Dual descent regularization algorithms in variable exponent Lebesgue spaces

Brigida Bonino[2†], Claudio Estatico[1*†] and Marta Lazzaretti[1,3†]

[1*]Dep. of Mathematics, University of Genova, Via Dodecaneso 35, Genova, 16146, Italy.
[2]Dep. of Mechanical Engineering, DIME, University of Genova, Via All'Opera Pia, 15, Genova, 16145, Italy.
[3]I3S Laboratory, CNRS, UCA, INRIA, 2000 Route des Lucioles, Sophia Antipolis, 06903, France.

*Corresponding author. E-mail: estatico@dima.unige.it;
Contributing authors: brigida.bonino@ge.imati.cnr.it;
lazzaretti@dima.unige.it;
†These authors contributed equally to this work.

Abstract

We consider one-step iterative algorithms to solve ill-posed inverse problems in the framework of variable exponent Lebesgue spaces $\boldsymbol{L^{p(\cdot)}}$. These unconventional spaces are particular (non-Hilbertian) Banach spaces which can induce adaptive local regularization in the resolution of inverse problems. We first study gradient descent iteration schemes in Banach spaces, where the classical Riesz theorem does not hold and, consequently, primal and dual spaces are no longer isometrically isomorphic. In particular, we prove that gradient methods in Banach spaces can be fully explained and understood in the context of proximal operator theory, with appropriate norm or Bregman distances as proximity measure, which shows a deep connection between regularization iterative schemes and convex optimization. We review the key concept of duality map, and provide an explicit formula of the duality map for the space $\boldsymbol{L^{p(\cdot)}}$. Then we apply the Landweber and the Conjugate Gradient methods, extended to Banach setting, to solve deblurring imaging problems in $\boldsymbol{L^{p(\cdot)}}$ and propose an effective strategy to select the pointwise variable exponent function $\boldsymbol{p(\cdot)}$. Our numerical tests show the advantages of considering variable exponent Lebesgue spaces w.r.t. both

the standard $\boldsymbol{L^2}$ Hilbert and the constant exponent Lebesgue space $\boldsymbol{L^p}$, in terms of both reconstruction quality and convergence speed.

Keywords: Variable exponent Lebesgue spaces, iterative regularization, proximal operators, adaptive regularization.

1 Introduction

Let us consider the functional linear equation

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

where $A : \mathcal{X} \longrightarrow \mathcal{Y}$ is a compact, hence linear and continuous, operator between two infinite-dimensional Banach spaces $\mathcal{X}$ and $\mathcal{Y}$, with data $y \in \mathcal{Y}$ and solution $x \in \mathcal{X}$. It is well known that problem (1) is ill-posed and regularization methods are required in order to mitigate the effects of noisy data in the reconstruction. In the Hilbert space setting, the problem has been thoroughly studied and many methods have been established. One-step iterative algorithms, such as Landweber or Conjugate Gradient methods, represent the main class of regularization schemes for functional equations, see e.g. [1–3], where the regularization parameter is just the number of iterations. We recall that an iterative method works as regularizer if an early-stopping strategy prevents over-fitting of the noise in the reconstructions, according to the well-known semi-convergence property [4]. This is often being referred to as implicit regularization, since no penalty term has to be introduced but regularization is achieved by performing a relatively small number of iterations

Solving problem (1) in Hilbert spaces has many computational advantages but it can lead to over-smoothness of the solutions, bad reconstructions of edges and sparse patterns, such as small objects or impulse signals [5, 6]. For these reasons, in this paper we consider (1) in Banach space setting and in particular in variable exponent Lebesgue spaces $L^{p(\cdot)}$. The latter are Lebesgue spaces defined in terms of a point-wise variable exponent $p(\cdot)$ and they are intrinsically endowed with useful space-variant properties. Hence, we propose to solve ill-posed inverse problems in $L^{p(\cdot)}$ to take advantage of the natural adaptivity of these spaces. A suitable variable exponent induces indeed an adaptive regularization in the reconstruction, having the possibility to enforce sparsity and preserve edges in certain parts of the domain, as when considering a constant exponent L^p space with $1 < p < 2$ [6], and enforcing high level of regularity in other smooth parts, as in a classical Hilbertian L^2 setting.

In particular, we propose to combine variable exponent Lebesgue space setting with one-step minimization iterative strategies, which heavily depend on the definition of a gradient descent step, as we will discuss in Section 2. From this standpoint, we first study the generalization of the Landweber algorithm,

which for Hilbert spaces $\mathcal{X}, \mathcal{Y}$ and initial guess $x_0 \in \mathcal{X}$, reads as follows

$$x_{k+1} = x_k - \alpha A^*(Ax_k - y)\,,$$

to the Banach space setting, where it becomes

$$x^*_{k+1} = x^*_k - \alpha_k A^* J^{\mathcal{Y}}_p(Ax_k - y) \in \mathcal{X}^*, \qquad x_{k+1} = J^{\mathcal{X}^*}_{r^*}(x^*_{k+1}) \in \mathcal{X}.$$

Here $x^*_0 = J^{\mathcal{X}}_r(x_0) \in \mathcal{X}^*$, and $J^{\mathcal{X}}_r$, $J^{\mathcal{Y}}_p$, $J^{\mathcal{X}^*}_{r^*}$ represent the so-called duality mappings, needed to correctly move from primal to dual spaces and vice-versa [7]. We analyze such an algorithm from a different point of view, making a link between regularization theory and convex optimization. Specifically, we read such an iterative algorithm in the context of proximal methods, defined in terms of a specific functional and an appropriate Bregman distance. At the best of our knowledge, this is a novel reinterpretation which allows a full understanding of the role of the geometrical properties of both $\mathcal{X}$ and $\mathcal{Y}$ spaces.

In Section 3, basic properties and main definitions of $L^{p(\cdot)}$ spaces are first recalled, both in the continuous and in the discrete setting. Later on, we discuss the relationship between primal and dual spaces and then the analytical expression of the duality mapping of $L^{p(\cdot)}$ is given and analytically proved.

Finally, in Section 4 we consider as application the resolution of exemplar signal and image deblurring problems in $L^{p(\cdot)}$. We consider both the Landweber and the Conjugate Gradient dual methods and compare their performances with standard Hilbert reconstruction strategies. Variable exponent Lebesgue spaces yields better restorations with both algorithms, expecially in reconstructing edges, thin details and uniform background.

2 Iterative regularization dual methods in Banach spaces

In the following, we analyse one-step iterative regularization algorithms in Banach spaces. In particular, we consider the Landweber dual method [7] by both the regularization theory point of view and the convex optimization one.

2.1 The (classic) Hilbertian approach

Before introducing iterative regularization methods in Banach spaces, for the sake of simplicity we briefly start dealing with the basic Hilbertian case, that is, $\mathcal{X}$ and $\mathcal{Y}$ are now Hilbert spaces. By virtue of the Riesz isomorphism Theorem, that allows to consider the adjoint operator $A^* : \mathcal{Y}^* \longrightarrow \mathcal{X}^*$ simply as $A^* : \mathcal{Y} \longrightarrow \mathcal{X}$ up to the canonical isometric isomorphisms, the simplest iterative regularization in Hilbert spaces is the Landweber algorithm

$$x_0 \in \mathcal{X}, \qquad x_{k+1} = x_k - \alpha A^*(Ax_k - y)\,, \tag{2}$$

where $\alpha \in (0, 2/\|A\|^2)$ is a fixed step length, which always guarantees convergence for noiseless data [7]. By considering the normal equation $A^*Ax = A^*y$, the algorithm can be interpreted as a fixed point method for the operator $G(x) = x - \alpha(A^*Ax - A^*y)$, as well as a gradient descent method for the minimization of the least square *residual functional* $f : \mathcal{X} \longrightarrow \mathbb{R}$ defined as

$$f(x) = \frac{1}{2}\|Ax - y\|_{\mathcal{Y}}^2 . \tag{3}$$

Indeed, since f is convex and differentiable, with $\nabla f(x_k) = A^*(Ax_k - y) \in \mathcal{X}$, iteration (2) is the one-step descent method

$$x_{k+1} = x_k - \alpha \nabla f(x_k). \tag{4}$$

The last equation can be equivalently written as

$$\begin{aligned} x_{k+1} &= \arg\min_{x \in \mathcal{X}} \frac{1}{2}\|x - \big(x_k - \alpha \nabla f(x_k)\big)\|_{\mathcal{X}}^2 \\ &= \arg\min_{x \in \mathcal{X}} \left\{ \frac{1}{2}\|x - x_k\|_{\mathcal{X}}^2 + \alpha \langle \nabla f(x_k), x \rangle \right\}, \end{aligned} \tag{5}$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product of $\mathcal{X}$. It is quite evident that the latter minimization problem is well defined, since the argument is coercive, being the sum of a quadratic functional and a linear one. Hence, since the two addenda are convex and differentiable,

$$\nabla\Big(\frac{1}{2}\| \cdot - x_k\|_{\mathcal{X}}^2 + \alpha \langle \nabla f(x_k), \cdot \rangle\Big)(x_{k+1}) = 0 .$$

It is interesting to notice that the minimization problem (5) can be recasted in the framework of the theory of proximal operators [8]. For a fixed $x' \in \mathcal{X}$ and a continuously differentiable functional g, if we define the proximal operator $\mathrm{prox}_{\nabla g} : \mathcal{X} \longrightarrow \mathcal{X}$ as

$$\operatorname*{prox}_{\nabla g}(x') = \arg\min_{x \in \mathcal{X}} \left(\frac{1}{2}\|x - x'\|_{\mathcal{X}}^2 + \langle \nabla g(x'), x \rangle \right) , \tag{6}$$

then the one-step gradient iteration (2) can be compactly written as

$$x_{k+1} = \operatorname*{prox}_{\alpha \nabla f}(x_k) ,$$

by virtue of (4) and (5). This heuristically shows that iteration (2) corresponds to the computation of a point which decreases $\langle \nabla f(x_k), x \rangle$ and simultaneously is close (i.e., proximal) to the previous iteration. The step size α can be here thought as the weight which balances between the two terms $\frac{1}{2}\|x - x_k\|_{\mathcal{X}}^2$ and $\langle \nabla f(x_k), x \rangle$.

2.2 The dual method in Banach spaces

In [7], a generalization of the Landweber method to non-Hilbertian Banach spaces had been first proposed. Such a generalization is not straightforward, because a Banach space is not necessarily isomorphous to its dual, so that the iteration formula (5) is not even formally consistent, being $x_k \in \mathcal{X}$ summed to $\nabla f(x_k) \in \mathcal{X}^* \neq \mathcal{X}$ [9]. The key tool for the generalization to Banach spaces is the duality map, which associates an element of a Banach space $\mathcal{U}$ with an element of its dual $\mathcal{U}^*$. About that, we introduce its simplest formulation.

Definition 1 (Duality map) Let $\mathcal{U}$ be a Banach space and $p > 1$ a fixed constant. Given the convex functional $h_p(u) = \frac{1}{p}\|u\|^p_{\mathcal{U}}$, where $u \in \mathcal{U}$, its subdifferential $\partial h_p : \mathcal{U} \longrightarrow 2^{\mathcal{U}^*}$ is a multi-valued operator called *duality map* with gauge function $t \mapsto t^{p-1}$, denoted as

$$J^{\mathcal{U}}_p = \partial h_p = \partial\Big(\frac{1}{p}\|\cdot\|^p_{\mathcal{U}}\Big)\,.$$

The previous definition is based on the Asplund Theorem [6]. By direct computation, it holds

$$J^{\mathcal{U}}_p(u) = \{u^* \in \mathcal{U}^* : \ \langle u^*, u\rangle = \|u\|_{\mathcal{U}}\|u^*\|_{\mathcal{U}^*},\ \|u^*\|_{\mathcal{U}^*} = \|u\|^{p-1}_{\mathcal{U}}\}\,, \qquad \forall u \in \mathcal{U}\,,$$

where now $\langle u^*, u\rangle = \langle u, u^*\rangle = u^*(u) \in \mathbb{R}$ denotes the duality pair in Banach spaces. The latter equality can be equivalently considered as definition of duality map $J^{\mathcal{U}}_p$ [10]. We notice that $\langle u^*, u\rangle = \|u\|^p$ for any $u^* \in J^{\mathcal{U}}_p(u)$, which is formally similar to the equality $\langle h, h\rangle = \|h\|^2$ of any Hilbert space, where $J_2(h) = h$.

Duality maps satisfy important properties [10]. We just recall that, being the subdifferential of a convex functional, $J^{\mathcal{U}}_p$ is a monotone operator, that is $\langle J^{\mathcal{U}}_p(u) - J^{\mathcal{U}}_p(u'), u - u'\rangle \geq 0$, and it be can proven that $J^{\mathcal{U}}_p(-u) = -J^{\mathcal{U}}_p(u)$ and $J^{\mathcal{U}}_p(\lambda u) = \lambda^{p-1} J^{\mathcal{U}}_p(u)$ for $\lambda \geq 0$.

In this work, we focus our attention on Banach spaces with single valued duality maps, in order to obtain easy readability. To this aim, we recall that a Banach space $\mathcal{U}$ is:

i) smooth if, for every $u \neq 0$, there exists an unique $x^* \in \mathcal{U}^*$ such that $\|u^*\| = 1$ and $\langle u^*, u\rangle = \|u\|$;
ii) strictly convex if $\|\beta u_1 + (1-\beta)u_2\| < 1$, for all $\beta \in (0,1)$ and $u_1, u_2 \in \mathcal{U}$, with $u_1 \neq u_2$ and $\|u_1\| = \|u_2\| = 1$, that is, the boundary of the unit ball contains no line segment.

If (and only if) the Banach space $\mathcal{U}$ is smooth, then the duality map $J^{\mathcal{U}}_p$ is single valued. Moreover, if $\mathcal{U}$ is also reflexive and strictly convex then $\mathcal{U}^*$ is smooth and $J^{\mathcal{U}}_p$ is invertible [7]. Its inverse map is given by

$$(J^{\mathcal{U}}_p)^{-1} = J^{\mathcal{U}^*}_{p^*}\,, \tag{7}$$

where $J_{p^*}^{\mathcal{U}^*}$ is the duality map of the dual space $\mathcal{U}^*$ with gauge function $t \mapsto t^{p^*-1}$, being p^* the Hölder conjugate of p, that is $\frac{1}{p} + \frac{1}{p^*} = 1$. We highlight that in (and only in) any Hilbert space $\mathcal{H}$, for $p = 2$ the duality map reduces to the identity operator, up to canonical isometric isomorphisms, so that $J_2^{\mathcal{H}}(x) = \partial(\|\cdot\|_{\mathcal{H}}^2/2)(x) = x$.

