \|x_k^*\| R_k < +\infty.$$

Since the series is convergent, then we should have true at least one of the following two conditions:

1. $(x_k)_k$ is a null sequence.

$$\frac{1}{p}\|x\|^p = \Delta_p(0, x) = \lim_{k\longrightarrow+\infty} \Delta_p(x_k, x) < \frac{1}{p}\|x\|_p^p$$

   But this is a contradiction, so that $(x_n)_n$ cannot be vanishing.
2. the numerical sequence $(\tau_k)_n$ is a null sequence.
   Since $\tau_k = \frac{\alpha_k\|A\|R_k^{r-1}Q_k}{\|x_k^*\|}$, if $\tau_k \longrightarrow 0$ then $\alpha_k \longrightarrow 0$ because $R_k \geq \varepsilon$, and $\|x_k^*\| \leq c$.
   But the step size α_k, which solves the one-variable minimization problem (10), cannot be a vanishing sequence when the residual does not vanish, that is, if $R_k \leq \varepsilon$ definitively, as supposed. Indeed, from (14) and (23), we have that

$$\langle p_k^*, x_k - \bar{x}\rangle \leq -(1-C)R_k^r \leq -(1-C)\varepsilon^r < 0\,,$$

   which means that any functional p_k^* is a descent functional with negative slope uniformly bounded away from zero. Moreover, since the duality map j_q is continuous, then the map $\phi_k : \mathbb{R} \longrightarrow \mathbb{R}$ defined as

$$\alpha \longrightarrow \phi_k(\alpha) = \|Aj_q(x_k^* + \alpha p_k^*) - b\|^r$$

   is continuous with respect to α, with $\phi_k(0) = R_k^r \geq \varepsilon^r$ and $\phi_k'(0) \leq c < 0$, with $c = c(\varepsilon)$ a fixed constant. On this grounds, by sign permanence Theorem, we have that there exists a $\delta > 0$ such that $\phi_n(\alpha)$ is decreasing in $[0, \delta]$, so that $\alpha_k \geq \delta > 0$ for any k. So, we have necessarily that $\liminf_{n\longrightarrow+\infty} R_n = 0$.
   On the other hand, by construction of the step size based on the one dimensional minimization problem (10), we recall that $0 < R_{n+1} \leq R_n$, so that

$$\lim_{n\longrightarrow+\infty} R_n = 0. \tag{31}$$

3. Third step: the sequence $(x_n)_n$ has a Cauchy subsequence.
By boundedness of $(x_n)_n$ and $(j_p(x_n))_n$ we can find a subsequence $(x_{n_k})_k$ such that

(a) the sequence of the norm $(\|x_{n_k}\|)_k$ is convergent,
(b) the sequence $(j_p(x_{n_k}))_k$ is weakly convergent,

The sequence $(R_{n_k})_k$ is also null by (31). We want to show that $(x_{n_k})_k$ is a Cauchy (sub)sequence. By virtue of [20] Theorem 2.12 (e), we have for all $k, l \in \mathbb{N}$, with $l > k$,

$$\Delta_p(x_{n_k}, x_{n_l}) = \frac{1}{q}\left(\|x_{n_k}\|^p - \|x_{n_l}\|^p\right) + \langle j_p(x_{n_l}) - j_p(x_{n_k}), x_{n_l}\rangle .$$

The first addendum satisfies

$$(\|x_{n_k}\|^p - \|x_{n_l}\|^p) \longrightarrow 0 \quad \text{as} \quad k \longrightarrow +\infty \quad \text{for [20] Theorem 2.12 (d) (ii).}$$

Regarding the second term, we have

$$\langle j_p(x_{n_l}) - j_p(x_{n_k}), x_{n_l}\rangle = \langle j_p(x_{n_l}) - j_p(x_{n_k}), \bar{x}\rangle + \langle j_p(x_{n_l}) - j_p(x_{n_k}), x_{n_l} - \bar{x}\rangle . \tag{32}$$

Here

$$\langle j_p(x_{n_l}) - j_p(x_{n_k}), \bar{x}\rangle \longrightarrow 0 \quad \text{as} \quad k \longrightarrow +\infty \quad \text{for (b), and}$$

$$\begin{aligned}
|\langle j_p(x_{n_l}) - j_p(x_{n_k}), x_{n_l} - \bar{x}\rangle| &= \left|\sum_{n=n_k}^{n_l-1} \langle j_p(x_{n+1}) - j_p(x_n), x_{n_l} - \bar{x}\rangle\right| \\
&= \left|\sum_{n=n_k}^{n_l-1} \alpha_n \langle p_n^*, x_{n_l} - \bar{x}\rangle\right| \\
&\le \left|\sum_{n=n_k}^{n_l-1} \alpha_n \left\| j_r(Ax_n - b) + \sum_{j=0}^{n-1} c_{j,n} j_r(Ax_j - b)\right\| \|Ax_{n_l} - b\|\right| \\
&\le \sum_{n=n_k}^{n_l-1} \alpha_n R_n^{r-1} Q_n R_n = \sum_{n=n_k}^{n_l-1} \frac{\tau_n \|x_n^*\| R_n}{\|A\|} \to 0 \quad \text{as } k \to +\infty \text{ by (30).}
\end{aligned}$$

Hence we can write

$$\langle j_p(x_{n_l}) - j_p(x_{n_k}), x_{n_l}\rangle \longrightarrow 0,$$

which prove that the (sub)sequence $(x_{n_k})_k$ is a Cauchy sequence. Finally, since it is Cauchy, it converges strongly to a point $\tilde{x} \in \mathcal{X}$.

