HALF-LINEAR REGULARIZATION FOR NONCONVEX IMAGE RESTORATION MODELS

Bartomeu Coll, Joan Duran and Catalina Sbert
Department of Mathematics and Computer Science
University of Balearic Islands
Ctra. Valldemossa Km. 7.5, 07122 Palma de Mallorca, Spain

(Communicated by Jean-Michel Morel)

Abstract. Image restoration is the problem of recovering an original image from an observation of it in order to extract the most meaningful information. In this paper, we study this problem from a variational point of view through the minimization of energies composed of a quadratic data-fidelity term and a nonsmooth nonconvex regularization term. In the discrete setting, existence of minimizer is proved for arbitrary linear operators. For this kind of problems, fully segmented solutions can be found by minimizing objective nonconvex functionals. We propose a dual formulation of the model by introducing an auxiliary variable with a double function. On one hand, it marks the edges and it ensures their preservation from smoothing. On the other hand, it makes the criterion half-linear in the sense that the dual energy depends linearly on the gradient of the image to be recovered. This leads to design an efficient optimization algorithm with wide applicability to several image restoration tasks such as denoising and deconvolution. Finally, we present experimental results and we compare them with TV-based image restoration algorithms.

1. Introduction. An important problem in image processing is the reconstruction of an original image from a degraded observation of it. This can be done by removing some kind of distortion or perturbation in order to recover the underlying true image. Such a process, called image restoration, has been studied at length from different points of view [11, 27, 28, 43, 44]. Most of the problems related to the degradation are caused by the noise (restoration by denoising) and by the blur (restoration by deblurring).

A simple model for image restoration is based on the following steps: use a priori information of the degradation, propose a model for the degradation process, and formulate and evaluate objective criteria of goodness. In the variational framework, this leads to the minimization of a functional involving a first regularization term related to the gradient of the image to be recovered $u$ and a second one which measures the fidelity to the data $f$. The regularization is written in terms of a

2010 Mathematics Subject Classification. 68U10, 94A08, 65F22, 90C26, 90C46, 65K10, 65K05.
Key words and phrases. Restoration, denoising, deconvolution, segmentation, inverse problems, edge-preserving regularization, nonconvex optimization, weighted total variation, duality.

These authors were supported by the Ministerio de Ciencia e Innovación under grant TIN2011-27539 and by the Conselleria d’Educació, Cultura i Universitats of the Govern de les Illes Balears under grant “Grup Competitiu 2011”. J. Duran had a fellowship of the Conselleria d’Educació, Cultura i Universitats of the Govern de les Illes Balears for the realization of his Ph.D. thesis, which was selected under an operational program co-financed by the European Social Fund.
so-called potential function $\phi$ and the problem consists in the minimization of an energy functional such as

$$J(u) = \int_{\Omega} \phi(|\nabla u|) + \frac{\lambda}{2} \|Au - f\|^2_2,$$

where $u$ and $f$ belong to a suitable space of functions, $A$ is a linear operator modeling the degradation of $u$, and $\lambda$ is a trade-off parameter between the two terms.

The most common model is the widely mentioned total variation smoothing functional of Rudin, Osher and Fatemi [43], where $\phi(t) = t$. Although convex functions are often used for the regularization term, Aubert and Kornprobst [3] pointed out that nonconvexity gives better results in numerical tests. Since the pioneering work of S. Geman and D. Geman [21], different nonconvex potential functions have been considered either in a statistical or variational framework [3, 6, 19, 32, 37, 39]. For this reason, many image restoration problems are often converted into nonconvex optimization problems. However, their practical interest is limited by theoretical deficiencies about the existence of solutions and the convergence of minimizing sequences. Furthermore, nonconvexity requires complex computational stages involving backward-parabolic Euler-Lagrange equations that lead to ill-posed problems. For instance, it is well-known that Perona-Malik PDE [39] is formally ill-posed.

An important effort has been done in order to avoid the numerical intricacies arising from nonconvex optimization. Robini et al. [41, 42] combined simulated annealing with deterministic continuation for the reconstruction of piecewise constant images. On the other hand, Nikolova et al. [36, 37] proposed a nonsmooth continuation GNC-like approach in the context of linear inverse problems. The segmentation and the restoration problems were solved jointly by minimizing a nonconvex and nonsmooth energy. However, there is no guarantee for the convergence of the continuation method. Furthermore, an important drawback is the use of a large number of additional variables that increase the computational complexity of the algorithm. Other interesting approaches for dealing with nonconvex optimization can be found in [9], where the authors proposed a gradient sampling algorithm, and in [15], where a smoothing nonlinear conjugate gradient method was introduced. Moreover, a Newton-type solution algorithm was proposed in [25], while in [1, 8], the authors introduced a proximal alternating linearized minimization algorithm by completely depart from the convex setting.