On these grounds, for fixed parameters $p, r > 1$, the Landweber iteration scheme of the seminal paper [7] for the solution of (1) in smooth, reflexive and strictly convex Banach spaces $\mathcal{X}$ and $\mathcal{Y}$ reads as in Algorithm 1, where $x_0 \in \mathcal{X}$

Algorithm 1 Landweber (dual) method in Banach spaces

Parameters: $p, r > 1$, $\{\alpha_k\}_k$ such that $\alpha_k > 0$.
Initialization: Start with $x_0 \in \mathcal{X}$, $x_0^* = J_r^{\mathcal{X}}(x_0) \in \mathcal{X}^*$.

FOR $k = 0, 1, \ldots$ REPEAT

$$\begin{aligned} x_{k+1}^* &= x_k^* - \alpha_k A^* J_p^{\mathcal{Y}}(Ax_k - y)\,, \\ x_{k+1} &= J_{r^*}^{\mathcal{X}^*}(x_{k+1}^*)\,. \end{aligned} \tag{8}$$

is the initial guess (the null vector $x_0 = 0 \in \mathcal{X}$ can be used), $\alpha_k > 0$ is a proper variable step length and r^* is the Hölder conjugate of r.

By analogy, we notice that the descent step of the Hilbertian scheme (2) is now performed in the dual space $\mathcal{X}^*$, since both x_k^* and $A^* J_p^{\mathcal{Y}}(Ax_k - y)$ of (8) belong to $\mathcal{X}^*$. Hence, this algorithm will be referred to as *dual method*.

The dual method (8) has been introduced by a pure formal approach. Anyway, its relation with gradient methods is evident, as sketched in [6]. Indeed, by simple application of the chain rule for (sub-)differentiation of the p-power residual

$$f(x) = \frac{1}{p}\|Ax - y\|_{\mathcal{Y}}^p\,, \tag{9}$$

we have

$$\begin{aligned} \partial f(x) &= \left(\left(\partial\left(\frac{1}{p}\|\cdot\|_{\mathcal{Y}}^p\right)\Big|_{Ax-y}\right)^* \partial(Ax - y)\right)^* \\ &= \left(\left(J_p^{\mathcal{Y}}(Ax - y)\right)^* A\right)^* = A^* J_p^{\mathcal{Y}}(Ax - y)\,, \end{aligned} \tag{10}$$

where the superscript star denotes the adjunction of linear operators. This shows that iterative step (8) can be written as

$$x_{k+1}^* = x_k^* - \alpha_k \partial f(x_k)\,,$$

which is computed in the dual space $\mathcal{X}^*$, with a strong analogy with the iterative step (4) of Hilbert setting, though computed in the primal space $\mathcal{X}$ in that case.

On the contrary, the role of the duality maps $J_r^{\mathcal{X}}$ and $J_{r^*}^{\mathcal{X}^*}$ of Algorithm 1, related to the "geometry" induced by the norms of the Banach space $\mathcal{X}$ and its dual $\mathcal{X}^*$, is not evident, because both duality maps do not appear in the sub-differential (10) of the residual functional (9). Hence, we just aim at giving here a simple explanation of the whole Landweber iteration scheme (8) in Banach spaces, that is, involving the duality maps $J_r^{\mathcal{X}}$ and $J_{r^*}^{\mathcal{X}^*}$ too. Our approach of the following sections, although formally consistent, allows us an heuristic rationale which has never been explicitly introduced before.

2.3 The dual method as Bregman-proximal iterative minimization

As first study, by considering the functional (9) and its subdifferential (10) in smooth, reflexive and strictly convex Banach spaces, inspired by (5) we write

$$x_{k+1} = \arg\min_{x\in\mathcal{X}} \left\{ \frac{1}{r}\|x - x_k\|_{\mathcal{X}}^r + \alpha_k \langle \partial f(x_k), x\rangle \right\}, \tag{11}$$

which leads to

$$\partial\Big(\frac{1}{r}\|\cdot - x_k\|_{\mathcal{X}}^r + \alpha_k\langle \partial f(x_k), \cdot\rangle\Big)(x_{k+1}) = 0\,,$$

that is, to the iterative gradient method

$$J_r^{\mathcal{X}}(x_{k+1} - x_k) + \alpha_k \partial f(x_k) = 0\,,$$

where the Definition 1 of duality map has been applied to $\frac{1}{r}\|\cdot\|_{\mathcal{X}}^r$. The latter can be solved explicitly in our hypotheses, since $(J_r^{\mathcal{X}})^{-1} = J_{r^*}^{\mathcal{X}^*}$, leading to

$$x_{k+1} = x_k - \tilde{\alpha}_k J_{r^*}^{\mathcal{X}^*}(\partial f(x_k)),$$

where $\tilde{\alpha}_k = \alpha_k^{r^*-1}$, that is, by (10)

$$x_{k+1} = x_k - \tilde{\alpha}_k J_{r^*}^{\mathcal{X}^*}(A^* J_p^{\mathcal{Y}}(Ax_k - y))\,. \tag{12}$$

This iteration generalizes the so called *primal method* of [6, Section 5.3.1], (there $p = r$). The primal method (12) solves the minimization problem (11) in Banach spaces for the p-power residual functional (9), as well as (2) solves the minimization problem (5) in Hilbert spaces for the square residual functional (3). However, since (12) and (8) are different, we understand that the direct extension to Banach space of the procedure for the solution of the minimization problem (5) in Hilbert space does not give rise to what is known as the Landweber method in Banach space, that is, the dual method (8).

The correct generalization requires indeed a different way to measure distances. It is known that, in Banach spaces, Bregman distances are more

appropriate than norm distances, since Bregman distance inherits the rich geometrical properties of the involved Banach space [11]. Bregman distances are widely used to measure the distance between Ax and y instead of norm functional such as $\|Ax - y\|^p$, as well as the distance between the k-th iteration and the generalized solution in many proofs of convergence. Generally, the Bregman distance associated to a convex functional is defined as the difference between the functional and its linear approximation as follows [12].

Definition 2 (Bregman distance) Let $g : \mathcal{U} \longrightarrow \mathbb{R}$ be a convex and continuously-differentiable functional on a Banach space $\mathcal{U}$. The *Bregman distance* $B_g^{\mathcal{U}}(\cdot, u) : \mathcal{U} \longrightarrow [0, +\infty)$ of g at $u \in \mathcal{U}$ is defined as

$$B_g^{\mathcal{U}}(u', u) = g(u') - \Big(g(u) + \langle \partial g(u), u' - u\rangle\Big), \qquad \forall u' \in \mathcal{U}.$$

For $g(\cdot) = \frac{1}{r}\|\cdot\|^r$, with $r > 1$, the Bregman distance will be denoted as $B_r^{\mathcal{U}}$. Since $\partial g = J_r^{\mathcal{U}}$, it becomes

$$B_r^{\mathcal{U}}(u', u) = \frac{1}{r}\|u'\|^r - \frac{1}{r}\|u\|^r - \langle J_r^{\mathcal{U}}(u), u' - u\rangle, \tag{13}$$

If the Banach space is a Hilbertian one $\mathcal{H}$, then $B_2^{\mathcal{H}}(u', u) = \frac{1}{2}\|u' - u\|^2$. However, in general Bregman distance and conventional norm distances $\frac{1}{r}\|u' - u\|^r$ are different, since Bregman distance does not satisfy symmetry nor the triangle inequality. Some basic results about Bregman distance in a uniformly smooth and uniformly convex Banach space $\mathcal{U}$ can be found in [6], and [7, Theorem 2.12]. We remind here just a few ones.

(i) $B_r^{\mathcal{U}}(u', u) \geq 0$, and $B_r^{\mathcal{U}}(u', u) = 0$ if and only if $u = u'$;
(ii) $B_r^{\mathcal{U}}$ is continuous in both the arguments;
(iii) $B_r^{\mathcal{U}}(u', u)$ can be written as $B_r^{\mathcal{U}}(u', u) = \frac{1}{r}\|u'\|^r + \frac{1}{r^*}\|u\|^r - \langle J_r^{\mathcal{U}}(u), u'\rangle$, since $\langle J_r^{\mathcal{U}}(u), u\rangle = \|u\|^r$;
(iiv) the following three statements are equivalent
  (a) $\lim_{n\longrightarrow+\infty} \|u_n - u\| = 0$,
  (b) $\lim_{n\longrightarrow+\infty} \|u_n\| = \|u\|$ and $\lim_{n\longrightarrow+\infty}\langle J_r^{\mathcal{U}}(u_n), u'\rangle = \langle J_r^{\mathcal{U}}(u), u'\rangle$, for any $u' \in \mathcal{U}$,
  (c) $\lim_{n\longrightarrow+\infty} B_r^{\mathcal{U}}(u, u_n) = 0$.

In this regard, by using the Bregman distance $B_r^{\mathcal{X}}$ instead of the norm distance $\frac{1}{r}\|\cdot\|_{\mathcal{X}}^r$ in (11), we can write

$$\begin{aligned} x_{k+1} &= \arg\min_{x\in\mathcal{X}} \Big\{ B_r^{\mathcal{X}}(x, x_k) + \alpha_k \langle \partial f(x_k), x\rangle \Big\} \\ &= \arg\min_{x\in\mathcal{X}} \Big\{ \frac{1}{r}\|x\|_{\mathcal{X}}^r + \frac{1}{r^*}\|x_k\|_{\mathcal{X}}^r - \left\langle J_r^{\mathcal{X}}(x_k), x\right\rangle + \alpha_k \left\langle \partial f(x_k), x\right\rangle \Big\}, \end{aligned}$$

which is again convex, differentiable end coercive, so that

$$\partial\Big(\frac{1}{r}\|\cdot\|_{\mathcal{X}}^{r}+\frac{1}{r^{*}}\|x_k\|_{\mathcal{X}}^{r}-\left\langle J_r^{\mathcal{X}}(x_k),\cdot\right\rangle+\alpha_k\left\langle\partial f(x_k),\cdot\right\rangle\Big)(x_{k+1})=0\,.$$

Formally, the latter equality leads to the following iterative gradient-type iteration

$$J_r^{\mathcal{X}}(x_{k+1})-J_r^{\mathcal{X}}(x_k)+\alpha_k\partial f(x_k)=0\,,$$

which can be written as

$$x_{k+1}=J_{r^*}^{\mathcal{X}^*}\Big(J_r^{\mathcal{X}}(x_k)-\alpha_k\partial f(x_k)\Big)\,.$$

We notice that this is exactly the Landweber dual method in Banach space (8) for the minimization of the p-power residual functional (9), since the subdifferential of f is the single element (10) and $x_k^*=J_r^{\mathcal{X}}(x_k)$. Thus, we have proven the following proposition, which gives a comprehensive understanding of (8).

Proposition 1 *Let $\mathcal{X}$ and $\mathcal{Y}$ be two smooth, reflexive and uniformly convex Banach spaces. Then, for $p,r>1$, the Landweber iterative step (8) in Banach spaces, defined in [7],*

$$x_{k+1}=J_{r^*}^{\mathcal{X}^*}\Big(x_k^*-\alpha_k A^*J_p^{\mathcal{Y}}(Ax_k-y)\Big)\,,$$

for the iterative solution of the ill-posed linear equation $Ax=y$ as in (1) , corresponds to the solution of the following minimization problem

$$x_{k+1}=\arg\min_{x\in\mathcal{X}}\Big\{B_r^{\mathcal{X}}(x,x_k)+\alpha_k\langle\partial f(x_k),x\rangle\Big\}$$

where $f:\mathcal{X}\longrightarrow\mathbb{R}$ is the functional $f(x)=\frac{1}{p}\|Ax-y\|_{\mathcal{Y}}^{p}$ and $B_r^{\mathcal{X}}$ denotes the Bregman distance (13) of $\mathcal{X}$.

Based on the rationale of (6), Proposition 1 allows us to write the Landweber iterative step in Banach spaces, that is (8), as solution of a proximal problem. More precisely, by defining the Bregman-proximal operator $\operatorname{prox}_{\partial f}^{B_r^{\mathcal{X}}}:\mathcal{X}\longrightarrow\mathcal{X}$ as

$$\underset{\partial g}{\overset{B_r^{\mathcal{X}}}{\operatorname{prox}}}(x')=\arg\min_{x\in\mathcal{X}}\left(B_r^{\mathcal{X}}(x,x')+\langle\partial g(x'),x\rangle\right)\,,$$

the basic iteration (8) reduces to

$$x_{k+1}=\underset{\alpha_k\partial f}{\overset{B_r^{\mathcal{X}}}{\operatorname{prox}}}(x_k)\,,$$

which gives another useful interpretation of the dual algorithm in the context of convex optimization.

In the following sections, the dual algorithm (8) will be applied to unconventional Banach spaces, namely the variable exponent Lebesgue spaces. Before going on, we report here the properties of the (conventional, i.e. constant exponent) Lebesgue spaces $L^p(\Omega)$, with $\Omega \subseteq \mathbb{R}^d$ measurable. For $1 < p < +\infty$, $L^p(\Omega)$ is a smooth, reflexive and strictly convex Banach space (more deeply, uniformly smooth and uniformly convex too [7]). Hence the duality map $J_s^{L^p} : L^p(\Omega) \longrightarrow (L^p(\Omega))^*$ of Definition 1 is single valued, and it holds

$$\langle J_s^{L^p}(u), v\rangle = \int_\Omega \|u\|_p^{s-p} |u(t)|^{p-1} \operatorname{sign}(u(t))\, v(t)\, d\mu(t)\,, \qquad \forall u, v \in L^p(\Omega)\,,$$

where $\mu(t)$ denotes the Lebesgue measure in $\mathbb{R}^d$. By virtue of the isometric isomorphism between $(L^p(\Omega))^*$ and $L^{p^*}(\Omega)$, we simply write

$$J_s^{L^p}(u) = \|u\|_p^{s-p} |u|^{p-1} \operatorname{sign}(u)\,, \tag{14}$$

so that $J_s^{L^p}(u) \in L^{p^*}(\Omega)$ for any $u \in L^p(\Omega)$, up to the canonical isometric isomorphisms.