4. Forth step: the sequence $(x_n)_n$ converges strongly to the minimum p-norm solution $\bar{x}$

We know that the subsequence $(x_{n_k})_{n_k}$ converges to an element $\tilde{x} \in \mathcal{X}$. We prove now that $\tilde{x} = \bar{x}$ and $\lim_{n\longrightarrow+\infty} \|x_n - \bar{x}\| = 0$.

From $\|Ax_{n_k} - b\| = \|A(x_{n_k} - \bar{x})\| = R_{n_k} \longrightarrow 0$ as $k \longrightarrow +\infty$, then, by continuity of A, we have that $(\tilde{x} - \bar{x}) \in \mathcal{N}(A)$ and $j_p(\tilde{x}) \in \mathcal{R}(A^*)$. Hence, by recalling Theorem 2, $\tilde{x} = \bar{x}$. It remains to prove that not only the subsequent $(x_{n_k})_k$ converges to $\bar{x}$, but the full sequence $(x_n)_n$ converges strongly to $\bar{x}$.

By continuity of the Bregman distance, we can state that

$$\lim_{k\longrightarrow+\infty} \Delta_{n_k} = \lim_{k\longrightarrow+\infty} \Delta_p(\bar{x}, x_{n_k}) = \Delta_p(\bar{x}, \lim_{k\longrightarrow+\infty} x_{n_k}) = \Delta_p(\bar{x}, \bar{x}) = 0.$$

Hence the sequence $(\Delta_n)_n$ is convergent (by first point of the proof), and has a subsequence converging to zero. Theorem 2.12 (d) of [20], allows us to state that $x_n \longrightarrow \bar{x}$ strongly, that is,

$$\|x_n - \bar{x}\| \longrightarrow 0 \quad \text{as} \quad n \longrightarrow +\infty\,,$$

which concludes the proof.

4 CG and its regularization properties: the noisy case

In the previous section, we proved that the CG method, in the case $\mathcal{X} = L^p$ with $1 < p < 2$ and $\mathcal{Y} = L^r$, converges to the minimum p-norm solution of (1) for exact data $b \in \mathcal{R}(A)$. In the same setting, now we consider the case of noisy data $b_\delta \in \mathcal{Y}$ with a known noise level $\delta > 0$, that is,

$$\|b - b_\delta\| \leq \delta.$$

According to the well known semi-convergence behavior of any iterative regularization algorithm, in the case of noisy data an early stop of the iterations prevents noise amplification in the restoration process. Hence, our algorithm for noisy data is based on the discrepancy principle stopping rule, that is, the iterations are stopped as soon as $\|Ax_n - b_\delta\| \leq \tau\delta$, where $\tau > 1$ is a fixed constant value. This way, the restoration is accepted as soon as its residual is equal or less than τ times the magnitude of the noise of the (noisy) data b_δ (indeed, searching for a solution x_n such that the corresponding residual is smaller than the noise level δ is unreasonable). After introducing the algorithm, in the subsequent theorem we will prove that it belongs to the class of iterative regularization methods, by showing that its iterations either stop at or converge strongly to the minimum $p-$norm solution $\bar{x}$ of (1) as the noise level δ goes to zero.

Algorithm 2 ***Regularization by Conjugate Gradient method in Banach spaces***

Choose two constants $C \in (0,1)$ and $\tau \in (1, +\infty)$ such that $C + \tau^{-1} \in (0,1)$, and let

$$\gamma = \frac{Cr}{(2+\tau^{-1})^r - 1 + Cr} \tag{33}$$

be a relaxation parameter and d an arbitrary constant satisfying

$$0 < d \leq \left((1-\gamma)(1-\tau^{-1}) - \frac{(2+\tau^{-1})^r - 1}{r}\gamma \right) \frac{1}{\|A\|}.$$

Set $n = 0$, $x_0^* = 0$, $p_0^* = A^* j_r(b_\delta)$, $R_0 = \|b_\delta\|$, and

$$\alpha_0 = \arg \min_{\alpha \in \left[0, \frac{q^{p-1}(1-\tau^{-1})^{p-1} R_0^{p-r}}{\|A\|^p}\right]} \|Aj_q(x_0^* + \alpha p_0^*) - b_\delta\|^r, \tag{34}$$

While $R_n > \tau\delta$ **do**:

Update $n \leftarrow n + 1$.