In this work, we are interested in expanding the scope of edge-preserving convex approximations to nonsmooth nonconvex regularizers. By imposing some conditions on the potential function to preserve the main information of the image, that is, forward diffusion in quasi-constant areas and enhancing around edges, we can interpret the dynamic on the associated Euler-Lagrange equation. In this sense, our proposed model acts as the weighted total variation, where the weight depends on the norm of the gradient. We also introduce an auxiliary variable $\omega$ that transforms the primal problem into a dual one by means of an augmented energy $J^*$ which is linear with respect to the norm of the gradient of $u$. For this reason, we refer to our approach half-linear regularization. More precisely, we shall have that

$$\min_u J(u) = \min_u \left\{ \min_{\omega} J^*(u, \omega) \right\},$$

where the solution of $\min_{\omega} J^*(u, \omega)$ will be explicitly given by $\omega = \phi'(|\nabla u|)$. We shall finally propose a dual algorithm with $\omega$ acting as edge detector.
Our proposal has a direct relationship with two different approaches to solve the optimization problem in the nonconvex case. On one hand, Menard et al. [2, 24] proposed a weighted total variation model by using the ideas of anisotropic diffusion and the total variation. On the other hand, Charbonnier et al. [14] provided a deterministic algorithm to find the solution of the optimization problem by minimizing a dual energy, which involves an isotropic weighted Laplacian. This is the so-called half-quadratic criterion and it is based on the pioneering papers by German et al. [19, 20]. In our case, the minimization of the dual energy involves an anisotropic term given by the weighted total variation. From a numerical point of view, the weighted TV is not straightforward to minimize since it is not differentiable at zero. However, TV-regularizers can better preserve sharp edges and object boundaries, which are usually the most important features to recover, than Laplacian-like regularizers.

Now, we briefly streamline the novelty of our approach and our contributions. This paper has four major contributions, which are subsequently outlined:

- We prove the existence of minimizer for general nonconvex discrete energies by only imposing that the linear operator modeling the degradation of the image does not annihilate constant functions as the gradient does. Because the coercivity is not guaranteed for some of the functionals under consideration, the novelty lies on the original mathematical tools used in the proof.
- The main contribution of the paper is the new-proposed half-linear regularization for nonsmooth nonconvex energies satisfying a suitable set of conditions for general edge preservation.
- Nikolova et al. [36] proved that the minimization of nonsmooth nonconvex energies provides piecewise constant images with neat edges. We relax the conditions imposed there and, thus, we extend the theorem to more general functionals.
- The last contribution is an optimization dual algorithm that exploits the properties of the half-linear regularization. We establish convergence results for the new-proposed scheme. Numerical experiments confirm the role of the dual variable as edge detector and exhibit the efficiency of our approach to restore piecewise constant images from noisy and blurred data.

The rest of the paper is organized as follows. Section 2 displays the general minimization problem we focus on and it establishes the existence of minimizer for general nonconvex energies. We discuss in Section 3 the set of conditions imposed on the potential function under which the half-linear regularization applies. This section contains theoretical results that gives meaning to our approach and allow us to introduce a robust optimization algorithm in Section 4. Section 5 exhibits the applicability of the algorithm for image denoising and deconvolution. We compare our results with TV-based restoration algorithms, the half-quadratic criterion among themselves. We conclude with some final remarks.

2. The general minimization problem. The most commonly studied degradation model in image restoration relates the observed data \( f \) to the underlying true image \( u \) by means of

\[
 f = Au + \eta,
\]

where \( \eta \) accounts for additive white Gaussian noise and \( A \) is a linear operator modeling the degradation of \( u \) caused, for instance, by optical blurring. Image
denoising models may be seen as a special case by taking $A$ as the identity operator. Unfortunately, getting $u$ from (1) is in some cases an ill-posed inverse problem in the sense of Hadamard [26]. Indeed, $A^*A$ is not always one-to-one and, even in this case, its eigenvalues may be small causing numerical instability. Prior information on the underlying image can then help the restoration of $u$. A flexible way to do that is regularization. In this case, the solution is defined as the minimizer of an energy functional.

The general problem we consider in this paper lies on the minimization of an objective functional composed of a nonsmooth nonconvex regularization term and a quadratic data-fidelity term:

$$J(u) = \int_{\Omega} \phi(|\nabla u|) + \frac{\lambda}{2} \|Au - f\|^2_X.$$  

In this setting, $\phi$ is a nonsmooth nonconvex potential function and $\lambda$ is a positive weighting constant that controls the trade-off between a good fit to $f$ and the amount of oscillation of $u$. An important requirement in image restoration is that $\phi$ allows the recovery of large differences at the locations of edges and smoothes differences elsewhere.