The discrete analogue of $L^p(\Omega)$ Lebesgue spaces are the p-summable sequence spaces l^p. For the real sequence $(x_n)_{n\in\mathbb{N}} \in l^p$, where $x_n \in \mathbb{R} \;\; \forall n$, we denote the norm as $\|x\|_p = \sqrt[p]{\sum_{n\in\mathbb{N}} |x_n|^p}$. The duality map of $x \in l^p$ with the gauge function $t \mapsto t^{s-1}$ is denoted as $J_s^p(x)$. By means of the corresponding isometric isomorphisms between $(l^p)^*$ and l^{p^*}, $J_s^p(x)$ is the real sequence

$$J_s^p(x) = \left(\|x\|_p^{s-p} |x_n|^{p-1} \operatorname{sign}(x_n) \right)_{n\in\mathbb{N}} \in l^{p^*}\,, \tag{15}$$

with a straightforward analogy with the continuous case (14) of $L^p(\Omega)$ [6].

3 The variable exponent Lebesgue spaces setting

Definitions, key concepts and fundamental properties of variable exponent Lebesgue spaces $L^{p(\cdot)}(\Omega)$ will be now reported (for a thorough dissertation about these spaces, see [13, 14]). Then, an explicit formulation of their duality maps is given, and the descent regularization algorithms in these spaces will be shown at the end of the section.

Let $\Omega \subseteq \mathbb{R}^d$, with $d \in \mathbb{N}$, $d \geq 1$, be a Lebesgue measurable subset with positive measure and let $p(\cdot) : \Omega \longrightarrow [1, +\infty]$ be a Lebesgue measurable function. The variable exponent Lebesgue space $L^{p(\cdot)}(\Omega)$ is defined in terms of the function $p(\cdot)$, which represents the exponent function. These spaces

can be used as solution spaces for ill-posed inverse problems, since they induce a spatially adaptive regularization thanks to their intrinsic space-variant geometrical properties.

3.1 Definition of Luxemburg norm

We define the set of all possible variable exponents as

$$\mathcal{P}(\Omega) := \{p(\cdot) : \Omega \longrightarrow [1, +\infty] \mid p(\cdot) \text{ is Lebesgue measurable}\}.$$

Given an exponent function $p(\cdot) \in \mathcal{P}(\Omega)$, the essential infimum and essential supremum of $p(\cdot)$ are two quantities which play an important role in the properties of the space $L^{p(\cdot)}(\Omega)$ itself and are here denoted by

$$p_- := \operatorname*{ess\,inf}_{u\in\Omega} p(u) \qquad \text{and} \qquad p_+ := \operatorname*{ess\,sup}_{u\in\Omega} p(u).$$

In the following, for the sake of simplicity we only consider exponents $p(\cdot)$ with $1 < p_- \leq p_+ < +\infty$. For the general case, see e.g. [13].

In classical $L^p(\Omega)$ spaces with a constant exponent $p \in (1, +\infty)$, for any Lebesgue measurable function $x : \Omega \longrightarrow \mathbb{R} \cup \{+\infty\}$, its norm is $\|x\|_p = \left(\int_\Omega |x(t)|^p \mathrm{d}t\right)^{1/p}$ and consequently the space is defined as the set of functions with finite p-norm: $x \in L^p(\Omega) \iff \|x\|_p < +\infty$. However, when considering a point-wise variable exponent $p(\cdot)$, the definition of the norm is not straightforward. Indeed, comparing to the above classical definition, it is possible to compute the quantity $\int_\Omega |x(t)|^{p(t)} \mathrm{d}t$ even with a variable exponent, but it is not clear which specific value of $p(\cdot)$ should be used to compute its radical. Hence, in order to characterise a norm in $L^{p(\cdot)}(\Omega)$ spaces, it is necessary to first introduce the so-called modulars.

Definition 3 Let $\mathcal{F}(\Omega)$ be the set of all Lebesgue measurable functions $x : \Omega \longrightarrow \mathbb{R} \cup \{+\infty\}$. Given $p(\cdot) \in \mathcal{P}(\Omega)$ with $p_+ < +\infty$, the function $\rho_{p(\cdot)} : \mathcal{F}(\Omega) \longrightarrow [0, +\infty]$ defined by

$$\rho_{p(\cdot)}(x) = \int_\Omega |x(t)|^{p(t)} \mathrm{d}t$$

is called modular associated to the exponent function $p(\cdot)$.

The modular $\rho_{p(\cdot)}(x)$ can be seen as the generalization of the p-power of the norm $\|x\|_p^p$ in $L^p(\Omega)$ with constant exponent $p \in (1, +\infty)$. We can now characterise the space $L^{p(\cdot)}(\Omega)$ and give a definition of its norm.

Definition 4 The space $L^{p(\cdot)}(\Omega)$ is the set of functions $x \in \mathcal{F}(\Omega)$ such that

$$\rho_{p(\cdot)}\left(\frac{x}{\lambda}\right) \leq 1,$$

for some $\lambda > 0$. For any $x \in L^{p(\cdot)}(\Omega)$, we define $\|\cdot\|_{L^{p(\cdot)}(\Omega)} : L^{p(\cdot)}(\Omega) \longrightarrow \mathbb{R}$ as

$$\|x\|_{L^{p(\cdot)}(\Omega)} := \inf\left\{\lambda > 0 : \rho_{p(\cdot)}\left(\frac{x}{\lambda}\right) \leq 1\right\}. \tag{16}$$

Theorem 2 *[13] The function $\|\cdot\|_{L^{p(\cdot)}(\Omega)}$ defined in (16) is a norm on $L^{p(\cdot)}(\Omega)$ and the space $\left(L^{p(\cdot)}(\Omega), \|\cdot\|_{L^{p(\cdot)}(\Omega)}\right)$ is a Banach space.*

The norm defined in (16) is often referred to as Luxemburg norm and it can be extended to the comprehensive framework of Orlicz spaces [13] with an analogue definition. It can be considered a general definition of norm, that is, the norm is the (smallest) scaling factor which normalizes the modular. To simplify the notation, we often will write $\|\cdot\|_{L^{p(\cdot)}}$ in the place of $\|\cdot\|_{L^{p(\cdot)}(\Omega)}$.

We can now observe that when $p(\cdot) \equiv p \in (1, +\infty)$ is constant, the classical notion of norm $\|x\|_p$ in $L^p(\Omega)$ can be easily retrieved:

$$\rho_{p(\cdot)}\left(\frac{x}{\lambda}\right) = \rho_p\left(\frac{x}{\lambda}\right) = \frac{1}{\lambda^p}\rho_p(t) = \frac{1}{\lambda^p}\|x\|_p^p$$

so that the infimum in (16) is equal to $\|x\|_p$. We remark that the computation of the p-radical of the integral is necessary to ensure the homogeneity property $\|\beta x\|_p = |\beta|\|x\|_p$, for any $\beta \in \mathbb{C}$. With a variable exponent, such computation is obviously not possible and in turn the 1-d minimization problem (16) has to be solved. However, the quantity (16) can be bounded by both the p_- and p_+ radicals of the modular. At a certain extent, we can consider the norm as a specific $\tilde{p}$-radical (which depends on x) of the modular with $p_- \leq \tilde{p} \leq p_+$, as stated by the following Proposition.

Proposition 2.1 *[13, Lemma 3.2.5][15, Th. 1] Let $p(\cdot) \in \mathcal{P}(\Omega)$ with $p_+ < +\infty$.*

(i) If $x \neq 0$, then $\|x\|_{L^{p(\cdot)}} = \beta$ if and only if $\rho_{p(\cdot)}(x/\beta) = 1$.

(ii) $\|x\|_{L^{p(\cdot)}} < 1\,(= 1, > 1)$ if and only if $\rho_{p(\cdot)}(x) < 1\,(= 1, > 1)$.

(iii) If $\|x\|_{L^{p(\cdot)}} > 1$, then $\rho_{p(\cdot)}(x)^{1/p_+} \leq \|x\|_{L^{p(\cdot)}} \leq \rho_{p(\cdot)}(x)^{1/p_-}$.

(iv) If $0 < \|x\|_{L^{p(\cdot)}} \leq 1$, then $\rho_{p(\cdot)}(x)^{1/p_-} \leq \|x\|_{L^{p(\cdot)}} \leq \rho_{p(\cdot)}(x)^{1/p_+}$.

An important consequence is that, when the essential infimum of $p(\cdot)$ is finite, $p_+ < +\infty$, then there holds that: $\rho_{p(\cdot)}(x) < +\infty \iff \|x\|_{L^{p(\cdot)}} < +\infty$. This equivalence is obvious in L^p spaces, but it does not always hold true in general $L^{p(\cdot)}$ spaces. A classical example is the following: consider $\Omega = [1, +\infty)$, $x(t) = 1$ and $p(t) = t$, $\forall t \in \Omega$, hence $p_+ = +\infty$. Then, by Definition 3, $\rho_{p(\cdot)}(x) = \int_1^{+\infty} 1^t \,\mathrm{d}t = \int_1^{+\infty} 1 \,\mathrm{d}t = +\infty$, but for all $\lambda > 1$ there holds $\rho_{p(\cdot)}\left(\frac{x}{\lambda}\right) = \int_1^{+\infty} \frac{1}{\lambda^t}\mathrm{d}t = \frac{1}{\lambda \log(\lambda)} < +\infty$, so that by Definition 4 and Theorem 2, the norm is finite, i.e., $\|x\|_{L^{p(\cdot)}} < +\infty$ (in particular, by numerical

computation, $\|x\|_{L^{p(\cdot)}} \approx 1.763$). In $L^{p(\cdot)}$ spaces, modular and norm are thus truly different objects. Finally, we collect now some useful properties.

Theorem 3 *[13, Theorem 3.4.7, Theorem 3.4.9][15, Lemma 1] Given $p(\cdot) \in \mathcal{P}(\Omega)$ such that $1 < p_- \leq p_+ < +\infty$, then $L^{p(\cdot)}(\Omega)$ is a smooth, uniformly convex, hence reflexive, Banach space.*

Proposition 3.1 *[16] Given $p(\cdot) \in \mathcal{P}(\Omega)$ and $q(\cdot) \in \mathcal{P}(\Omega)$, with $p_+ < +\infty$ and $q_+ < +\infty$, then the following natural immersion holds*

$$L^{q(\cdot)}(\Omega) \hookrightarrow L^{p(\cdot)}(\Omega)$$

if and only if

$$p(t) \leq q(t) \quad a.e.\ in\ \Omega.$$

3.2 Dual space and duality mappings

Sections 2.2 and 2.3 highlighted the key concepts of dual space and duality mapping in the development of gradient descent iteration schemes in Banach spaces. Given a Banach space $\mathcal{U}$, to define the Landweber dual method of Algorithm 1, it is necessary to have an explicit expression of both the duality mapping $J_r^{\mathcal{U}} : \mathcal{U} \longrightarrow \mathcal{U}^*$ and its inverse $\left(J_r^{\mathcal{U}}\right)^{-1} : \mathcal{U}^* \longrightarrow \mathcal{U}$. Recall that for any Banach space $\mathcal{U}$, the inverse of $J_r^{\mathcal{U}}$ is given by $\left(J_r^{\mathcal{U}}\right)^{-1} = J_{r^*}^{\mathcal{U}^*}$. In particular, in conventional fixed exponent $L^p(\Omega)$ spaces, the isometric isomorphism between the space $\left(L^p(\Omega)\right)^*$ and $L^{p^*}(\Omega)$ leads to

$$\left(J_r^{L^p}\right)^{-1} = J_{r^*}^{(L^p)^*} = J_{r^*}^{L^{p^*}},$$

so that we have an analytical expression of the inverse duality map just by using (14) or (15).

In this subsection, we define the dual space and the duality mapping of $L^{p(\cdot)}(\Omega)$, and show that, differing from the constant exponent case, an isomorphism between $(L^{p(\cdot)}(\Omega))^*$ and $L^{p^*(\cdot)}(\Omega)$ does not hold true in general. For a comprehensive review of these arguments, we refer again to [13].

Definition 5 Let $G : L^{p(\cdot)}(\Omega) \longrightarrow \mathbb{R}$ be a linear functional. G is bounded if

$$\sup\left\{|G(u)| :\ u \in L^{p(\cdot)}(\Omega), \|u\|_{L^{p(\cdot)}} \leq 1\right\} < +\infty.$$

The dual space of $L^{p(\cdot)}(\Omega)$ is the set

$$(L^{p(\cdot)}(\Omega))^* = \{G : L^{p(\cdot)}(\Omega) \longrightarrow \mathbb{R} \ :\ G \text{ is linear and bounded}\},$$

which is a Banach space with the previous finite supremum as norm $\|G\|_{(L^{p(\cdot)}(\Omega))^*}$.

For $1 < p_- \leq p_+ < +\infty$, the Hölder conjugate of $p(\cdot)$ is a Lebesgue measurable function $p^*(\cdot) \in \mathcal{P}(\Omega)$ such that

$$\frac{1}{p(t)} + \frac{1}{p^*(t)} = 1 \qquad \text{a.e. in } \Omega.$$

With a strong formal analogy to the constant exponent case, for any $z \in L^{p^*(\cdot)}(\Omega)$, there exists a unique $G \in (L^{p(\cdot)}(\Omega))^*$ such that

$$G(u) = \int_\Omega z(t)u(t)\mathrm{d}t \qquad \forall\ u \in L^{p(\cdot)}(\Omega).$$

Thus, we can denote unambiguously G as G_z.

Definition 6 [13, Definition 2.7.1] The associate space of $L^{p(\cdot)}(\Omega)$, denoted by $\mathcal{A}(L^{p^*(\cdot)}(\Omega))$, is the space of functions $z \in L^{p^*(\cdot)}(\Omega)$ such that

$$\sup\left\{\int_\Omega |z(t)||u(t)|\mathrm{d}t :\ u \in L^{p(\cdot)}(\Omega), \|u\|_{L^{p(\cdot)}} \leq 1\right\} < +\infty. \tag{17}$$

The function $\|\cdot\|'_{p^*(\cdot)} : \mathcal{A}(L^{p^*(\cdot)}(\Omega)) \longrightarrow \mathbb{R}$ defined by

$$\|z\|'_{p^*(\cdot)} := \sup\left\{\int_\Omega |z(t)||u(t)|\mathrm{d}t :\ u \in L^{p(\cdot)}(\Omega), \|u\|_{L^{p(\cdot)}} \leq 1\right\}$$

is a norm on $\mathcal{A}(L^{p^*(\cdot)}(\Omega))$.