Compute

$$\begin{cases} x_n^* = x_{n-1}^* + \alpha_{n-1} p_{n-1}^*, \\ x_n = j_q\left(x_n^*\right), \\ p_n^* = -A^* j_r\left(A x_n - b_\delta\right) + \beta_n p_{n-1}^*, \end{cases} \tag{35}$$

where

$$\alpha_n = \arg\min_{\alpha\in[0,T_n]} \|A j_q(x_n^* + \alpha p_n^*) - b_\delta\|^r \tag{36}$$

and

$$\beta_n = \gamma \frac{R_n^r}{R_{n-1}^r}, \tag{37}$$

with

$$R_n = \|A x_n - b_\delta\|,$$

$$T_n = \min\left\{ \frac{R_n^{2-r}\left(V_n - d\|A\|Q_n\right)}{G_q 2^{q-2}\|x_n^*\|^{q-2}(q-1)\|A\|^2 Q_n^2}, \frac{\|x_n^*\|}{\|A\| R_n^{r-1} Q_n} \right\},$$

$$V_n = 1 - \tau^{-1} - \frac{(2+\tau^{-1})^r - 1}{r}\gamma Q_{n-1},$$

$$Q_n = \frac{1-\gamma^{n+1}}{1-\gamma},$$

and G_q *the constant value of the Banach space defined in [20].*

End while

Theorem 4 *Let* $\mathcal{X} = L^p$ *with* $1 < p < 2$ *and* $\mathcal{Y} = L^r$*, with* $r > 1$*. The sequence of the iterations* $(x_n)_n$ *of Algorithm 2 is a regularization method.*

Proof According to [20], to show that Algorithm 2 is a regularization method we consider a sequence of "less and less" noisy data $(b_h)_{h\in\mathbb{N}}$ such that

$$\|b - b_h\| \le \delta_h,$$

with $\lim_{h\longrightarrow+\infty} \delta_h = 0,$ and $0 < \delta_{h+1} < \delta_h$. We then will prove that for any data δ_h there exists an iteration x_n such that $\|Ax_n - b_h\| < \tau\delta_h$, and that the sequence of iterates $(x_n)_n$ either stops at or converges strongly to the minimum $p-$norm solution $\bar{x}$ of (1) as the noise level δ_h goes to zero. To this aim, we consider the following adaptive setting of the noisy data $(b_h)_h$ as the iterations go on.

REPEAT, for $n = 0, 1, 2, \ldots$, the following steps:

IF for all $h > h_{n-1}$ (where $h_{-1} = 0$) the discrepancy principle is satisfied, that is,

$$\|Ax_n - b_h\| < \tau\delta_h\,, \tag{38}$$

THEN STOP the iterations;
ELSE choose the smallest integer value $h_n > h_{n-1}$ such that $\|Ax_n - b_{h_n}\| \ge \tau\delta_{h_n}$, and consider Algorithm 2 with noisy data $b_\delta = b_{h_n}$, that is

$$\begin{cases} x_n^* = x_{n-1}^* + \alpha_{n-1} p_{n-1}^* \\ x_n = j_q\left(x_n^*\right) \\ p_n^* = -A^* j_r\left(A x_n - b_{h_n}\right) + \beta_n p_{n-1}^* \end{cases} \tag{39}$$

We notice that the discrepancy principle related to δ_h is satisfied for any $h_n < h \leq h_{n-1}$, and the algorithm find a regularized solution for the associated data b_h with $h_n < h \leq h_{n-1}$. On the other hand, if the inequality (38) holds for all $h > h_{n-1}$ at a certain iteration n, then $x_n = \bar{x}$ as in the case of exact data (indeed the sequence of noise levels is such that $\lim_{h \longrightarrow +\infty} \delta_h = 0$), that is, the iterations stop at the minimum $p-$norm solution of (1).

So we have to consider only the case where inequality (38) is never fulfilled. In such a case $R_n = \|Ax_n - b_{h_n}\| > 0$ for all n. The proof follows the same steps of the noiseless case of Theorem 3, so we report the basic different arguments.

First we show that, since $R_n = \|Ax_n - b_{h_n}\| > \tau\delta_{h_n}$, the sequence of Bregman distances $(\Delta_n)_n$ is non-decreasing and convergent.

As already shown at the beginning of the first step of the proof of Theorem 3, we have

$$\Delta_{n+1} = \frac{1}{q}\|x_n^* + \alpha_n p_n^*\|^q + \frac{1}{p}\|\bar{x}\|^p - \langle x_n^*, \bar{x}\rangle - \alpha_n\langle p_n^*, \bar{x}\rangle.$$

For $n = 0$, it becomes

$$\begin{aligned}
\Delta_1 &= \frac{1}{q}\alpha_0^q\|A^* j_r(b_{h_0})\|^q + \frac{1}{p}\|\bar{x}\|^p - \alpha_0\langle j_r(b_{h_0}), A\bar{x}\rangle \\
&= \frac{1}{q}\alpha_0^q\|A^* j_r(b_{h_0})\|^q + \Delta_0 - \alpha_0\langle j_r(b_{h_0}), b_{h_0}\rangle + \alpha_0\langle j_r(b_{h_0}), b_{h_0} - b\rangle \\
&\leq \frac{1}{q}\alpha_0^q\|A\|^q R_0^{(r-1)q} + \Delta_0 - \alpha_0 R_0^r + \alpha_0 R_0^{r-1}\delta_{h_0}\,.
\end{aligned}$$

Similarly to (12), this gives the upper-bound for $\alpha_0 > 0$, in order to obtain a reduction of the Bregman distance at the first iteration. Indeed, since $R_0 > \tau\delta_{h_0}$ we have that, if

$$\frac{1}{q}\alpha_0^q\|A\|^q R_0^{(r-1)q} - \alpha_0 R_0^r + \alpha_0 R_0^r\tau^{-1} < 0 \Longleftrightarrow \alpha_0 < \frac{q^{p-1}\left(1-\tau^{-1}\right)^{p-1} R_0^{p-r}}{\|A\|^p}$$

then $\Longrightarrow \Delta_1 < \Delta_0\,.$

For $n > 0$, we have again (13)

$$\begin{aligned}
\Delta_{n+1} \leq \Delta_n + G_q\int_0^1 & \frac{(\|x_n^* + t\alpha_n p_n^*\| \vee \|x_n^*\|)^q}{t}\varrho_q\left(\frac{t\|\alpha_n p_n^*\|}{\|x_n^* + t\alpha_n p_n^*\| \vee \|x_n^*\|_q}\right) dt \\
& + \alpha_n\langle p_n^*, x_n - \bar{x}\rangle\,.
\end{aligned}$$

We study the two latter addenda of this inequality.