Let us now introduce the discrete minimization problem we are interested in. To simplify, we assume that an image is denoted by a two-dimensional matrix of size $N \times N$. We denote by $X$ the Euclidean space $\mathbb{R}^{N \times N}$ endowed with the usual scalar product $\langle u, v \rangle_X = \sum_{i,j} u_{i,j} v_{i,j}$. If $u \in X$, the discrete gradient is a vector in $Y = X \times X$ given at each pixel by $(\nabla u)_{i,j} = (\nabla u_{1})_{i,j}, (\nabla u_{2})_{i,j}$, with

$$\nabla u_{1}^{1}_{i,j} = \begin{cases} 
\frac{u_{i+1,j} - u_{i,j}}{\alpha} & \text{if } i < N, \\
0 & \text{otherwise},
\end{cases}$$

and

$$\nabla u_{2}^{2}_{i,j} = \begin{cases} 
\frac{u_{i,j+1} - u_{i,j}}{\alpha} & \text{if } j < N, \\
0 & \text{otherwise},
\end{cases}$$

for each $i, j \in \{1, \ldots, N\}$. Here $\alpha > 0$ is a real scaling parameter that tunes the value of the gradient above which a discontinuity is detected. The standard Euclidean norm of $\mathbb{R}^2$ will be simply denoted by $|\cdot|$ and, thus, $|\nabla u|_Y^2 = \sum_{i,j} |(\nabla u)_{i,j}|^2$. With this notation, the general problem we consider is

$$\min_{u \in X} J(u),$$

where $J$ is the discrete functional given by

$$J(u) = \sum_{1 \leq i,j \leq N} \phi(|(\nabla u)_{i,j}|) + \frac{\lambda}{2} \|Au - f\|^2_X.$$  

2.1. Existence of minimizer. We start our investigations of the problem (3)-(4) by establishing the existence of solution. A basic requirement is

(A0) $\ker(A) \cap \ker(\nabla) = \{0\},$

which is equivalent to suppose that $A$ does not annihilate constant functions as the gradient operator does. Note that if $\phi$ is a bounded nonconvex potential function, then (4) is not coercive unless we impose $A$ to be invertible. However, this hypothesis does not apply when $A$ is, for instance, a convolution operator with a point spread kernel. Therefore, the coercivity of the functional cannot be used. In the following result, we provide a novel and original proof instead.
Theorem 2.1. Let $\phi : [0, +\infty) \to [0, +\infty)$ be a continuous and increasing function satisfying $\phi(0) = 0$. If (A0) holds, then there exists $\hat{u} \in X$ such that

$$J(\hat{u}) = \min_{u \in X} J(u).$$

Proof. Since $\phi(t) \geq 0$ for $t \in (0, +\infty)$ and $\phi(0) = 0$, then the functional $J(u)$ is bounded below by zero for every $u \in X$ and, thus, its infimum exists. Let us denote it by $m = \inf_{u \in X} J(u)$.

Define $G(u) = \frac{1}{2} \|Au - f\|_X^2$ and consider the optimization problem $\inf_{u \in X} G(u)$. It is easy to prove that $G$ has at least a minimizer. If $A$ is of full rank, then $J$ is coercive and continuous so that it reaches its minimum and the proof would be finished. Otherwise, the solution set is given by the affine variety $u_0 + \ker(A)$, where $u_0$ is a particular solution of the associated equation $A^*(Au - f) = 0$. Let $\mu_0 = \min_{u \in X} G(u)$ be. For a fixed $\mu \geq \mu_0$, consider the subset

$$M_{\mu} = \{u \in X : G(u) \leq \mu\}$$

and define the functional

$$J_1(u) = \sum_{1 \leq i,j \leq N} \phi(|(\nabla u)_{i,j}|), \quad \forall u \in M_{\mu}.$$ (5)

Let us prove that $J_1$ reaches its minimum on $M_{\mu}$. Fix $u_0 \in M_{\mu_0}$ and define $\tau = \max_{1 \leq i,j \leq N} |(\nabla u_0)_{i,j}|$. We can assume $\tau > 0$, otherwise $u_0$ would be a minimizer of (5) so we would done. Since $M_{\mu}$ is unbounded, we can move sufficiently far from $u_0$ in $M_{\mu}$ so that any $v = u - u_0$, with $u \in M_{\mu}$, tends to be parallel to $\ker(A)$. Therefore, (A0) implies that there exists $C > 0$ such that

$$\|v\|_X > C \Rightarrow |(\nabla v)_{i,j}| \geq 2\tau, \quad \forall (i,j) \in \{(i,j) : |(\nabla v)_{i,j}| \neq 0\}. \quad (6)$$

On the other hand, take $R > C$ and define the compact set $K = M_{\mu} \cap B_R(u_0)$, where $B_R(u_0)$ denotes the closed ball of radius $R$ centered at $u_0$. Because $J_1$ is continuous, there exists $u_\mu \in K$ such that

$$J_1(u_\mu) = \inf_{u \in K} J_1(u) =: k_\mu.$$ (7)

It remains to prove that $u_{\mu}$ is in fact a minimizer of (5) on $M_{\mu}$. For that, fix $u \in M_{\mu} \setminus K$ and define $v = u - u_\mu$. Therefore, one has that $\|v\|_X \geq R > C$ and (6) yields

$$|(\nabla u)_{i,j}| = |(\nabla (u_\mu + v))_{i,j}| \geq \tau \geq |(\nabla u_\mu)_{i,j}|, \quad \forall i,j \in \{1, \ldots, N\}.$$ (8)

The previous inequalities and the increase of $\phi$ imply $J_1(u) \geq J_1(u_\mu) \geq J_1(u_\mu)$, so $u_\mu$ is a minimizer of (5) on $M_{\mu}$.