First of all, observe that $\mathcal{A}(L^{p^*(\cdot)}(\Omega))$ might happen to be a proper subset of $L^{p^*(\cdot)}(\Omega)$. Moreover, since

$$\begin{aligned}\|G_z\|_{(L^{p(\cdot)}(\Omega))^*} &= \sup\left\{|G_z(u)| :\ u \in L^{p(\cdot)}(\Omega), \|u\|_{L^{p(\cdot)}} \leq 1\right\} = \\ &= \sup\left\{\int_\Omega |z(t)||u(t)|\mathrm{d}t :\ u \in L^{p(\cdot)}(\Omega), \|u\|_{L^{p(\cdot)}} \leq 1\right\},\end{aligned} \tag{18}$$

(17) is equivalent to requiring that the linear operator G_z is bounded. In this case the two norms of $z \in \mathcal{A}(L^{p^*(\cdot)}(\Omega))$ and $G_z \in L^{p^*(\cdot)}(\Omega)$ are the same. Thus, there exists an isometric embedding $\mathcal{A}(L^{p^*(\cdot)}(\Omega)) \hookrightarrow (L^{p(\cdot)}(\Omega))^*$ between the associate space and the dual space of $L^{p(\cdot)}(\Omega)$. However, as shown by the following Proposition combined with (18), $L^{p^*(\cdot)}(\Omega)$ and $(L^{p(\cdot)}(\Omega))^*$ are not isometrically isomorphic.

Proposition 3.2 *[13, Corollary 3.2.14] For all $z \in \mathcal{A}(L^{p^*(\cdot)}(\Omega))$, there holds*

$$\frac{1}{2}\|z\|_{L^{p^*(\cdot)}} \leq \|z\|'_{p^*(\cdot)} \leq 2\|z\|_{L^{p^*(\cdot)}},$$

and the bounds are optimal.

For any $x \neq 0$, in [15, 17] the authors proved that $\|\cdot\|_{L^{p(\cdot)}}$ is Gateaux-differentiable for $1 < p_- \leq p_+ < +\infty$ (hence the space $(L^{p(\cdot)}(\Omega), \|\cdot\|_{L^{p(\cdot)}})$ is smooth), and in [18, 19] that it is also Fréchet differentiable. From these results, following similar arguments, we explicitly derive an analytical expression for the duality mapping $J_s^{L^{p(\cdot)}}$ of $L^{p(\cdot)}(\Omega)$ spaces.

Theorem 4 *Let the exponent function $p(\cdot) \in \mathcal{P}(\Omega)$ be such that $1 < p_- \leq p_+ < +\infty$. Then, for each $x \in L^{p(\cdot)}(\Omega)$ and for any $s \in (1, +\infty)$, the duality mapping $J_s^{L^{p(\cdot)}} : L^{p(\cdot)}(\Omega) \longrightarrow (L^{p(\cdot)}(\Omega))^*$ is the linear operator such that*

$$\langle J_s^{L^{p(\cdot)}}(x), h\rangle = \frac{1}{\int_\Omega \frac{p(t)|x(t)|^{p(t)}}{\|x\|_{L^{p(\cdot)}}^{p(t)}}\mathrm{d}t} \int_\Omega \frac{p(t)\,\mathrm{sign}\,(x(t))|x(t)|^{p(t)-1}}{\|x\|_{L^{p(\cdot)}}^{p(t)-s}} h(t)\mathrm{d}t, \tag{19}$$

for any $h \in L^{p(\cdot)}(\Omega)$.

Proof By Definition 1, we know that $J_s^{L^{p(\cdot)}} = \partial\left(\frac{1}{s}\|\cdot\|_{L^{p(\cdot)}}^s\right)$. Taking into account the smoothness of $(L^{p(\cdot)}(\Omega), \|\cdot\|_{L^{p(\cdot)}})$, in the following of the proof we will focus on the computation of the Gâteaux derivative the functional $x \in L^{p(\cdot)}(\Omega) \longmapsto \|x\|_{L^{p(\cdot)}}^s$ (without the fixed scaling factor $\frac{1}{s}$ for simplicity), for any $x_0 \in L^{p(\cdot)}(\Omega)$.

We now first consider $x_0 \neq 0$. We have to prove that, for any possible direction $h \in L^{p(\cdot)}(\Omega)$, the real function $\sigma \mapsto \|x_0 + \sigma h\|_{L^{p(\cdot)}}^s$, with $\sigma \in \mathbb{R}$, is differentiable at $\sigma = 0$. We will use the implicit function theorem as follows.

Let $k > 1$ be a fixed real number, $D = (-1, 1) \times \left(\frac{1}{k}\|x_0\|_{L^{p(\cdot)}}^s, k\|x_0\|_{L^{p(\cdot)}}^s\right)$ and consider the function $\phi : D \to \mathbb{R}$ defined by means of the convex modular function $\rho_{p(\cdot)}$ which characterizes the Luxemburg norm (16)

$$\phi(\sigma, \lambda) = \rho_{p(\cdot)}\left(\frac{x_0 + \sigma h}{\lambda^{1/s}}\right) - 1 = \int_\Omega \frac{|x_0(t) + \sigma h(t)|^{p(t)}}{\lambda^{p(t)/s}}\mathrm{d}t - 1. \tag{20}$$

In the sequel, we will demonstrate the following statements, which are the hypothesis of the implicit function Theorem:

i) $\phi \in C^1(D)$;
ii) $\phi\left(0, \|x_0\|_{L^{p(\cdot)}}^s\right) = 0$;
iii) $\frac{\partial \phi}{\partial \lambda}\left(0, \|x_0\|_{L^{p(\cdot)}}^s\right) < 0$.

Indeed, once proven i), ii) and iii), the implicit function Theorem guarantees that there exist neighbourhoods U of 0 and V of $\|x_0\|_{L^{p(\cdot)}}^s$ such that $U \times V \subset D$ and a unique C^1-mapping $\lambda : U \to V$ which satisfies $\lambda(0) = \|x_0\|_{L^{p(\cdot)}}^s$, $\phi(\sigma, \lambda(\sigma)) = 0$ for any $\sigma \in U$, and

$$\lambda'(\sigma) = -\frac{\frac{\partial \phi}{\partial \sigma}(\sigma, \lambda(\sigma))}{\frac{\partial \phi}{\partial \lambda}(\sigma, \lambda(\sigma))}, \quad \forall \sigma \in U\,. \tag{21}$$

The equality $\phi(\sigma, \lambda(\sigma)) = 0\ \forall \sigma \in U$, rewritten as

$$\rho_{p(\cdot)}\left(\frac{x_0 + \sigma h}{\lambda(\sigma)^{1/s}}\right) = 1, \quad \forall \sigma \in U\,,$$

together with Definition 4 of the norm in $L^{p(\cdot)}(\Omega)$, allows us to derive that

$$\lambda(\sigma) = \|x_0 + \sigma h\|^s_{L^{p(\cdot)}}, \quad \forall \sigma \in U. \tag{22}$$

Hence, from (21) and (22) we have that $\lambda'(0)$ exists and

$$\lambda'(0) = \lim_{\sigma \to 0} \frac{\|x_0 + \sigma h\|^s_{L^{p(\cdot)}} - \|x_0\|^s_{L^{p(\cdot)}}}{\sigma} = -\frac{\frac{\partial \phi}{\partial \sigma}\left(0, \|x_0\|^s_{L^{p(\cdot)}}\right)}{\frac{\partial \phi}{\partial \lambda}\left(0, \|x_0\|^s_{L^{p(\cdot)}}\right)}. \tag{23}$$

The functional $\|\cdot\|^s_{L^{p(\cdot)}}$ is thus Gâteaux differentiable at $x_0 \neq 0$, and the explicit computation of the ratio (23) will provide expressions (19) too.

We can now prove the statements i), ii) and iii).

i) To prove that $\phi \in C^1(D)$, let us consider the integrand $f : \Omega \times D \to \mathbb{R}$ of (20)

$$f(t;\ (\sigma, \lambda)) = \frac{|x_0(t) + \sigma h(t)|^{p(t)}}{\lambda^{p(t)/s}}, \quad t \in \Omega, \quad (\sigma, \lambda) \in D. \tag{24}$$

It is easy to show that, for any fixed $(\sigma, \lambda) \in D$, the map $t \mapsto f(t;\ (\sigma, \lambda))$ is integrable in Ω. Indeed, by definition of D, there hold $|\sigma| < 1$ and $\lambda \geq \frac{1}{k}\|x_0\|^s_{L^{p(\cdot)}} = \lambda_{\min} > 0$, which yields to

$$\frac{|x_0(t) + \sigma h(t)|^{p(t)}}{\lambda^{p(t)/s}} \leq \frac{k^{p(t)/s}(|x_0(t)| + |h(t)|)^{p(t)}}{\|x_0\|^{p(t)}_{L^{p(\cdot)}}} \leq \frac{k^{p_+/s}}{c}(|x_0(t)| + |h(t)|)^{p(t)}$$

with $c = \min\left(\|x_0\|^{p_-}_{L^{p(\cdot)}}, \|x_0\|^{p_+}_{L^{p(\cdot)}}\right)$ and $(|x_0(t)| + |h(t)|)^{p(t)}$ being integrable since $x_0,\ h \in L^{p(\cdot)}(\Omega)$ and $p_+ < +\infty$. Consequently, the function ϕ of (20) is well-defined.
We now show that for a.e. $t \in \Omega$, the map $(\sigma, \lambda) \mapsto f(t;\ (\sigma, \lambda))$, with $(\sigma, \lambda) \in D$, is a C^1-mapping. By formal computation, the partial derivatives of (24) are

$$\frac{\partial f}{\partial \sigma}(t;\ (\sigma, \lambda)) = \frac{p(t)\,|x_0(t) + \sigma h(t)|^{p(t)-1}\,\mathrm{sign}\,(x_0(t) + \sigma h(t))\ h(t)}{\lambda^{p(t)/s}}, \tag{25}$$

$$\frac{\partial f}{\partial \lambda}(t;\ (\sigma, \lambda)) = -\frac{p(t)\,|x_0(t) + \sigma h(t)|^{p(t)}}{s\lambda^{p(t)/s+1}}, \quad \forall (\sigma, \lambda) \in D. \tag{26}$$

Since $p_- > 1$ and $\lambda > 0$, it is evident from (25) and (26), that $(\sigma, \lambda) \mapsto \frac{\partial f}{\partial \sigma}(t;\ (\sigma, \lambda))$ and $(\sigma, \lambda) \mapsto \frac{\partial f}{\partial \lambda}(t;\ (\sigma, \lambda))$ are continuous mappings in D. Anyway, to explicitly compute both the numerator and the denominator of (21), that is, the partial derivatives of (20), we need to commute differentiation and integration operators. To this aim, we apply the Dominated Convergence Theorem, by searching for a function $g : \Omega \longrightarrow \mathbb{R}$, integrable on Ω, such that

$$\left|\frac{\partial f}{\partial \sigma}(t;\ (\sigma, \lambda))\right| \leq g(t), \qquad \left|\frac{\partial f}{\partial \lambda}(t;\ (\sigma, \lambda))\right| \leq g(t).$$

Similarly as before for the estimation of $|f(t;\ (\sigma,\lambda))|$, $(\sigma,\lambda)\in D$ implies that

$$\left|\frac{\partial f}{\partial\sigma}(t;\ (\sigma,\lambda))\right| \leq \frac{k^{p(t)/s}p(t)\left(|x_0(t)|+|h(t)|\right)^{p(t)}}{\|x_0\|^{p(t)}_{L^{p(\cdot)}}}$$
$$\leq \frac{p_+\cdot k^{p_+/s}}{c}\left(|x_0(t)|+|h(t)|\right)^{p(t)},$$

with $c=\min\left(\|x_0\|^{p_-}_{L^{p(\cdot)}},\|x_0\|^{p_+}_{L^{p(\cdot)}}\right)$, and that

$$\left|\frac{\partial f}{\partial\lambda}(t;\ (\sigma,\lambda))\right| \leq \frac{p_+\cdot k^{p_+/s+1}}{c_1}\left(|x_0(t)|+|h(t)|\right)^{p(t)}$$

with $c_1=\min\left(\|x_0\|^{p_-+s}_{L^{p(\cdot)}},\|x_0\|^{p_++s}_{L^{p(\cdot)}}\right)$. Thus, we can now consider

$$g(t)=\max\left(\frac{p_+\cdot k^{p_+/s}}{c},\frac{p_+\cdot k^{p_+/s+1}}{c_1}\right)\left(|x_0(t)|+|h(t)|\right)^{p(t)},$$

as dominating function, which is integrable on Ω, as already stated before. Hence differentiation and integration in ϕ commute, leading to

$$\frac{\partial\phi}{\partial\sigma}(\sigma,\lambda)=\frac{\partial}{\partial\sigma}\left[\int_\Omega f(t;\ (\sigma,\lambda))\,\mathrm{d}t-1\right]=\int_\Omega\left[\frac{\partial}{\partial\sigma}f(t;\ (\sigma,\lambda))\right]\mathrm{d}t$$
$$=\int_\Omega p(t)\frac{|x_0(t)+\sigma h(t)|^{p(t)-1}\,\mathrm{sign}\left(x_0(t)+\sigma h(t)\right)}{\lambda^{p(t)/s}}h(t)\mathrm{d}t, \quad (27)$$
$$\frac{\partial\phi}{\partial\lambda}(\sigma,\lambda)=\frac{\partial}{\partial\lambda}\left[\int_\Omega f(t;\ (\sigma,\lambda))\mathrm{d}t-1\right]=\int_\Omega\left[\frac{\partial}{\partial\lambda}f(t;\ (\sigma,\lambda))\right]\mathrm{d}t$$
$$=-\int_\Omega\frac{p(t)\cdot|x_0(t)+\sigma h(t)|^{p(t)}}{s\lambda^{p(t)/s+1}}\mathrm{d}t. \quad (28)$$

From (27) and (28), the continuity of $\frac{\partial\phi}{\partial\sigma}$ and $\frac{\partial\phi}{\partial\lambda}$ is straightforward.

ii) By Definition 4 of Luxemburg norm and Proposition 2.1 (i),

$$\phi\left(0,\|x_0\|^s_{L^{p(\cdot)}}\right)=\int_\Omega\left|\frac{x_0(t)}{\|x_0\|_{L^{p(\cdot)}}}\right|^{p(t)}\mathrm{d}t-1=0.$$

iii) We have similarly

$$\frac{\partial\phi}{\partial\lambda}\left(0,\|x_0\|^s_{L^{p(\cdot)}}\right)=-\int_\Omega p(t)\frac{|x_0(t)|^{p(t)}}{\|x_0\|^{p(t)+s}_{L^{p(\cdot)}}}\mathrm{d}t$$
$$\leq-\frac{p_-}{\|x_0\|^s_{L^{p(\cdot)}}}\int_\Omega\left|\frac{x_0(t)}{\|x_0\|_{L^{p(\cdot)}}}\right|^{p(t)}\mathrm{d}t=-\frac{p_-}{\|x_0\|^s_{L^{p(\cdot)}}}<0.$$

We can now obtain formula (19) from (23), plugging $\sigma = 0$ into (27) and (28). Recall that we are computing the gradient of $\|\cdot\|^s_{L^{p(\cdot)}}$, thus a multiplication by $\frac{1}{s}$ is needed to obtain (19).