A) Following the analogous computation of (15), since $R_k > \tau\delta_{h_k}$ we have

$$
\begin{aligned}
\langle p_n, x_n - \bar{x}\rangle &= \langle -A^*\Big(j_r(Ax_n - b_{h_n}) + \sum_{k=0}^{n-1} c_{k,n} j_r(Ax_k - b_{h_k})\Big), x_n - \bar{x}\rangle \\
&= -\langle j_r(Ax_n - b_{h_n}), Ax_n - b_{h_n}\rangle + \langle j_r(Ax_n - b_{h_n}), b - b_{h_n}\rangle \\
&\quad - \sum_{k=0}^{n-1} c_{k,n} \langle j_r\left(Ax_k - b_{h_k}\right), Ax_n - b\rangle \\
&\le -R_n^r + R_n^{r-1}\delta_{h_n} - \sum_{k=0}^{n-1} c_{k,n} \left[\frac{1}{r}\|Ax_k - b_{h_k}\|^r \right. \\
&\quad \left. - \frac{1}{r}\|Ax_k - b_{h_k} - Ax_n + b\|^r\right] \qquad (40) \\
&\le -R_n^r + \tau^{-1}R_n^r \\
&\quad - \sum_{k=0}^{n-1} c_{k,n}\left[\frac{1}{r}R_k^r - \frac{1}{r}\big(\|Ax_k - b_{h_k}\| + \|Ax_n - b_{h_n}\| + \delta_{h_n}\big)^r\right] \\
&\le (\tau^{-1} - 1)R_n^r + \sum_{k=0}^{n-1} c_{k,n}\left[\frac{1}{r}\big(R_k + (1+\tau^{-1})R_n\big)^r - \frac{1}{r}R_k^r\right] \\
&\le (\tau^{-1} - 1)R_n^r + \frac{(2+\tau^{-1})^r - 1}{r}\sum_{k=0}^{n-1} c_{k,n}R_k^r.
\end{aligned}
$$

B) The other term (which is equivalent to (16)) is estimated as in (17), where instead of $j_r\left(Ax_n - b\right)$ and $j_r\left(Ax_k - b\right)$, we respectively consider $j_r\left(Ax_n - b_{h_n}\right)$ and $j_r\left(Ax_k - b_{h_k}\right)$, obtaining again the same bound (18) and the same inequality (19).

Now reconsidering (13), we have

$$
\begin{aligned}
\Delta_{n+1} \le \Delta_n &+ G_q\left[2^{q-2}\|x_n^*\|^{q-2}(q-1)\alpha_n^2 R_n^{2(r-1)}Q_n^2\right] \\
&+ \alpha_n\left[(\tau^{-1} - 1)R_n^r + \frac{(2+\tau^{-1})^r - 1}{r}\sum_{k=0}^{n-1} c_{k,n}R_k^r\right],
\end{aligned}
$$

so that $\Delta_{n+1} \le \Delta_n$ if

$$
\begin{aligned}
&G_q\left[2^{q-2}\|x_n^*\|^{(q-2)}\alpha_n^2 R_n^{2(p-1)}Q_n^2\right] \\
&\quad + \alpha_n\left[(\tau^{-1} - 1)R_n^r + \frac{(2+\tau^{-1})^r - 1}{r}\sum_{k=0}^{n-1} c_{k,n}R_k^r\right] \le 0\,.
\end{aligned} \qquad (41)
$$

The relaxation parameter (33) , as shown by the same computations of (23), now leads to

$$
(1-\tau^{-1})R_n^r - \frac{(2+\tau^{-1})^r - 1}{r}\sum_{k=0}^{n-1} c_{k,n}R_k^r \ge 1 - (C + \tau^{-1}) > 0\,, \qquad (42)
$$

since $C + \tau^{-1} \in (0,1)$ by definition, so that (41) holds if

$$
0 \le \alpha_n \le \min\left\{\frac{(1-\tau^{-1})R_n^r - \frac{(2+\tau^{-1})^r - 1}{r}\sum_{k=0}^{n-1} c_{k,n}R_k^r}{G_q 2^{q-2}\|x_n^*\|^{q-2}(q-1)\|A\|^2 R_n^{2(r-1)}Q_n^2}, \frac{\|x_n^*\|}{\|A\|R_n^{r-1}Q_n}\right\}, \qquad (43)
$$

$\forall n = 1, 2, 3, \ldots$, where condition (18) has been considered. Thanks to the same results of (24), the sequence $(\Delta_n)_n$ is convergent.