Finally, consider the function $g : [\mu_0, +\infty) \to \mathbb{R}$ defined as $g(\mu) = \mu + k_\mu$. Since $g$ is continuous and coercive, there exists $\hat{\mu} \in [\mu_0, +\infty)$ such that

$$g(\hat{\mu}) = \inf_{\mu \in [\mu_0, +\infty)} g(\mu).$$ (9)

In what follows we shall prove that $g(\hat{\mu}) = m$. By definition of infimum, for any $\epsilon > 0$ there exists $v \in X$ such that $\|v\| < m + \epsilon$. Define $\mu^* = \frac{\epsilon}{2}\|Av - f\|_X^2$, whence one clearly has that $v \in M_{\mu^*}$, and let $u_{\mu^*}$ be a minimizer of $J_1$ on $M_{\mu^*}$. One deduces that

$$g(\mu^*) = \mu^* + \sum_{1 \leq i,j \leq N} \phi(|(\nabla u_{\mu^*})_{i,j}|) \leq \frac{\lambda}{2}\|Av - f\|_X^2 + \sum_{1 \leq i,j \leq N} \phi(|(\nabla v)_{i,j}|)$$

$$= J(v) < m + \epsilon.$$ (10)
Since \( g(\mu) \leq g(\mu^*) < m + \epsilon \) for any \( \epsilon > 0 \), then \( g(\mu) \leq m \). On the other hand, we have that
\[
m \leq J(u_{\hat{\mu}}) = \frac{1}{2}\|Au_{\hat{\mu}} - f\|_2^2 + J_1(u_{\hat{\mu}}) \leq \hat{\mu} + k_{\hat{\mu}} = g(\hat{\mu}).
\]
Then, \( g(\hat{\mu}) = m \) and \( u_{\hat{\mu}} \in M_{\hat{\mu}} \) is a minimizer of \( J \), which ends the proof.

Beyond guaranteeing the existence of solution for the minimization problem under consideration, the main contributions of Theorem 2.1 are the mathematical tools used in its proof.

3. **Half-linear regularization.** It is well known that minimization of nonconvex energies involves numerical difficulties that restrict the methods that can be used. In order to overcome these drawbacks, we reformulate the objective functional \( J \) as an augmented energy \( J^* \) by introducing a closed-form dual variable.

3.1. **Assumptions on the potential function.** In order to correctly introduce the half-linear regularization, we need to assume some specific conditions on the potential function. A qualitative reasoning on the Euler-Lagrange equation is developed here to justify the set of imposed conditions. The requirements are imposed to have forward diffusion (smoothing) in quasi-constant areas and backward diffusion (enhancement) around edges.

If we focus on the continuous case, suppose that some particular function \( u \) happens to be a minimizer of (2). Then it formally satisfies the Euler-Lagrange equation
\[
\begin{cases}
-\text{div} \left( \phi'(|\nabla u|) \frac{\nabla u}{|\nabla u|} \right) + \lambda A^* (Au - f) = 0, & \text{if } \nabla u \neq 0, \\
|\nabla u| = 0, & \text{otherwise}.
\end{cases}
\]
Note that the divergence term is clearly related to the weighted total variation regularizer and, thus, \( \phi'(|\nabla u|) \) acts as the “weight”. In the case that a pixel belongs to a quasi-constant area, we can assume that all neighboring gradients are close to zero. Consequently, we impose that
\[
\lim_{t \downarrow 0^+} \phi'(t) = M, \quad 0 < M < +\infty.
\]
Then, \( \phi'(|\nabla u|) \) is close to \( M \) and (7) behaves as the usual total variation near points where the norm of the gradient is small. On the contrary, consider now a pixel belonging to an edge. A discontinuity appears in its neighborhood and, thus, the gradient in the orthogonal direction of the edge is large. In this case we impose
\[
\lim_{t \to +\infty} \phi'(t) = 0
\]
and, thus, \( \phi'(|\nabla u|) \) is close to zero in the transversal direction of the edge so that the singularity is preserved. Furthermore, since we want to avoid that small variations in the gradient produce large changes in the value of the weight, we impose \( \phi' \) to be continuous. It also seems natural that there should be a one-to-one correspondence between values of the gradient and values of the weight, so that \( \phi' \) must be strictly monotonous.

We can summarize all conditions imposed on the potential function as follows:
(A1) \( \phi: [0, +\infty) \to [0, +\infty) \) strictly increases on \( (0, +\infty) \), with \( \phi(0) = 0 \).
(A2) \( \phi \) is twice continuously differentiable on \( (0, +\infty) \).
(A3) \( \phi' \) strictly decreases on \( (0, +\infty) \).
(A4) \( \lim_{t \to +\infty} \phi'(t) = 0 \).
(A5) \( \lim_{t \downarrow 0^+} \phi'(t) = M \), with \( 0 < M < +\infty \).