To conclude the proof, it remains to consider the case $x_0 = 0$. The real function $\sigma \mapsto \|x_0 + \sigma h\|^s_{L^{p(\cdot)}}$, with $\sigma \in \mathbb{R}$, becomes $\sigma \mapsto \|\sigma h\|^s_{L^{p(\cdot)}}$. We have

$$\lim_{\sigma\to 0} \frac{\|x_0 + \sigma h\|^s_{L^{p(\cdot)}} - \|x_0\|^s_{L^{p(\cdot)}}}{\sigma} = \lim_{\sigma\to 0} \frac{\|\sigma h\|^s_{L^{p(\cdot)}}}{\sigma} = \lim_{\sigma\to 0} \frac{|\sigma|^s \|h\|^s_{L^{p(\cdot)}}}{\sigma} = 0,$$

since $s > 1$, which proves the differentiability at the origin as well. □

It is easy to check that, if $p(\cdot) \equiv p$ is constant, with $1 < p < +\infty$, then $J^{L^{p(\cdot)}}_s$ coincides with $J^{L^p}_s$, the duality map (14) of classical (i.e. constant exponent) Lebesgue spaces $\langle J^{L^p}_s(x), h\rangle = \|x\|^{s-p}_p \int_\Omega \operatorname{sign}\big(x(t)\big)|x(t)|^{p-1}h(t)\mathrm{d}t$-

3.3 The discrete case: variable exponent sequence spaces

As well as for $L^p(\Omega)$ spaces, which have their discrete analogue l^p, as sketched at the end of Section 2.3, the same holds for $L^{p(\cdot)}(\Omega)$ spaces with variable exponent. Let $\mathcal{P}$ be the set of real sequences $(p_n)_{n\in\mathbb{N}}$, where $p_n \in \mathbb{R} \quad \forall n$, such that $1 < p_- = \inf_{n\in\mathbb{N}} p_n \leq p_+ = \sup_{n\in\mathbb{N}} p_n < +\infty$. $\mathcal{P}$ contains all the sequences we consider as variable exponents in the discrete setting. Given $(p_n)_{n\in\mathbb{N}} \in \mathcal{P}$, for any real sequence $x = (x_n)_{n\in\mathbb{N}}$, where $x_n \in \mathbb{R}\ \forall n$, the modular function of Definition 3 is here naturally defined as

$$\rho_{(p_n)}(x) = \sum_{n\in\mathbb{N}} |x_n|^{p_n}. \tag{29}$$

Definition 7 [13] The space $l^{(p_n)}$ is the set of real sequences $x = (x_n)_{n\in\mathbb{N}}$ such that $\rho_{(p_n)}\left(\frac{x}{\lambda}\right) < 1$ for some $\lambda > 0$. For any $x = (x_n)_{n\in\mathbb{N}} \in l^{(p_n)}$, the (Luxemburg) norm on $l^{(p_n)}$ is defined as

$$\|x\|_{(p_n)} := \inf\left\{\lambda > 0 :\ \rho_{(p_n)}\left(\frac{x}{\lambda}\right) \leq 1\right\}$$

and the space $\left(l^{(p_n)}, \|\cdot\|_{(p_n)}\right)$ is a Banach space.

Obviously, in this Banach space is defined the duality map of Definition 1. The duality map of $x \in l^{(p_n)}$ with the gauge function $t \mapsto t^{s-1}$, for $s > 1$, is a linear operator $J^{l^{(p_n)}}_s(x) : l^{(p_n)} \longrightarrow \left(l^{(p_n)}\right)^*$ and can be obtained rewriting (19) in the discrete form. Its explicit formula can be proved similarly to Theorem 4, since the formula and its proof is the generalization to the functional $\frac{1}{s}\|x\|^s_{(p_n)}$ of the Fréchet derivative of the norm $\|x\|_{(p_n)}$ introduced in [19].

Theorem 5 *Given $(p_n)_{n\in\mathbb{N}} \in \mathcal{P}$ a variable exponent, that is, with $1 < p_- \leq p_+ < +\infty$, then $\left(l^{(p_n)}, \|x\|_{(p_n)}\right)$ is smooth. For each $x = (x_n)_{n\in\mathbb{N}} \in l^{(p_n)}$ and for any*

$s > 1$, the duality mapping $J_s^{l^{(p_n)}}(x) : l^{(p_n)} \longrightarrow (l^{(p_n)})^*$ is the linear operator such that

$$\langle J_s^{l^{(p_n)}}(x), h\rangle = \frac{1}{\sum_{n\in\mathbb{N}} \frac{p_n|x_n|^{p_n}}{\|x\|_{(p_n)}^{p_n}}} \sum_{n\in\mathbb{N}} \frac{p_n \operatorname{sign}(x_n)|x_n|^{p_n-1}}{\|x\|_{(p_n)}^{p_n-s}} h_n. \tag{30}$$

for any $h = (h_n)_{n\in\mathbb{N}} \in l^{(p_n)}$.

Proof We just sketch the proof, being similar to the one of Theorem 4, reduced to the discrete case as in [19]. By Definition 1, we know that $J_s^{l^{(p_n)}} = \partial\left(\frac{1}{s}\|\cdot\|_{(p_n)}^s\right)$. Hence the proof consists in the computation of the gradient of the functional $x \in l^{(p_n)} \longmapsto \frac{1}{s}\|\cdot\|_{(p_n)}^s$. With the same rationale and steps of the proof of Theorem 4, for any fixed $x \in l^{(p_n)}$ and any possible direction $h \in l^{(p_n)}$, the function $\sigma \in \mathbb{R}, \sigma \mapsto \|x + \sigma h\|_{(p_n)}^s$ is proven to be differentiable at $\sigma = 0$, leading to the explicit formula (30). □

We denote by $(p_n^*)_{n\in\mathbb{N}}$ the Hölder conjugate sequence of $(p_n)_{n\in\mathbb{N}}$, that is, $\frac{1}{p_n} + \frac{1}{p_n^*} = 1$ for any $n \in \mathbb{N}$. Again with a strong analogy with the arguments of Section 3.2, there exists an isometric embedding between $\left(l^{(p_n)}\right)^*$ and its associate space $\mathcal{A}(l^{(p_n^*)}) \subseteq l^{(p_n^*)}$, which allows to write

$$J_s^{l^{(p_n)}}(x) = \frac{1}{\sum_{n\in\mathbb{N}} \frac{p_n|x_n|^{p_n}}{\|x\|_{(p_n)}^{p_n}}} \left(\frac{p_n \operatorname{sign}(x_n)|x_n|^{p_n-1}}{\|x\|_{(p_n)}^{p_n-s}}\right)_{n\in\mathbb{N}} \in l^{(p_n^*)}. \tag{31}$$

Anyway, as well as there is not an isometric isomorphism between $L^{p^*(\cdot)}(\Omega)$ and $(L^{p(\cdot)}(\Omega))^*$ (as stated by Proposition 3.2), in the discrete case there is not an isometric isomorphism between $l^{(p_n^*)}$ and $\left(l^{(p_n)}\right)^*$ too.

According to (7), with (14) and (15), the isometric isomorphism allowed us to obtain an exact and explicit formula for the inverse of the duality mapping in the classical case of $L^p(\Omega)$ and l^p spaces with constant exponent $p \in (1, +\infty)$ only. In particular, we can write

$$\left(J_s^{L^p}\right)^{-1} = J_{s^*}^{(L^p)^*} = J_{s^*}^{L^{p^*}} \quad \text{and} \quad \left(J_s^{l^p}\right)^{-1} = J_{s^*}^{(l^p)^*} = J_{s^*}^{l^{p^*}},$$

with s^* Hölder conjugate of s. In order to be able to use Algorithm 1 also in variable exponent spaces $L^{p(\cdot)}(\Omega)$ or $l^{(p_n)}$, where each isomorphism is not isometric, we nevertheless consider the following approximations

$$\left(J_s^{L^{p(\cdot)}}\right)^{-1} = J_{s^*}^{(L^{p(\cdot)})^*} \approx J_{s^*}^{L^{p^*(\cdot)}} \quad \text{and} \quad \left(J_s^{l^{(p_n)}}\right)^{-1} = J_{s^*}^{(l^{(p_n)})^*} \approx J_{s^*}^{l^{(p_n^*)}}.$$

These approximations provide an inexact but explicit formula of the duality maps for the dual space $l^{(p_n^*)}$. Although Proposition 3.2 states that the norms of $L^{p^*(\cdot)}(\Omega)$ and $(L^{p(\cdot)}(\Omega))^*$ are not isometric, it shows one-half and double size bounds among them. Unfortunately, since the duality maps are

the Fréchet derivative of the s-power norm functional of Definition 1, these bounds do not give any quantitative information about the goodness of the approximations. Anyway, due to continuity arguments, we can say that the approximation should be good for small ranges $[p_-, p_+]$ of exponent values, since for $p_- = p_+$, which coincides with the constant exponent case, the equality holds. In Algorithm 2 we rewrite the Landweber (dual) method for the discrete sequence case, which will be used in our numerical tests to solve (1) with $A \in \mathcal{L}(l^{(p_n)}, l^{(p_n)})$ and $y \in l^{(p_n)}$.

Algorithm 2 Landweber (dual) method in $l^{(p_n)}$

Parameters: $s, r > 1$, $(p_n)_{n\in\mathbb{N}} \in \mathcal{P}$.

Initialization: Start with $x_0 \in l^{(p_n)}$, $x_0^* = J_r^{l^{(p_n)}}(x_0) \in l^{(p_n^*)}$.

FOR $k = 0, 1, \ldots$ REPEAT

$$x_{k+1}^* = x_k^* - \alpha_k A^* J_r^{l^{(p_n)}}(Ax_k - y)\,,$$
$$x_{k+1} = J_{s^*}^{l^{(p_n^*)}}(x_{k+1}^*)\,,$$

where

$$\alpha_k = \arg\min_{\alpha>0} \|AJ_{s^*}^{l^{(p_n^*)}}\left(x_k^* - \alpha A^* J_r^{l^{(p_n)}}(Ax_k - y)\right) - y\|_{l^{(p_n)}}^r\,,$$

We remark that both the duality maps $J_r^{l^{(p_n)}}$ and $J_{s^*}^{l^{(p_n^*)}}$ can be computed by (31), so that the algorithm is completely implementable in closed form.

The same tools can be used to extend other one step minimization algorithms to Banach spaces, by exploiting their geometrical properties in order to enhance sparsity or quality of the edges. The basic idea is again to compute any iterative step in the dual space. We report in Algorithm 3 the extension of the conjugate gradient method, whose convergence in conventional constant exponent Lebesgue spaces l^p has been studied in [20].

The computation of the optimal step size α_k in $l^{(p_n)}$ is an expensive task for both algorithms, due to the involved Luxemburg norms. In the next numerical section we will discuss some sub-optimal approximations schemes for α_k, which require a much lower numerical complexity.

4 Numerical results

In this section we shoe numerically the regularization effectiveness of descent algorithms in variable exponent Lebesgue spaces for imaging reconstruction.

The idea of working in these particular Banach spaces comes from the attempt to combine simultaneously the benefits of both the classical methods in Hilbert space L^2 and the more recent approaches in conventional Banach

Algorithm 3 Conjugate gradient (dual) method in $l^{(p_n)}$

Parameters: $s, r > 1$, $(p_n)_{n\in\mathbb{N}} \in \mathcal{P}$, $0 < \gamma < \frac{1}{2}$.

Initialization: $x_0 \in l^{(p_n)}$, $x_0^* = J_r^{l^{(p_n)}}(x_0) \in l^{(p_n^*)}$, $d_0^* = A^* J_r^{l^{(p_n)}}(y)$.

FOR $k = 0, 1, \ldots$ REPEAT

$$x_{k+1}^* = x_k^* + \alpha_k d_k^*,$$
$$x_{k+1} = J_{s^*}^{l^{(p_n^*)}}(x_{k+1}^*),$$
$$d_{k+1}^* = -A^* J_r^{l^{(p_n)}}(Ax_{k+1} - y) + \beta_{k+1} d_k^*,$$

where

$$\alpha_k = \arg\min_{\alpha>0} \|AJ_{s^*}^{l^{(p_n^*)}}(x_k^* + \alpha d_k^*) - y\|_{l^{(p_n)}}^r,$$

and

$$\beta_{k+1} = \gamma \frac{\|Ax_{k+1} - y\|_{l^{(p_n)}}^r}{\|Ax_k - y\|_{l^{(p_n)}}^r}$$

spaces L^p. In fact, the exponent p, with $1 < p \leq 2$, can be heuristically considered as a regularization parameter. The blurred image is reconstructed with a different level of regularization according to the value of p: on the one hand, if $p \approx 1$, the restoration is more sparse, details and edges of the image are sharper, the background has low ringing effects, but the algorithm results to be less numerically stable; on the other hand, if $p \approx 2$ the algorithm is numerically stable, but the restored image is generally oversmoothed. Thus, to obtain a reliable reconstruction, each region of the input image should be differently regularised (e.g. depending on whether the region is just background, or there are small objects, edges, or large flat brightness areas). In the following subsections, an adaptive strategy is provided, where at each point of the blurred image a specific exponent p is automatically associated.