The second step, as well as for the noiseless case of Theorem 3, is to show that $\lim_{n \longrightarrow +\infty} R_n = 0$. Following exactly the same formal computation, the series (25) still converges. Similarly to (27) and (28), the constant d of (26) is now chosen such that

$$0 < d \leq \left((1-\gamma)(1-\tau^{-1}) - \frac{(2+\tau^{-1})^r - 1}{r}\gamma \right) \frac{1}{\|A\|} \,,$$

since now

$$0 < (1-\gamma)(1-\tau^{-1}) - \frac{(2+\tau^{-1})^r - 1}{r}\gamma \leq \frac{V_k}{Q_k} \,.$$

This leads to the same bound (29) on the step size α_n and the inequalities (30). On this ground, the reductio ad absurdum leading to (31) can be done similarly, by noticing that, thanks to (40) and (42),

$$\langle p_k^*, x_k - \bar{x} \rangle \leq -(1 - (C + \tau^{-1})) R_k^r \leq -(1 - (C + \tau^{-1}))\varepsilon^r < 0 \,,$$

that is, p_k^* would still be a descent functional with negative slope uniformly bounded away from zero.

The third step of the proof is to show that the sequence $(x_n)_n$ has a Cauchy subsequence. The proof is basically the same, where the convergence of the second addendum of (32) is given by

$$\begin{aligned}
|\langle j_p(x_{n_l}) - j_p(x_{n_k}), x_{n_l} - \bar{x} \rangle| &= \left| \sum_{n=n_k}^{n_l-1} \langle j_p(x_{n+1}) - j_p(x_n), x_{n_l} - \bar{x} \rangle \right| \\
&= \left| \sum_{n=n_k}^{n_l-1} \alpha_n \langle p_n^*, x_{n_l} - \bar{x} \rangle \right| \\
&= \left| \sum_{n=n_k}^{n_l-1} \alpha_n \langle j_r(Ax_n - b_{h_n}) + \sum_{j=0}^{n-1} c_{j,n} j_r(Ax_j - b_{h_j}), Ax_{n_l} - b \rangle \right| \\
&\leq \left| \sum_{n=n_k}^{n_l-1} \alpha_n \langle j_r(Ax_n - b_{h_n}) + \sum_{j=0}^{n-1} c_{j,n} j_r(Ax_j - b_{h_j}), Ax_{n_l} - b_{h_{n_l}} \rangle \right| \\
&\quad + \left| \sum_{n=n_k}^{n_l-1} \alpha_n \langle j_r(Ax_n - b_{h_n}) + \sum_{j=0}^{n-1} c_{j,n} j_r(Ax_j - b_{h_j}), b_{h_{n_l}} - b \rangle \right| \\
&\leq \sum_{n=n_k}^{n_l-1} \alpha_n R_n^{r-1} Q_n (R_{n_l} + \delta_{h_{n_l}}) \leq \sum_{n=n_k}^{n_l-1} \alpha_n R_n^{r-1} Q_n (R_n + \delta_{h_n}) \\
&\leq (1 + \tau^{-1}) \sum_{n=n_k}^{n_l-1} \frac{\tau_n \|x_n^*\| R_n}{\|A\|} \to 0 \quad \text{as} \quad k \to +\infty \,.
\end{aligned}$$

This ground, the proof can be completed as well as in the step four of the noiseless case.

5 Practical implementation for the step size computation

As already sketched, the classical CG method in Hilbert space gives a simple and closed form expression for the step size α_n of (10), which is the same optimal choice of the steepest descent method [2]. Indeed, in that Hilbertian case, the corresponding minimization problem

$$\arg\min_{\alpha} \|A\left(x_n + \alpha p_n\right) - b\|^2$$

is a simple one-dimensional quadratic and differentiable minimization problem, whose first derivative is linear. In our more general Banach settings, problem (10)

$$\begin{cases} \arg\min_\alpha \|Aj_q\left(x_n^* + \alpha p_n^*\right) - b\|^r = \arg\min_\alpha \Phi(\alpha) \\ \alpha \in [0, T] \end{cases}$$

does not allow a closed form explicit solution, according to Galois' results on the solution of a polynomial equation. Thus, the description of the practical implementation of the proposed algorithm leaves a practical question open.

In order to find the minimum of $\Phi(\alpha)$, we can evaluate the first derivative of $\Phi(\alpha)$

$$\Phi'(\alpha) = j_r\left(Aj_q\left(x_k^* + \alpha p_k^*\right) - b\right)^* \cdot \left((q-1)\, A\left(j_{q-1}(|x_k^* + \alpha p_k^*|) \circ p_k^*\right)\right) \tag{44}$$

where $\circ$ is the component-wise product, and we find the critical point by solving the equation $\Phi'(\alpha) = 0$. If $p \neq 2$ or $r \neq 2$, (44) is no longer a linear equation, so we cannot compute its solution explicitly. Iterative methods can be used, such as the simplest secant method

$$\alpha_{k+1} = \alpha_k - \Phi'(\alpha_k)\frac{\alpha_k - \alpha_{k-1}}{\Phi'(\alpha_k) - \Phi'(\alpha_{k-1})},\ k = 1, 2, ...$$

as first basic example. However, this approach becomes very expensive in computational load in real problems with large-scale matrices.