Requirements involved in (A1)-(A2) are typical in image restoration [3]. In general, (A3)-(A5) are conceived in such a way that small variations are smoothed, assuming they are due to noise, and high variations are preserved, supposing they are generated by edges. But noise can also produce high variations while edges with low contrast can generate low gradient modulus. This may be seen as a limitation of the proposed model, although it commonly appears in almost all edge-preserving regularizers such as the half-quadratic criterion [14, 19, 20].

**Remark 1.** Since \( \phi \) is given on \([0, +\infty)\), we can define the right derivative of \( \phi \) at zero by \( \phi'(0) := M \) due to (A2) and (A5). Therefore, \( \phi \) is continuously differentiable on \([0, +\infty)\). However, under the above conditions, the even extension of \( \phi \) on \( \mathbb{R} \) would be non-differentiable at zero.

Let us now study the above requirements from another point of view. Using the tangent and normal directions to the level lines along which the image intensity is constant, the Euler-Lagrange equation can be decomposed in terms of the directional derivatives. Hence, (7) can be rewritten as

\[
\begin{cases}
A^*Au - \frac{1}{\lambda} \left( \frac{\phi'(|\nabla u|)}{|\nabla u|} u_{TT} + \phi''(|\nabla u|)u_{NN} \right) = A^*f, & \text{if } \nabla u \neq 0, \\
|\nabla u| = 0, & \text{otherwise},
\end{cases}
\]

where \( u_{TT} \) and \( u_{NN} \) denote the second derivatives in the tangent and the normal direction, respectively. At first glance, (A1)-(A5) imply that we encourage forward diffusion along isophotes, that is, lines of constant grey level value, combined with backward diffusion along the gradient direction.

Perona and Malik [39] showed that edge enhancement and reconstruction of blurred images can be performed by imposing the normal diffusion coefficient to be negative. In our case, (A3) implies that \( \phi''(|\nabla u|) < 0 \) and, thus, an inverse diffusion might be observed in the normal direction when using the class of potential functions proposed in this paper. Therefore, the regularization term in (4) always sharpens, more or less, the edges of the image.

At locations where the gradients are low, (A5) ensures that the tangent diffusion coefficient is huge. Consequently, the diffusion takes place mainly following the level lines along which the image intensity is constant, while these curves are also enhanced. In a neighborhood of an edge, where the image has a strong gradient, condition (A4) implies that

\[
\lim_{t \to +\infty} \frac{\phi'(t)}{t} = 0 \quad \text{and} \quad \lim_{t \to +\infty} \phi''(t) = 0.
\]

Consequently, there is no diffusion neither in the tangent nor in the normal direction. Note that \( \frac{\phi'(t)}{t} \) tends to \( 0^+ \), whereas \( \phi''(t) \) tends to \( 0^- \) so that the edge is not only preserved but also slightly enhanced. After all, the diffusion process leads to constant regions surrounded by sharp edges.

3.2. **The main dual theorem.** The key of the half-linear regularization is to introduce a dual variable \( \omega \), with closed expression, which detects edges and image features. Our approach is based on the theorem below.

**Theorem 3.1.** Let \( \phi \) satisfy requirements (A1)-(A5) and let \( M \) be the real value given in (A5).
Remark 2. The proof of the theorem reveals the explicit expression of $\psi$, although the second statement makes clear that we do not really need $\psi$ to solve (9).
By assuming that (A1)-(A5) hold, the functional (4) can be written as
\[
J(u) = \min_{\omega \in (0, M]^N} J^*(u, \omega),
\]
where \(w = (w_{i,j})_{1 \leq i,j \leq N}\) and the dual energy is explicitly given by
\[
J^* (u, \omega) = \sum_{1 \leq i,j \leq N} (\omega_{i,j} |(\nabla u)_{i,j}| + \psi(\omega_{i,j})) + \frac{\lambda}{2} \|Au - f\|_X^2.
\]
Note that \(J^*\) is linear with respect to the norm of the gradient of the main variable and it is strictly convex with respect to the auxiliary variable when the main one is fixed\(^\text{1}\). In fact, \(J^*\) involves a weighted total variation term on \(u\) with \(\omega\) acting as the weight. Furthermore, from Theorem 3.1, the value of the dual variable that minimizes (15), when \(u\) is fixed, is explicitly given by
\[
\omega_{i,j} = \phi'(|(\nabla u)_{i,j}|), \quad \forall i,j \in \{1, \ldots, N\}.
\]
It is important to emphasize that the above closed formula and the strictly decreasing property of \(\phi'\) guarantee the role of \(\omega\) as an indicator of edges:
- If \(\omega_{i,j}\) is close to zero, then the gradient at pixel \((i,j)\) is large and, thus, it belongs to an edge.
- If \(\omega_{i,j}\) is close to \(M\), then the gradient at pixel \((i,j)\) is small and, thus, it belongs to a quasi-constant region.