4.1 A first signal deblurring test

Before showing image reconstruction examples, we start considering a simple 1D reconstruction of an heterogeneous signal, composed of spikes on the left and a smooth function on the right, see Figure 1 (left, black line). Given a Gaussian convolution operator $A : \mathcal{X} \longrightarrow \mathcal{Y}$ and the noisy and blurred data $y \in \mathcal{Y}$, we compute the first 2000 iterations of the minimization Algorithm 1 for the (regularized) solution $x \in \mathcal{X}$ of (1). The blurred data is corrupted by about 10% of Gaussian white noise, see Figure 1 (left, red line). Since we are in a test environment, the data sets always include the true signal x_{true}, in order to compare and evaluate the performances of the implemented algorithms in terms of the relative reconstruction errors (RREs), namely computing at each iteration k the quantity

$$\epsilon_k = \frac{\|x_k - x_{true}\|_2}{\|x_{true}\|_2}, \tag{32}$$

where x_k is the restored signal at the k-th iteration and $\|\cdot\|_2$ is the conventional l^2 norm.

In this first test, the simplest sub-optimal constant step size $\alpha_k = \alpha = 1/\|A\|_2^2$, suitable in conventional l^2 Hilbert context, is always applied. Anyway, we will discuss a more advanced step size choice in the following section related to image deblurring. In Table 1 we report on the left the minimum value of such RREs ϵ_k, and on the right the minimum value of the residual, also called cost function, (9), all computed by the same $L^{1.5}$ norm for comparison purposes. By means of Algorithm 1 (always with norm powers $p = r = 2$), we first show the conventional Hilbert space l^2 algorithm (i.e. $\mathcal{X} = \mathcal{Y} = l^2$) in the first row of the table. In the subsequent rows, we consider $\mathcal{X} = l^p$ constant exponent or $l^{(p_n)}$ variable exponent Banach spaces with $\mathcal{Y} = l^2$, as well as $\mathcal{X} = \mathcal{Y} = l^p$ or $\mathcal{X} = \mathcal{Y} = l^{(p_n)}$. Generally, the $l^{(p_n)}$ restorations lead to smaller RREs ϵ_k in about $100 - 200$ iteration, while the Hilbert and l^p ones do not achieve their minimum RREs within 2000 iterations. For instance, for the Hilbert case $\mathcal{X} = \mathcal{Y} = l^2$, the smallest RRE is 0.4883 and for the constant exponent Banach space $\mathcal{X} = \mathcal{Y} = l^{1.5}$, the smallest RRE is 0.3952, both at iteration 2000. On the other hand, for $\mathcal{X} = \mathcal{Y} = l^{(p_n)}$ with $p_- = 1.3$ and $p_+ = 2$, the smallest RRE 0.4594 is obtained at iteration 75. Similar comments can be done for the values of the cost function (9) on the right part of the table.

$\mathcal{X}$	$\mathcal{Y}$	$[p_-, p_+]$	*Min rel. error*		*Min cost func.*	
			value	*iteration*	*value*	*iteration*
l^2	l^2		0.4883	2000	203.4	2000
$l^{1.5}$	l^2		0.5335	2000	550.9	2000
$l^{(p_n)}$	l^2	$[1.3, 2]$	0.4189	151	739.3	87
$l^{(p_n)}$	l^2	$[1.5, 2]$	0.4700	149	603.1	110
$l^{(p_n)}$	l^2	$[1.4, 1.8]$	0.4296	235	632.1	205
$l^{1.5}$	$l^{1.5}$		0.3952	2000	198.8	2000
$l^{(p_n)}$	$l^{(p_n)}$	$[1.3, 2]$	0.4594	75	325.3	73
$l^{(p_n)}$	$l^{(p_n)}$	$[1.5, 2]$	0.4580	139	368.0	116
$l^{(p_n)}$	$l^{(p_n)}$	$[1.4, 1.8]$	0.4357	153	356.9	117

Table 1: Minimum relative reconstruction error (32) and minimum value of the cost function (9) with $p = 1.5$, of the first numerical test (see Fig. 1), for different choices of $\mathcal{X}$ and $\mathcal{Y}$.

In Figure 1, left side, the usual behaviour of l^2 Hilbert space reconstruction is shown in green. It is enough good in the smooth part on the right, but strongly bad in the spiked part on the left. The l^p reconstruction, with $p = 1.5$, is shown in the blue plot of Figure 1, right side. Now the situation is the opposite, the result is enough good in the piked part but strongly bad in the smooth part, since small exponents $1 < p < 2$ promote sparsity. The attempt of combine the good performances of both these two approaches by working in $l^{(p_n)}$ spaces can be viewed in the red plot of Figure 1, right side. Here, we can see that the restoration is good on both the two parts of the signal. The exponent function p_n is shown by dot-dashed black line: in the sparse part it

is a rescaling between $p_- = 1.3$ and $p_+ = 2$ of the blurred data, in order to enhance sparsity on the background regions, whilst in the smooth part is fixed to 2, in order to benefit from smoothness of Hilbertian setting.

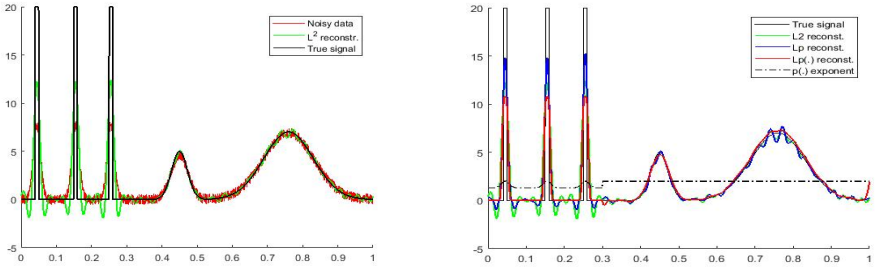


Figure 1: Left plots: true signal, blurred data of the first numerical test and $\mathcal{X} = \mathcal{Y} = l^2$ Hilbert restoration within 2000 iterations. Right plots: restoration with minimum cost function within 2000 iterations, for $\mathcal{X} = \mathcal{Y} = l^2$, $\mathcal{X} = \mathcal{Y} = l^{1.5}$, and $\mathcal{X} = \mathcal{Y} = l^{(p_n)}$. The exponent function is plotted in dot-dashed line with $p_- = 1.3$ and $p_+ = 2$.

In Tables 2 we report the results of the restoration of another signal, with peaks of different magnitudes and sign, as shown in Figure 2. To corroborate the effectiveness of the variable exponent approach, we directly minimize the modulus $\tilde{\rho}_{(p_n)}(\cdot)$ of the residual $Ax_k - y$, where

$$\tilde{\rho}_{(p_n)}(z) = \sum_{n \in \mathbb{N}} \frac{|z_n|^{p_n}}{p_n}.$$

Although the modular is not a norm, it is convex and we can consider its subdifferential as well in the place of the duality map of Algorithm 1 [21]. The modular leads to a simpler and faster implementation, and allow to confirm the flexibility and the robustness of the variable exponent approaches. In Table 2 the best results within 2000 iterations are given, and some values are also reported in Table 3 at the fixed iteration 500. The restoration of $l^{(p_n)}$ are always better than both the l^2 Hilbert and the l^p constant exponent ones, whether the relative error or the cost function are considered. The exponent function p_n is again shown by dot-dashed black line. Now it is set to $p_- = 1.3$ in the spiked part to enhance sparsity and $p_+ = 1.6, 2, 2.5$ in the smooth part to enhance regularity. In Fig. 2, left side, the restorations with the smallest RRE are shown, for $\mathcal{X} = \mathcal{Y} = l^2$, $\mathcal{X} = \mathcal{Y} = l^{1.3}$ and $\mathcal{X} = \mathcal{Y} = l^{(p_n)}$, with $p_- = 1.3$ and $p_+ = 2$. The $l^{(p_n)}$ restoration (red line) has the lowest ringing effects and oscillations in both the spiked and the smooth part, while the l^2 restoration (green line) is good on the smooth part only, and the $l^{1.3}$ restoration (blue line) is good on the spiked part only, as expected. The same behaviour can

be seen on the right plots of the figure, where the results at iteration 500 are shown. We highlight here the valuable quality of the restored signal for the $l^{(p_n)}$ setting. The corresponding numerical values are reported in Table 3. The corresponding convergence histories of the relative error and the cost function are shown in Figure 3. These plots show that, although the mimization of the cost function (9) behaves similarly (left image), the relative errors (32) are different (right image), expecially for the Hilbertian l^2 case. This is a typical behaviour of iterative regularization methods for ill-posed problems, such as signal and image deblurring.

$\mathcal{X}$	$\mathcal{Y}$	$[p_-, p_+]$	*Min rel. error*		*Min cost func.*	
			value	*iteration*	*value*	*iteration*
l^2	l^2		0.3770	2000	162.52	2000
$l^{1.3}$	l^2		0.2305	2000	158.64	2000
$l^{1.5}$	l^2		0.2767	2000	159.26	2000
$l^{(p_n)}$	l^2	[1.3, 1.6]	0.2230	2000	158.80	2000
$l^{(p_n)}$	l^2	[1.3, 2]	0.2201	2000	158.99	2000
$l^{(p_n)}$	l^2	[1.3, 2.5]	0.2195	2000	159.44	2000
$l^{1.3}$	$l^{1.3}$		0.2402	1222	157.75	2000
$l^{1.5}$	$l^{1.5}$		0.2722	2000	158.25	2000
$l^{(p_n)}$	$l^{1.6}$	[1.3, 1.6]	0.2219	2000	158.06	2000
$l^{(p_n)}$	$l^{1.3}$	[1.3, 2]	0.2174	1744	157.91	2000
$l^{(p_n)}$	$l^{1.5}$	[1.3, 2.5]	0.2143	2000	158.76	2000
$l^{(p_n)}$	$l^{(p_n)}$	[1.3, 1.6]	0.2180	1875	157.95	2000
$l^{(p_n)}$	$l^{(p_n)}$	[1.3, 2]	0.2103	2000	158.68	2000
$l^{(p_n)}$	$l^{(p_n)}$	[1.3, 2.5]	0.2084	2000	159.95	2000

Table 2: Minimum relative errors (32) and minimum values of the cost function (9), with $p = 1.5$, of the second numerical test (see Fig. 2), for different choices of $\mathcal{X}$ and $\mathcal{Y}$.

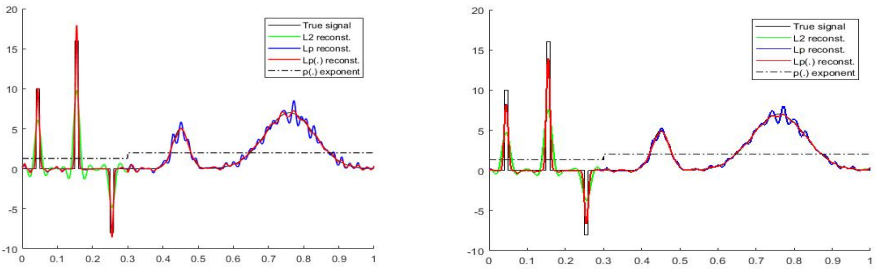


Figure 2: Left plots: restorations with minimum relative errors within 2000 iterations. Right plots: restorations at iteration 500. Second numerical test for $\mathcal{X} = \mathcal{Y} = l^2$, $\mathcal{X} = \mathcal{Y} = l^{1.3}$, and $\mathcal{X} = \mathcal{Y} = l^{(p_n)}$. The exponent function is plotted in dot-dashed line, with $p_- = 1.3$ and $p_+ = 2$.

$\mathcal{X}$	$\mathcal{Y}$	$[p_-, p_+]$	*Rel. error*	*Cost func.*
l^2	l^2		0.4209	188.64
$l^{1.3}$	$l^{1.3}$		0.2670	160.20
$l^{(p_n)}$	$l^{(p_n)}$	$[1.3, 2]$	0.2583	161.94

Table 3: Relative reconstruction errors (32) and cost functions (9) with $p = 1.5$, at the iteration 500 for the second test (see Fig. 2, left plots)

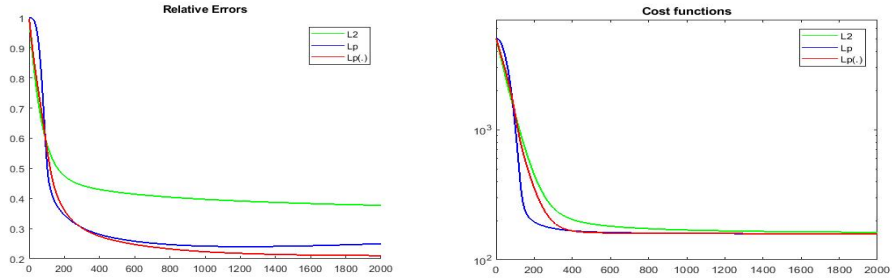


Figure 3: Converge histories of the example of Figure 2. Left plots: relative restoration errors (32). Right plots: cost function values (9), with $p = 1.5$.

4.2 Image deblurring tests

The well-known discrete model of the imaging reconstruction problem associated to (1) is given by the linear equation $Ax = y$, where A is the square matrix version of the Point Spread Function (PSF) of the acquisition tool, x is (the vectorization of) the unknown (true) image, and y is (the vectorization of) the blurred and noisy image. We consider here just monochromatic images, where the light intensity is associated to grey levels.

Since the introduction of variable exponent Lebesgue spaces for imaging is independent on the specific descent algorithm applied, we now provide results of the two Algorithms 2 and 3, namely the Landweber and Conjugate Gradient (CG) methods. After discussing the choice of the exponent function and of the step size, we show the results related to two different numerical data sets. The first is the largely used "satellite" data set, the latter is a "rectangles" data set we have specifically designed to test the algorithms.