To avoid this high computational load, we use the recent and efficient derivative-free algorithm for bound-constrained optimization (BCDFO) developed in [11], which is based on the iterative trust-region method. The idea behind a trust region method is very simple: the $k+1$ iteration $\alpha_{k+1} = \alpha_k + s_k$ is an approximate solution of the quadratic subproblem

$$m_k(\alpha_k + s_k) = \Phi(\alpha_k) + g_k^T s_k + \frac{1}{2} s_k^T H_k s, \tag{45}$$

inside the trust region

$$\mathcal{B}_\infty(\alpha_k, \Delta_k) = \{\alpha \in \mathbb{R} | \|\alpha - \alpha_k\|_\infty \leq \Delta_k\}\,.$$

where g_k is an approximation of the gradient of Φ at point α_k , H_k is an approximation of its Hessian and Δ_k is a positive scalar representing the size of the trust region.

The point $\alpha_k + s_k$ is acceptable and Δ_k is increased if

$$\varrho_k = \frac{\Phi(\alpha_k) - \Phi(\alpha_k + s_k)}{m_k(\alpha_k) - m_k(\alpha_k + s_k)} > \eta\,,$$

where $\eta > 0$ is a suitable constant. If $\varrho_k < \eta_1$, the size of the trust region Δ_k is decreased and $s_k = 0$. The key fact is that the function m_k is evaluated not using the true derivatives but interpolating known function's values at a given set Y_k. The geometry of this space has to cover the space to guarantee the convergence from an arbitrary starting point [24].
In our case, the BCDFO is a particular implementation where bounds on the variables are handled by an active set strategy (for details and convergence proofs, see [11]). Although efficient, the algorithm BCDFO to solve the one dimensional minimization problem (10) cannot be considered as optimal, so that the appropriate computation of the step size still remains an open problem.

6 Numerical results

In this section, first numerical experiments illustrate the performance of the proposed conjugate gradient technique in Banach spaces. We choose a basic image restoration problem [2] and we compare the technique with the Landweber method in Banach spaces (2).
Consider the operator equation of the first kind

$$g(t_1, t_2) = \int\int G_1(y_1, t_1) f(y_1, y_2) G_2(y_2, t_2) dy_1 dy_2 \, , \tag{46}$$

where

$$G_1(y_1, t_1) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(y_1-t_1)^2}{2\sigma_1^2}}$$

and

$$G_2(y_2, t_2) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y_2-t_2)^2}{2\sigma_2^2}} \, .$$

The linear system $Ax = b$ obtained by the discretization of the operator equation (46) is considered. Fig.1(a) shows the exact solution $x^\dagger$ for the given right-hand side b shown in Fig.1(b).
The exact solution $x^\dagger$ presents discontinuities characterized by different dimen-

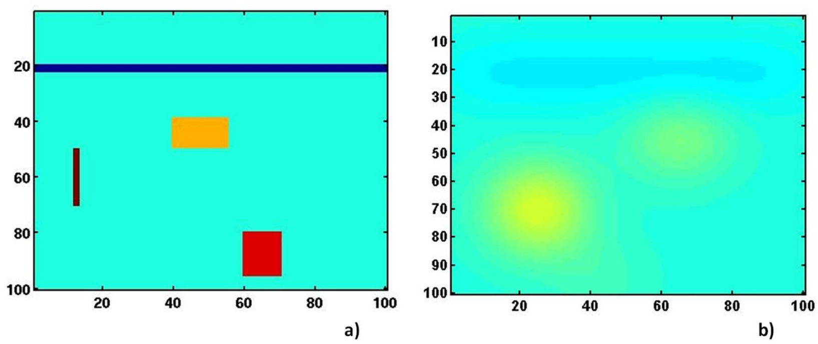

Fig. 1 Exact solution $x^\dagger$ of the linear system $Ax = b$ (a), exact data b (b).

sions and different intensities. Since the solution is almost-sparse, we look for the minimum $p-$norm solution with $p \in (1, 2)$, in order to promote sparsity in the restored solution. For more information about the role of the sparsity in inverse problems, we refer to [25], [8], [15].
To simulate noisy data b_δ, we add white Gaussian noise and measure the noise level $\|b - b_\delta\| = \delta$. When the noise is Gaussian, it is well known that the least square solution is the maximum likelihood estimate of x. For this reason, $\| Ax - b_\delta \|_2^2$ is used. In our numerical tests, the parameters of the proposed CG algorithm are selected as follows:

$\gamma = \frac{Cr}{(2+\tau^{-1})^r - 1 + Cr}$, with $C = 0.9$ and $\tau = 10$.
$d = d \leq \left((1-\gamma)(1-\tau^{-1}) - \frac{(2+\tau^{-1})^r - 1}{r}\gamma\right) \frac{1}{\|A\|} - 10^{-6}$,
G_q is evaluated using the estimate given in [20].

Fig. 2 shows the reconstructed solution by using the conjugate gradient method in l^2 (a), and by using the proposed approach in $l^{1.8}$ (b), in $l^{1.5}$ (c) and in $l^{1.2}$ (d). As stopping criterion, the discrepancy principle is used [12].
By direct visual inspection, it is clear that smaller p promotes sparsity and reconstruction of abrupt discontinuities. The reconstructed solutions in Banach spaces are less over-smoothed than the typical of Hilbert reconstructions, and few artifacts and oscillations are present, both in the spot images and in the background.