Table 1 lists some examples of nonconvex potential functions satisfying (A1)-(A5), and Figure 1 displays their graphs as well as the graphs of the corresponding weight functions.

| \(\phi(t)\) | \(\tilde{\omega}\) | \(M\) | \(\beta\) | \(\psi(\omega)\) |
|-----------|-------------|-----|-------|-------------|
| \(\phi_1(t) = \frac{t}{\sqrt{1 + t^2}}\) | \(\frac{1}{(1 + t^2)^{\frac{1}{2}}}\) | 1 | 1 | \((1 - \omega^2)^{\frac{1}{2}}\) |
| \(\phi_2(t) = \frac{t}{1 + t}\) | \(\frac{1}{(1 + t)^2}\) | 1 | 1 | \((1 - \sqrt{\omega})^2\) |
| \(\phi_3(t) = \ln(1 + t)\) | \(\frac{1}{1 + t}\) | 1 | +\(\infty\) | \(\omega - \ln \omega - 1\) |
| \(\phi_4(t) = (1 + t)^\gamma - 1, 0 < \gamma < 1\) | \(\frac{\gamma}{(1 + t)^{1-\gamma}}\) | \(\gamma\) | +\(\infty\) | \(\omega - 1 + (1 - \gamma) \left(\frac{\omega}{\gamma}\right)^{\frac{1}{\gamma}}\) |

Table 1. Examples of potential functions satisfying (A1)-(A5), with the associated weight functions \(\tilde{\omega}\) and the strictly convex functions \(\psi: (0, M) \to (0, \beta)\).

Theorem 3.1 is based on similar results related to the half-quadratic regularization \([14, 19, 20]\). However, it is important to distinguish between the scope of both approaches. Functions on which half-quadratic applies do not meet all assumptions (A1)-(A5). Moreover, our potential functions and those given in \([14, 19, 20]\) have different behaviors near zero, although both are edge-preserving in the general

\(^{1}\text{Note that } J^* \text{ is strictly convex with respect to } u \text{ whenever } A \text{ is invertible, which is not in general the case under consideration. Accordingly, the convex optimization problem } \min_{u \in X} J^*(u, \omega) \text{ may have several solutions. Even though } J^* \text{ is convex in each variable when the other is fixed, this does not guarantee the overall convexity.}
sense. Another important difference relies on their associated Euler-Lagrange equations. The proposed model approximates the potential function by $t \to \omega |t|$, which involves weighted total variation where the weight function is defined as $\phi'(t)$. On the contrary, the half-quadratic criterion uses the transformation $t \to \omega t^2$, which leads to weighted Laplacian with the weight being $\phi''(t)/t$, closely related to quadratic functionals.

After all, Theorem 3.1 is a significant contribution to the minimization of energy functionals in the image restoration field because it allows to deal with potential functions that are not treatable by the half-quadratic approach. More concretely, we consider nonsmooth nonconvex functionals that are suitable for the restoration of piecewise constant images with neat edges or for the segmentation of noisy and blurred data. This shall be developed in the following sections.

Finally, it is obvious that minimizing quadratic functionals is simpler than minimizing the weighted total variation. Moreover, a well-known drawback in the latter case is the staircasing effect. Nevertheless, Rudin, Osher and Fatemi [43] observed that quadratic regularization does not allow recovering sharp discontinuities as the total variation does. Hence, we meet a meaningful benefit of the proposed model.

### 3.3. Properties of the minimizers

The interest in the class of potential functions described in the previous subsections is motivated by theoretical results, which assert that nonconvex nonsmooth regularization provides better possibilities for restoring images composed of constant regions surrounded by neat edges [34, 36, 37].

We prove a similar result for the optimization problem (3)-(4) by assuming the imposed hypothesis on $\phi$. The main interest of this result is that fully “segmented” solutions of arbitrary linear inverse problems can be found by minimizing an objective functional where the regularization term is constructed using a nonconvex potential function. To prove that, we need to assume the following additional condition:

\[(A6) \quad \frac{\phi''(t)}{t^2} \ \text{strictly increases on} \ (0, +\infty) \ \text{and} \ \lim_{t \to 0^+} \frac{\phi''(t)}{t^2} = -\infty.\]

We can state now the following result, which proof is outlined in Appendix A.
Theorem 3.2. Suppose that assumptions (A0)-(A6) hold. Then, any local minimizer \( \hat{u} \in X \) of the functional \( J \) given in (4) satisfies

\[
(16) \quad \text{either } \quad |(\nabla \hat{u})_{i,j}| = 0 \quad \text{or} \quad |(\nabla \hat{u})_{i,j}| \geq \theta, \quad \forall 1 \leq i, j \leq N,
\]

where \( 0 < \theta < +\infty \) is the unique real value that solves \( \phi''(\theta) = -\lambda \mu \theta^2 \| f \|^2_X \), with \( \mu \) being a constant such that \( \mu \geq 1 \).