4.2.1 Setting the exponent values

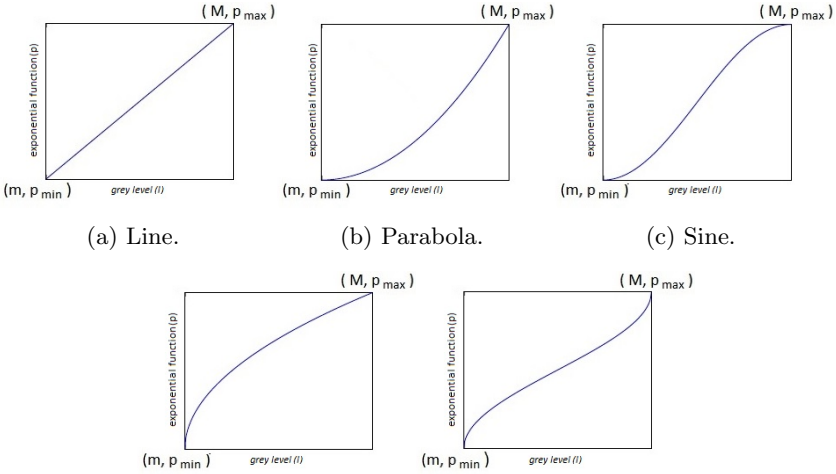
The key step for the application of descent methods in $L^{p(\cdot)}$ is the definition of the exponent function $p(\cdot)$. In particular, in image restoration the exponent sequence $p = (p_n)$ is (the vectorization of) a matrix with the same dimensions of the images x and y, that is, p is a finite sequence reshaped as matrix.

In our proposal, the values of the exponent matrix p are assigned according to the grey levels, that is to the light brightness of each pixels of the blurred data y, by some interpolation procedure. We assume here that the background

is characterized by dark pixels, that is, with the lowest values of light brightness, as in astronomical imaging. Hence, let m and M denote respectively the minimum and the maximum values of the blurred image y. According to the notation of Section 3.3, given p_- and p_+, with $1 < p_- \leq p_+ \leq 2$, respectively the chosen and prefixed minimum and the maximum values of p, the exponent matrix p is constructed as follows: p_- is assigned to the pixels with the lowest brightness intensity m, p_+ is assigned to the pixels with the highest intensity M, and the other values between p_- and p_+ are assigned by interpolation. This way, low grey levels of the images are associated with small values of the exponent p (useful for background restoration and sparsity promotion), whilst high grey levels are associated with large values of p (useful for stability and correct restoration of the right brightness intensities in flat regions). Basically, the interpolating function can be arbitrary chosen, with the only assumption that it must be an increasing function passing through (m, p_-) and (M, p_+). At this purpose, we have identified and tested the following 5 elementary functions endowed with different basic interpolation behaviours:

(a) line on $[0, 1]$,
(b) parabola on $[0, 1]$,
(c) sine $[-\frac{\pi}{2}, \frac{\pi}{2}]$,
(d) square root on $[0, 1]$,
(e) arc-sine on $[-1, 1]$,

properly dilated onto (m, p_-) and (M, p_+) by linear homotopy (see Fig. 4).



(a) Line. (b) Parabola. (c) Sine.

(d) Square root. (e) Arc-sine

Figure 4: Plots of the functions used as interpolation functions.

With regard to the choice of the p_- and p_+, the stability of the algorithms heavily depends on the dimension and type of signal or image given as input. As already stated for the constant exponent case in [22], higher instabilities generally emerge for exponents $p < 1.2$. As a consequence, in the following tests we just consider $p_- \geq 1.2$.

4.2.2 Setting the step size

The implementation of any one-step descent method generally involves the one-dimensional minimization of the residual required for the computation of the optimal step size, that is, the solution of the minimization problem for obtaining α_k in our Algorithms 2 and 3. In the Hilbertian context, from basic numerical linear algebra arguments, we know that the constant choice $\alpha_k = \alpha \in (0, 2/\|A\|^2)$, although not optimal, ensures convergences of Algorithm 2, while the optimal explicit formula $\alpha_k = \|A^*(Ax_k - y)\|^2/\|Ad_k\|^2$ is provided for Algorithm 3, namely the CGLS method.

In Banach spaces there is not any explicit formula, so that the minimization problem has to be numerically solved. Anyway, in the constant exponent case L^p, the norm is easy and fast to compute, so that the minimization can be efficiently implemented. On the contrary, in the variable exponent framewrok, the computation of the Luxemburg norm of Definition 7 is expensive, so that the solution of the minimizaton problem for the step size heavily increases the computational time of each iteration.

To overcome this issue, we have tested some alternative strategies that avoid the use of the $l^{(p_n)}$ norm. Among these, we have empirically proved that the problem can be addressed using the values α'_k computed minimising the (constant exponent) l^p norm with $p = p_{av}$, where p_{av} is the a weighted average, with respect to the current iteration, of the p_n values defined as

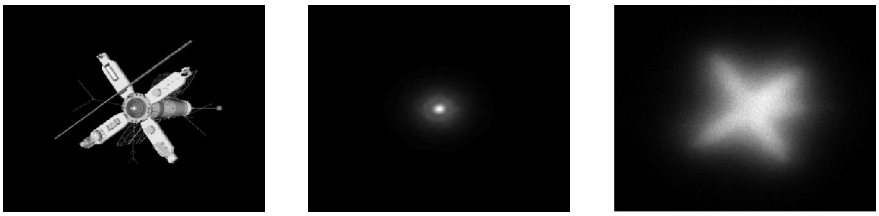
$$p_{av} = \frac{\log(\rho_{(p_n)}(x_k))}{\log(\|x_k\|_{(p_n)})}, \tag{33}$$

where $\rho_{(p_n)}$ is the modular function (29). Basically, p_{av} is the constant radical which, locally on x_k, would give the same value of the Luxemburg norm if applied to the modular, that is $\|x_k\|_{(p_n)}^{p_{av}} = \rho_{(p_n)}(x_k)$.

4.2.3 Satellite data set

In this section the widespread 256×256 satellite data set of Fig. 5 is used as reference deblurring problem. The noise on the blurred image is white Gaussian of about 5% [9].

We start dealing with Landweber dual Algorithm 2, where the simplest sub-optimal constant step size $\alpha_k = \alpha = 1/\|A\|_2^2$ of the previous 1D test is again applied. Fig. 6, left plot, shows the relative reconstruction errors (32) (RREs) (32) among some choices of all the tested values p_- and p_+, in the first 150 iterations. We see that, in this more involving 2D case, the largest



(a) Exact image (b) PSF (c) Blurred image

Figure 5: Satellite data set.

size of the exponent interval $[p_-, p_+] = [1.2, 2]$ leads to instabilities and fast semiconvergence (blue line). The choice $[p_-, p_+] = [1.2, 1.3]$ is able to provide both stability and low RREs (light blue line). In Table 4 we report the RREs (32) every 50 iterations using the different interpolating functions with $p_- = 1.2$ and $p_+ = 1.3$.

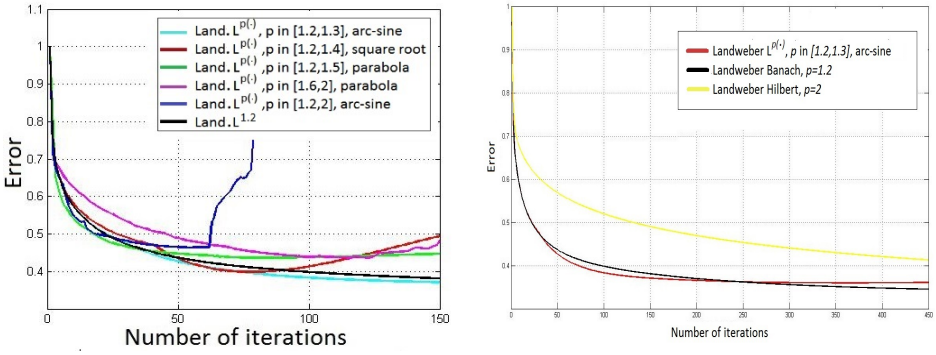


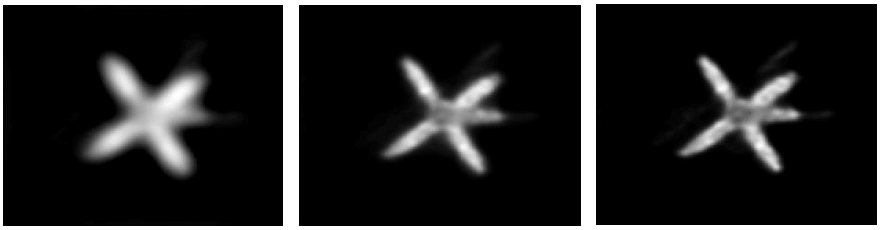
Figure 6: RREs (32) vs number of iterations for Landweber algorithm in $L^{p(\cdot)}$. Left plots: $L^{p(\cdot)}$ with different intervals $[p_-, p_+]$ and different interpolation functions. Right plots: $L^{p(\cdot)}$ with $[p_-, p_+] = [1.2, 1.3]$ and arc-sine interpolation, compared with constant exponent spaces $L^{1.2}$ and L^2.

In Fig. 6, right plot, RREs (32) of the first 450 iterations are shown, with the use of the arc-sine as interpolating function, which has been found to be the best promising one. The graph is compared with the RREs obtained by the Landweber algorithms in constant exponent Lebesgue space $L^{1.2}$ and in conventional Hilbert L^2 one. It is evident that the Landweber method in the space $L^{p(\cdot)}$ completely outperforms the classical approach in the Hilbert space, but also allows to obtain a smaller reconstruction error in the first iterations than working in $L^{1.2}$.

Figure 7 shows the images resulting after 150 iterations. The reconstructions are consistent with what is deduced from the study of the RREs. As a

Iteration	50	100	150	200
Interpolation	*Relative error*			
Line	0.4334	0.3858	0.3766	0.3815
Parabola	0.4295	0.3874	0.3741	0.3680
Sine	0.4246	0.3849	0.3753	0.3733
Square root	0.4253	0.3813	0.3751	0.3812
Arc-sine	0.4265	0.3831	0.3708	0.3656
$L^{1.2}$	0.4364	0.3982	0.3809	0.3698
L^2	0.5686	0.5196	0.4901	0.4692

Table 4: RREs (32) every 50 iterations for the Landweber algorithm in $L^{p(\cdot)}$ with $[p_-, p_+] = [1.2, 1.3]$, using the 5 interpolating functions of Fig. 4.



(a) Landweber in L^2 (b) Landweber in $L^{1.2}$ (c) Landweber in $L^{p(\cdot)}$

Figure 7: Reconstructions by Landweber algorithm at iteration 150 in L^2, in $L^{1.2}$, and in $L^{p(\cdot)}$ with $1.2 \leq p \leq 1.3$ and arc-sine interpolation.

matter of fact, quality of the deblurred image Fig. 7c is better than Fig. 7a in L^2, and although minimal, there are also visual benefits compared to reconstruction in $L^{1.2}$ of Fig. 7b. In particular, edges are sharper and thin details, such as the diagonal shaft, are more evident in Fig. 7c than Fig. 7b.

Dealing with the Conjugate Gradient (CG) method of Algorithm 3, it is expected to speed-up the convergence, hence to reduce the number of iterations with respect the Landweber one, which is generally very slow. For the CG, we study the computation of the step size α_k, which is a critical tasks. First we compare the RREs of the method with respect to the usage of the exact step size α_k and its approximation α'_k obtained by minimizing the (constant exponent) $L^{p_{av}}$ norm of the residual, where p_{av} is the exponent average (33).

Surprisingly enough, the results obtained are very similar, both in terms of the values of the step sizes and with respect to the quality of the reconstructed images, although the time for each iteration decreases from 40 sec to 12 sec by using the sub-optimal choice α'_k. This behaviour can be well observed in Fig. 8, where the RREs of CG Algorithm 3 with α_k and α'_k for $k = 0, ..., 150$ almost completely overlap each other. Moreover, in Fig. 9 the computed values of α_k and α'_k are showed. The trend of the two sequences is very similar, except for

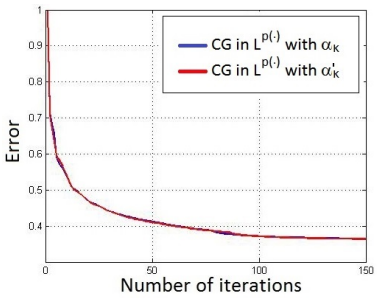


Figure 8: RREs (32) for CG in $L^{p(\cdot)}$ using step sizes α_k and α'_k.

few indexes, which denotes a reliable outcome. According to these considerations, the tests we consider in the following always will use the sequences of values $\{\alpha'_k\}$ as descent steps, since much faster to compute.

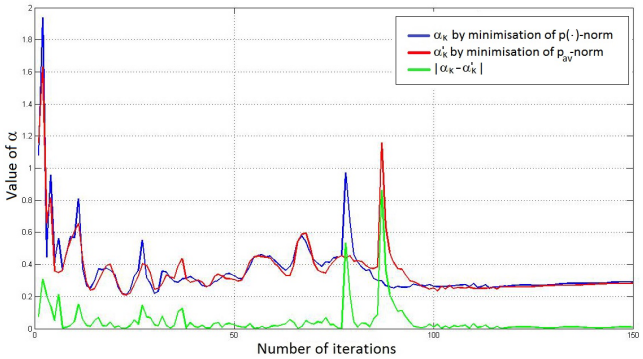


Figure 9: α_k values obtained minimising the residual in norm $p(\cdot)$, α'_k values obtained minimising the residual in norm p_{av} (33), and their distances.

Iteration	40	80	120	150
Interpolation	*Relative error*			
Line	0.4218	0.3852	0.3668	0.3640
Parabola	0.4328	0.3906	0.3681	0.3665
Sine	0.4205	0.3925	0.3761	0.3633
Square root	0.4228	0.3817	0.3712	0.3755
Arc-sine	0.4178	0.3824	0.3655	0.3569
L^2	0.3564	0.4104	0.6115	1.0104
Landweber	0.4485	0.4022	0.3768	0.3766

Table 5: RREs (32) for the CG algorithm in $L^{p(\cdot)}$ with $1.2 \leq p \leq 1.3$, using the 5 interpolation functions of Fig. 4.