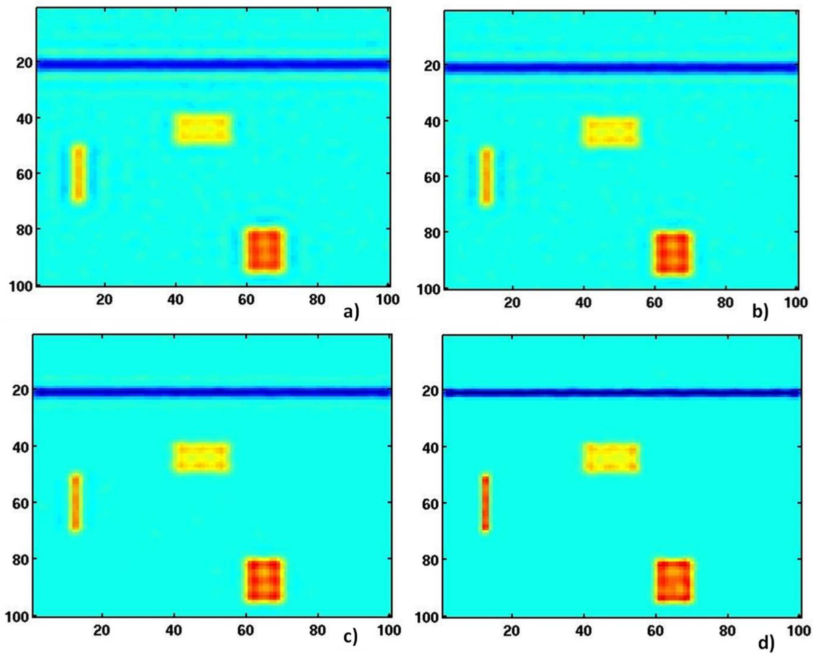

Fig. 2 Reconstructed solution using conjugate gradient in Hilbert space (a) and using the generalization of the conjugate gradient in $l^{1.8}$ (b), $l^{1.5}$ (c) and $l^{1.2}$ (d).

Table 1 RMSE for CG and Landweber methods in l^p

Method	$p = 2$	$p = 1.8$	$p = 1.5$	$p = 1.2$
CG	0.4091	0.3568	0.3246	0.2454
LW	0.3957	0.3552	0.3311	0.2452

Table 1 shows the Root Mean Square Error (RMSE):

$$RMSE(X_{rec}, X_{ref}) = \frac{\| X_{rec} - X_{ref} \|_2^2}{\| X_{ref} \|_2^2} \tag{47}$$

where X_{ref} and X_{rec} are the reference and the reconstructed solutions of Fig. 2, respectively. The RMSEs of the proposed CG and the Landweber methods are very similar, and the corresponding reconstructions look almost identical (so that, to save space, we do not show the reconstructions of the Landweber method).
Anyway, as we can see in Fig.3, the proposed CG method considerably outperforms the Landweber one both in terms of iterations number and computational time, for each value of γ (7). We recall that $\gamma = \frac{Cr}{(2+\tau^{-1})^r - 1 + Cr}$, where $c \in [0, 1)$, then, in this case, $\gamma \in [0, 0.4)$ for $r = 2$.
It must be pointed out that the computational cost of each iteration is completely different for Landweber and the proposed technique. In the latter, a 1-dimensional non-linear problem (10) must be minimized at each iteration.
The dependence on the γ and p parameters in terms of computational time is

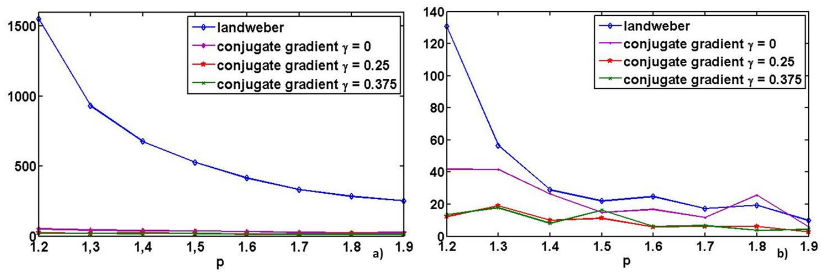

Fig. 3 Iteration number(a) and computational time (b) against the p parameter for Conjugate gradient and Landweber method. Note that the conjugate gradient method with $\gamma = 0$ is the steepest descent one.

analyzed.
In fig.4, the iteration numbers and computational time against c and p are shown. The computational time decreases for increasing p and γ values. Like in Hilbert space, the conjugate gradient method in l^p outperforms the Landweber method and the steepest descent method. These results are confirmed by different experiments (not shown here to save space) and an application of the proposed technique to a real remote sensed data could be found in [16]. On these grounds, we can summarize that the conjugate gradient method is really more efficient of the Landweber method especially for large scale matrices . Moreover, its convergence speed can be further improved by means of the applications of a dual preconditioner as in [9].

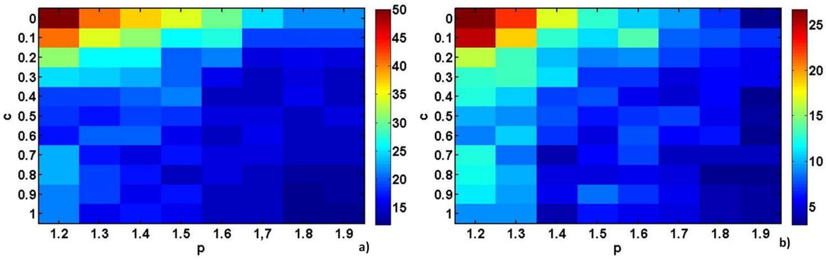

Fig. 4 Iteration number (a) and computational time (b) against the c and p parameter for Conjugate gradient method. Note that the conjugate gradient method with $\gamma = 0$ is the steepest descent one.