Although \( \theta \) is not given explicitly, this theorem provides interesting theoretical justification. We can consider the restoration of piecewise constant images where the number of the regions and their values are not fixed in advance from noisy and blurred data obtained at the output of a linear operator.

Theorem 3.2 is based on a similar result in [36] where Nikolova et al. showed the same property but under different conditions. To compare both results, any twice continuously differentiable function satisfying Nikolova’s requirements also meets (A1)-(A6) and, thus, our theorem follows. However, \( \phi_1 \) in Table 1 is a good counterexample to show that the other way around is not always satisfied.

4. The proposed dual algorithm. For solving the minimization problem (3)-(4) we propose Algorithm 1, which exploits the properties of the half-linear regularization by alternating minimizations over \( u \) and \( \omega \).

Algorithm 1 Proposed dual algorithm

\[
\begin{aligned}
& u^0 = f \\
& \text{repeat} \\
& \quad \omega^{n+1} = \arg \min_{\omega} J^* (u^n, \omega) = \phi' (|\nabla u^n|) \\
& \quad u^{n+1} = \arg \min_u J^* (u, \omega^{n+1}) \\
& \text{until convergence}
\end{aligned}
\]

On one hand, when \( u^n \) is known and fixed, Theorem 3.1 states that the minimizer of \( J^* (u^n, \omega) \) with respect to the dual variable \( \omega \in (0, M]^{N \times N} \) is explicitly given by

\[
(17) \quad \omega_{i,j}^{n+1} = \phi' (| (\nabla u^n)_{i,j} |), \quad \forall i,j \in \{1, \ldots, N\}.
\]

On the other hand, once \( \omega^{n+1} \) is known and fixed, the minimization problem

\[
(18) \quad u^{n+1} = \arg \min_{u \in X} J^* (u, \omega^{n+1})
\]

is equivalent to the minimization of the objective functional \( J \) and it underlies a total variation regularizer weighted by \( \omega^{n+1} \). The term of the weighted total variation of earlier expression is

\[
\text{WTV}(u) = \sum_{1 \leq i,j \leq N} w_{i,j} |(\nabla u)_{i,j}|,
\]

where \( \omega \) plays the role of weight function. In fact, this is a discretization of the standard weighted total variation [10] defined in the continuous setting for a function \( u \in L^1(\Omega) \) by

\[
\text{WTV}(u) = \sup \left\{ \int_\Omega u(x) \text{div} (\omega(x)\xi(x)) : \xi \in C^1(\Omega, \mathbb{R}^2), \| \xi \|_\infty \leq 1 \right\}.
\]

In fact, Theorem 3.2 can be generalized for potential functions \( \phi \) being \( C^2 \) on \((0, +\infty) \setminus H\), where \( H \) is a finite set of points (possibly empty) with \( 0 \notin H \) and such that if \( t \in H \) then \( \phi''(t^-) > \phi''(t^+) \) and \( -\infty < \lim_{r \to t^-} \phi''(r) < \lim_{r \to t^+} \phi''(r) \leq 0 \).
4.1. Convergence analysis. Algorithm 1 converts the original nonconvex minimization problem into a sequence of convex minimization problems. As usual, convergence is a key theoretical issue for iterative schemes. However, nonconvexity introduces theoretical deficiencies about the existence of a global minimizer and the convergence of minimizing sequences \( \{u_n\}_{n \geq 0} \). In our case, it can be shown that \( \{u_n\}_{n \geq 0} \) converges to a stationary point of (4) provided its stationary points are isolated. Otherwise, the accumulation points of \( \{u_n\}_{n \geq 0} \) can be guaranteed to get arbitrarily close to the set of all stationary points of \( J \).

For every \( u \in X \) let us define the mapping

\[
M(u) = \arg \min_{v \in X} J^*(u, \omega_u),
\]

where \( \omega_u = \partial'(|\nabla u|) \). Note that \( M \) is a point-to-set map that associates with every point \( u \in X \) a subset of \( X \) (see [29] for more details about this class of mappings). This is due to the fact that \( J^* \) is not strictly convex with respect to its first variable and, thus, it may have several minimizers. The procedure (17)-(18) is thus equivalent to the iterative application of (19) from an initial point \( u^0 \in X \),

\[
u^{n+1} \in M(u^n),
\]

where it is not relevant which of the points in the set \( M(u^n) \) is assigned to \( u^{n+1} \).

The fixed points of \( M \) satisfy \( u \in M(u) \). By the construction in (19)-(20), all fixed points are minimizers of \( J^* \) with respect to its first variable, but they are also stationary points of \( J \) as stated in the following proposition. Its proof is outlined in Appendix B.

**Proposition 1.** Let \( S \) be the set of all fixed points of the mapping, i.e., \( S = \{ u \in X : u \in M(u) \} \). Then \( u \in S \) if and only if \( u \) is a stationary point of the functional \( J \) defined in (4).