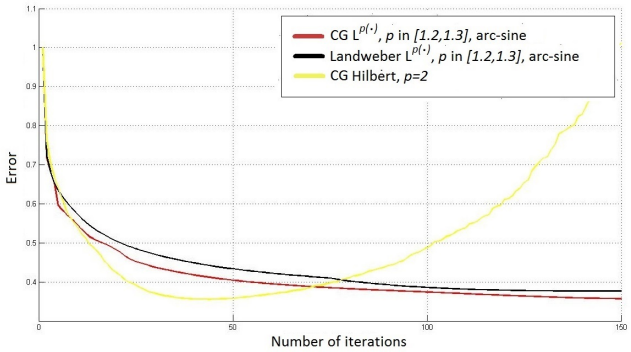
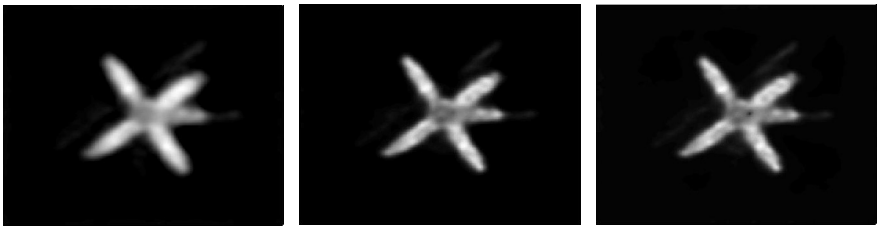


Figure 10: RREs (32) for CG and Landweber in $L^{p(\cdot)}$, with $1.2 \leq p \leq 1.3$ and arc-sine interpolation, and conventional CG in L^2.

Again considering $1.2 \leq p \leq 1.3$, Table 5 reports the RREs with respect to the use of the 5 interpolation functions of Section 4.2.1, which confirms that arc-sine outperforms all the other choices. The convergence history is shown in Fig. 10 for CG and Landweber both in $L^{p(\cdot)}$, together with the conventional CG method in L^2. In practice, both the methods in variable exponent Lebesgue space are not faster than the classic CG method in L^2. However, $L^{p(\cdot)}$ gives rise to higher regularization stability, limiting the phenomenon of semi-convergence, which in Hilbert causes a fast growth of the error. The restorations after 150 iterations are shown in Figure 11. It can be seen that the reconstruction of the CG in $L^{p(\cdot)}$ (Fig. 11c) is slightly better than the reconstruction obtained with Landweber (Fig. 11b), since small details in the arms of the satellite are more visible.

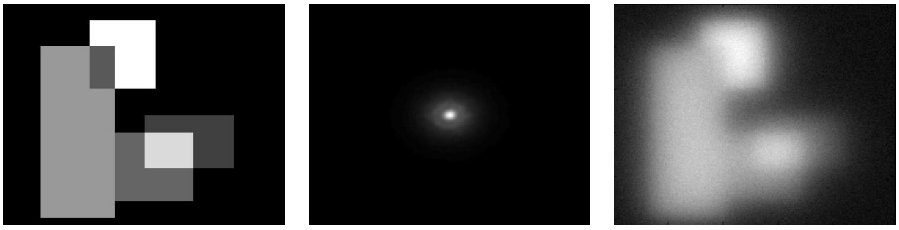


(a) CG in L^2 (b) Landweber in $L^{p(\cdot)}$ (c) CG in $L^{p(\cdot)}$

Figure 11: Reconstructions by CG at iteration 150 in L^2 and in $L^{p(\cdot)}$ with $1.2 \leq p \leq 1.3$ and arc-sine interpolation.

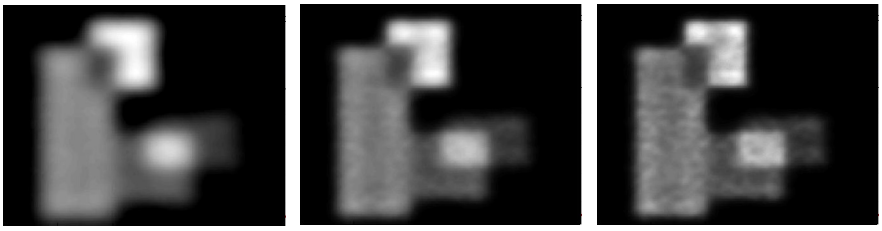
4.2.4 Rectangles data set

A second data set is used to focus the analysis on both edges detection and smooth reconstruction of uniform areas. The data set is hence characterised by (*i*) large flat regions with constant grey value and (*ii*) high discontinuities, as shown in Figure 12, where the PSF is the same of the satellite data set.



(a) Exact image (b) PSF (c) Blurred image

Figure 12: Rectangles data set.



(a) Landweber in L^2 (b) Landweber in $L^{p(\cdot)}$ (c) $L^{p(\cdot)}$ with update

Figure 13: Reconstructions by Landweber algorithm at iteration 100 in L^2, in $L^{p(\cdot)}$ with $1.3 \leq p \leq 1.6$, arc-sine interpolation and updated square root interpolation.

By our experiments, a good interval range for p is $[p_-, p_+] = [1.3, 1.6]$. In Figure 13a the restoration with Landweber in L^2 can be compared to the restoration in $L^{p(\cdot)}$ of Fig. 13b, again with arc-sine as interpolation function, both after 100 iterations. Figure 13c instead shows the reconstructed image obtained using the square root as interpolation function, and employing an update of the exponent matrix p every n iterations, with $n = 30$. With this updating strategy, the iterative scheme does not change, but the matrix p is redefined after each n iterations relying on the $(n-1)$-th reconstruction instead of the blurred fixed data y. By means of the $(n-1)$-th reconstruction, it is possible to take advantage of less blurred images to associate optimally the smaller values of p to the small grey levels and the higher values of p to the

high grey levels, making the details more defined and the background more uniform. This technique is very promising if applied for the reconstruction of images with discontinuities and uniform areas, such as the rectangles image. The benefit of working in the variable exponent Lebesgue spaces is therefore evident in Fig. 13. In particular the reconstructed figures are more defined than in L^2, especially the rectangle on the right, since edges are sharper and flat areas are less affected by ringing.

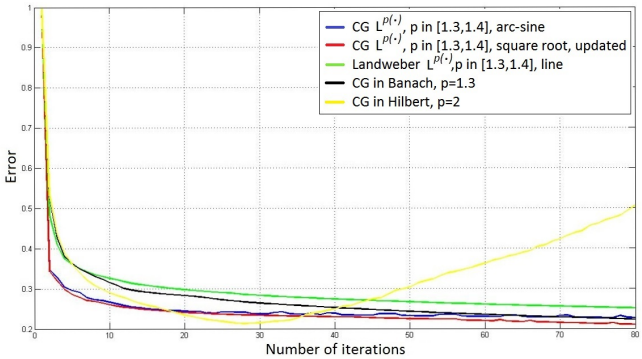
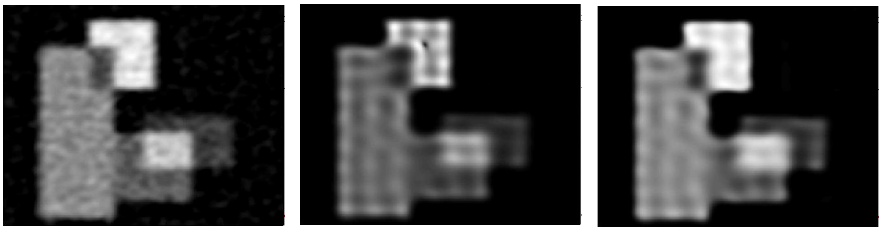


Figure 14: RREs (32) for CG in various $L^{p(\cdot)}$ cases, compared with CG in $L^{1.3}$, and CG in L^2.

Regarding the CG method, the situation is analogous to the satellite data set. Indeed, while the use of the CG in the Hilbert space is very fast, but also suffers from the fast semi-convergence (i.e. now after about 30 iterations the error increases), the use of the CG in $L^{p(\cdot)}$ allows a better stability of the algorithm, as shown by the convergence history of Fig. 14. In addition, the main advantages are in the ability to restore the homogeneity of the areas with uniform grey level and of the background. Figure 15 shows the best image resulting from the use of the CG in L^2 after 30 iterations (Fig. 15a), compared with the reconstruction obtained after 80 iterations with the CG in $L^{p(\cdot)}$, both with the fixed matrix p constructed through arc-sine (Fig. 15b) and with the matrix p updated every 40 iterations constructed through square root (Figure 15c). From the figures, it is clear that the application of the Conjugate Gradient algorithm in $L^{p(\cdot)}$ with the update of the matrix p provides the best outcome: the 6 rectangles are well recognisable with quite uniform and separate colours, the edges are visible, and the background is uniform too.

5 Conclusion

In this paper, we proposed to solve ill-posed inverse problems in the unusual setting of variable exponent Lebesgue spaces $L^{p(\cdot)}$ by one-step iterative algorithms. We also gave a comprehensive review of the generalization of one-step



(a) CG in L^2 (b) CG in $L^{p(\cdot)}$ (c) CG with update

Figure 15: CG algorithm: best reconstructions in L^2 (iteration 30), reconstruction at iteration 80 in $L^{p(\cdot)}$ with arc-sine interpolation, and in $L^{p(\cdot)}$ with updated square root interpolation, both with $1.3 \leq p \leq 1.4$.

iterative methods to Banach spaces for the resolution of ill-posed inverse problems. In this case, it is necessary to introduce duality mappings to embed primal and dual spaces and to define gradient descent iterations. To this aim, we first proved an explicit formula for the computation of the duality map in $L^{p(\cdot)}$ spaces. Moreover, we studied two possible approaches: performing the gradient step in the primal space (primal method) or in the dual space (dual method).

We showed the deep connection between these regularization iterative schemes and convex optimization by rewriting the Landweber iterations in Banach spaces as a proximal algorithm. In our insight, the primal method can be seen as a proximal iteration defined in terms of the usual Euclidean distance, whilst the dual method as a proximal iteration defined via the Bregman distance, generally more suitable to Banach spaces. This novel interpretation gives a useful heuristic and intuitive meaning of the iterative regularization in Banach spaces.

In the numerical section, we applied the Landweber dual method (Algorithm 2) and the Conjugate Gradient method (Algorithm 3) to solve signal and imaging deblurring problems in $L^{p(\cdot)}$. We showed as $L^{p(\cdot)}$ can be effectively used as solution spaces, since they allow to induce adaptive and space-variant regularization, without the use of specific penalty terms. We proposed a simple yet effective way to choose a variable exponent $p(\cdot)$ in order to enforce smoothness in the background of images as well as to preserve discontinuities at the edges and enforce sparsity. These first numerical results validated the choice of $L^{p(\cdot)}$ spaces for the resolution of deblurring imaging problem. Conjugate gradient method combined with an updating strategy for the variable exponent led to significantly improved images w.r.t. the classical L^2 reconstruction.

Further investigations about suitable stopping rules and choices of the exponent map $p(\cdot)$ will be object of future work, together with the application of other optimization algorithms in the same setting of $L^{p(\cdot)}$ spaces.

Acknowledgments. This work is partially supported by GNCS-INDAM, Italy.

- Conflict of interest: The authors declare that they have no conflict of interest.
- Availability of data and materials: Data will be made available on request.

References

[1] Neubauer, A.: Tikhonov-regularization of ill-posed linear operator equations on closed convex sets. J. Approx. Theory **53**(3), 304–320 (1988)

[2] Eicke, B.: Iteration methods for convexly constrained ill-posed problems in hilbert space. Numer Funct Anal Optim **13**(5-6), 413–429 (1992)

[3] Piana, M., Bertero, M.: Projected landweber method and preconditioning. Inverse Probl. **13**(2), 441–463 (1997)

[4] Natterer, F.: The mathematics of computerized tomography. John Wiley (1986)

[5] Bredies, K.: A forward–backward splitting algorithm for the minimization of non-smooth convex functionals in Banach space. Inverse Probl. **25**(1), 015005 (2008)

[6] Schuster, T., Kaltenbacher, B., Hofmann, B., Kazimierski, K.S.: Regularization methods in banach spaces. De Gruyter (2012)

[7] Schöpfer, F., Louis, A.K., Schuster, T.: Nonlinear iterative methods for linear ill-posed problems in Banach spaces. Inverse Probl. **22**(1), 311–329 (2006)

[8] Parikh, N., Boyd, S.: Proximal algorithms. Found. Trends Optim. **1**(3), 127–239 (2014)

[9] Brianzi, P., Di Benedetto, F., Estatico, C.: Preconditioned iterative regularization in banach spaces. Comput Optim Appl **54**(2), 263–282 (2013)

[10] Cioranescu, I.: Geometry of banach spaces, duality mappings and nonlinear problems. Springer (1990)

[11] Butnariu, D., Resmerita, E.: Bregman distances, totally convex functions, and a method for solving operator equations in banach spaces. Abstr. Appl. Anal. (2006)

[12] Bregman, L.M.: The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. USSR Comput. Math. Math. Phys. **7**(3), 200–217 (1967)

[13] Diening, L., Harjulehto, P., Hästö, P., Ruzicka, M.: Lebesgue and Sobolev Spaces with Variable Exponents. Lecture Notes in Math. Springer,

Germany (2011)

[14] Cruz-Uribe, D.V., Fiorenza, A.: Variable lebesgue spaces. Springer Birkhäuser Basel (2013)

[15] Dinca, G., Matei, P.: Geometry of Sobolve spaces with variable exponent and a generalization of the p-Laplacian. Analysis and Applications **07**(04), 373–390 (2009)

[16] Ciarlet, P.G., Dinca, G., Matei, P.: Operator equations and duality mappings in sobolev spaces with variable exponents. Chinese Annals of Mathematics, Series B **34**(5), 639–666 (2013)

[17] Dinca, G., Matei, P.: Geometry of Sobolev spaces with variable exponent: smoothness and uniform convexity. C. R. Math. **347**(15), 885–889 (2009)

[18] Matei, P.: A nonlinear eigenvalue problem for the generalized Laplacian on Sobolev spaces with variable exponent. Rom. J. Math. Comput. Sci. **2**(2), 70–82 (2012)

[19] Matei, P.: On the Fréchet differentiability of Luxemburg norm in the sequence spaces $\ell^{p(n)}$ with variable exponents. Rom. J. Math. Comput. Sci. **4**(2), 167–179 (2014)

[20] Estatico, C., Gratton, S., Lenti, F., Titley-Peloquin, D.: A conjugate gradient like method for p-norm minimization in functional spaces. Numer. Math. **137**, 895–922 (2017)

[21] Lazzaretti, M., Calatroni, L., Estatico, C.: Modular-proximal gradient algorithms in variable exponent lebesgue spaces. SIAM J. Sci. Comp., (in press) (2022)

[22] Estatico, C., Fedeli, A., Pastorino, M., Randazzo, A.: Quantitative Microwave Imaging Method in Lebesgue Spaces With Nonconstant Exponents. IEEE Trans. Antennas Propag. **66**(12), 7282–7294 (2018)