7 Conclusions

An iterative method, based on a generalization of the conjugate gradient method for the minimization in l^p Banach spaces is proposed.
We demonstrate that the conjugate gradient method converges strongly to the minimum $p-$norm solution and that together with the discrepancy principle as stopping rule, the proposed method is a regularization method.
Numerical experiments, undertaken on an image restoration problem, show that the method is robust in terms of reconstruction accuracy and it is faster than the Landweber method and the steepest descent one.

Acknowledgements The work of C. Estatico is partly supported by PRIN 2012 2012MTE38N and GNCS-INdAM. The work of F. Lenti is supported by the AVENUE project of the RTRA STAE foundation.

References

1. Asplund E.: Positivity of duality mappings, Bull. Amer. Math. Soc., 73, 200-203 (1967).
2. Bertero M., Boccacci P.: Introduction to inverse problems in imaging, 98-127, IOP publishing, London, 1998.
3. Bonesky K., Kazimiersky K., Maass P., Schöpfer F., Schuster T.: Minimization of Tikhonov functional in Banach spaces, Abstr. Appl. Anal., (2008).
4. Bregman L.M.: The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming, USSR Comp. Math. Math. Phys., 7, 200-217 (1967).
5. Burger M., Osher S.: Convergence rates of convex variational regularization, Inverse problems, 20, 1411-1421, (2004).
6. Butnariu D., Resmerita E.: Bregman distances, totally convex functions and a method for solving operator equations in Banach spaces, Abstr. Appl. Anal., 2006, 1-39, (2006).
7. Cioranescu I., Geometry of Banach spaces, duality mappings and nonlinear problems, 1-87, Mathematics and its Applications, 62, Kluwer Academic Publishers, Dordrecht, 1990.
8. Daubachies L., Defrise M., De Mol C.: An iterative thresholding algorithm for linear inverse problems with sparsity constraint, Comm. Pure Appl. Math., 57, 1413-1457, (2004).
9. Brianzi P., Di Benedetto F., Estatico C.: Preconditioned iterative regularization in Banach spaces, Comput. Optim. Appl., 54, 263-282, (2013).
10. Grasmair M.: Generalized Bregman distances and convergences rate for non-convex regularization methods, Inverse problems, 26, 1-20, (2010).
11. Gratton S., Toint P.L., Tröltzsch A.: An active-set trust-region method for derivative-free nonlinear bound-constrained optimization, Optim. Methods Softw., 26, 873-894, (2011).

12. Hansen P.C.: Rank-deficient and discrete ill-posed problem: numerical aspects of linear inversions, Monographs on Mathematical Modeling and Computation, SIAM, Philadelphia, 1998.
13. Herzog R., Wollner W.: A conjugate direction method for linear systems in Banach spaces, J. inverse ill-posed problems, (published online on Nov. 2016).
14. Kien B.T.: The normalized duality mapping and two related characteristic properties of a uniformly convex Banach space, Acta Math. Vietnam., 27, 53-67, (2002).
15. Lenti F., Nunziata F., Estatico C., Migliaccio M.: Analysis of reconstructions obtained solving l^p- penalized minimization problems, IEEE Trans. Geosci. Remote Sens., 53, 4876-4886, (2015).
16. Lenti F., Nunziata F., Estatico C., Migliaccio M.: Conjugate gradient method in Hilbert and Banach spaces to enhance the spatial resolution of radiometer data, IEEE Trans. Geosci. Remote Sens., 54, 397-406, (2016).
17. Lindestrauss J., Tzafriri L.: Classical Banach spaces, 53-104, Springer, Berlin, 1996.
18. Milicic P.M.: On moduli of expansion of the duality mapping of smooth Banach spaces, J. Inequ. Pure Appl. Math., 3, 1-7, (2002).
19. Resmerita E.: Regularization of ill-posed problems in Banach spaces: convergence rates, Inverse problems, 21, 1303-1314, (2005).
20. Schöpfer F., Louis A.K, Schuster T.: Nonlinear iterative methods for linear ill-posed problems in Banach spaces, Inverse problems, 22, 311-329, (2006).
21. Schöpfer F., Schuster T.: Fast regularizing sequential subspace optimization in Banach spaces, Inverse problems, 25, 1-22, (2009).
22. Schöpfer F., Schuster T., Louis A.K.: Metric and Bregman projections onto affine subspaces and their computation via sequential subspace optimization methods, J. inverse ill-posed problems, 15, 479-506, (2007).
23. Schöpfer F., Schuster T., Louis A.K.: An iterative regularization method for the solution of the split feasibility problem in Banach spaces, Inverse problems, 24, 1-20, (2008).
24. Scheinberg K., Toint, Ph.L.: Self-correcting geometry in model based algorithms for derivative-free unconstrained optimization, Technical report, TR09/06, Department of Mathematics, University of Namur, Belgium, (2009).
25. Schuster T., Kaltenbacher B., Hofmann B., Kazimierski K.: Regularization methods in Banach spaces, Radon series on Computational and Applied Mathematics, 10, de Gruyter, Berlin, 2012.
26. Xu Z.B., Roach G.F.: Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces, J. Math. Anal. Appl., 157, 189-210, (1991).