We can state now several results about the convergence of the proposed algorithm that strengthen theoretically it. Our analysis proceeds in much the same way as [17, 38]. We first show that the sequence \( \{J(u^n)\} \) is decreasing and convergent, and that the sequence \( \{u^n\} \) has accumulation points. Under additional conditions, we finally prove that \( \{u^n\} \) converges to a stationary point of (4). The following theorem contains the statements about the convergence of our algorithm. All proofs are contained in Appendix C.

**Theorem 4.1.** Let conditions (A0)-(A5) hold. From any initial point \( u^0 \in X \), let \( \{u^n\}_{n \geq 0} \) be any sequence generated by Algorithm 1 as described in (19)-(20). Then, the following statements hold:

i) The sequence \( \{J(u^n)\}_{n \geq 0} \), with \( J(u^n) = J^*(u^n, \omega^{n+1}) \), decreases and converges to \( J(\tilde{u}) \), where \( \tilde{u} \) is a stationary point of \( J \).

ii) The sequence \( \{u^n\}_{n \geq 0} \) has convergent subsequences and their limits are stationary points of \( J \).

iii) \( \|u^{n+1} - u^n\|_X \to 0 \) as \( n \to +\infty \).

iv) Either \( \{u^n\}_{n \geq 0} \) converges to a stationary point of \( J \), or the accumulation points of \( \{u^n\}_{n \geq 0} \) form a continuum\(^3\) in \( S \).

v) If all stationary points of \( J \) are isolated, then \( \{u^n\}_{n \geq 0} \) converges to a stationary point of \( J \).

vi) There exists a sequence \( \{\tilde{u}^n\}_{n \geq 0} \) of stationary points of \( J \) such that it satisfies \( \lim_{n \to \infty} \|u^n - \tilde{u}^n\|_X = 0 \).

\(^3\)A nonempty compact connected metric space is called a continuum.
vii) Let \( \hat{u} \) be any isolated stationary point of \( J \). If \( \hat{u} \) is a strong local minimizer of \( J \), then there exists an open neighbourhood \( B \) of \( \hat{u} \) such that \( \{ u^n \}_{n \geq 0} \) converges to \( \hat{u} \) for any \( u^0 \in B \).

**Remark 3.** From previous theorem, we can state the main conclusions about the convergence of the new-proposed dual algorithm:

- From (v), the algorithm converges to a stationary point of \( J \) if all stationary points are isolated.
- If it is not possible to ensure that all stationary points of \( J \) are isolated, then from (ii) and (iv) either \( \{ u^n \} \) converges or its accumulation points form a continuum in \( S \). The algorithm also behaves well in the latter situation. Indeed, by (i) and (vi), \( \{ J(u^n) \} \) converges and \( \{ u^n \} \) gets arbitrary close to \( S \).
- We cannot rigorously exclude the possibility that the sequence \( \{ u^n \} \) converges to a saddle point, if such exists. However, this is very unlikely since, by (vii), any isolated stationary point that is a minimum is locally attractor whereas maxima and saddle points are unstable. In fact, convergence to a maximum is only possible if initialized there due to (i).

Finally, let us emphasize that the above convergence analysis does not depend on the numerical scheme used to solve the convex minimization problem (18).

### 4.2. Algorithmic details.

There exist many numerical algorithms for solving (18) efficiently, for instance [4, 5, 16, 23, 33, 46]. We use the first-order primal-dual algorithm proposed by Chambolle and Pock in [13, 40], which can be applied to a wide class of convex optimization problems comprising ours. Throughout this subsection we shall make frequent use of the so-called proximity operator, which was introduced by Moreau in [30] as a generalization of the notion of a convex projection operator.

Let \( K : X \to Y \) be a continuous linear operator. Let \( F : Y \to [0, +\infty] \) and \( G : X \to [0, +\infty] \) be proper, convex and lower-semicontinuous functions. Consider the generic nonlinear primal problem

\[
\min_{u \in X} \left\{ F(Ku) + G(u) \right\}
\]

and its associated saddle-point problem

\[
\min_{u \in X} \max_{p \in Y} \left\{ \langle Ku, p \rangle_Y + G(u) - F^*(p) \right\}, \tag{21}
\]

where \( F^* \) is the convex conjugate of \( F \). Every solution \( (\hat{u}, \hat{p}) \in X \times Y \) of (21) satisfies \( K\hat{u} \in \partial F^*(\hat{p}) \) and \( -(K^*\hat{p}) \in \partial G(\hat{u}) \), so the iterates of the algorithm proposed by Chambolle and Pock [13] are

\[
\begin{align*}
p^{n+1} &= (I + \eta \partial F^*)^{-1}(p^n + \eta Kv^n), \\
u^{n+1} &= (I + \tau \partial G)^{-1}(u^n - \tau K^*p^{n+1}), \\
v^{n+1} &= u^{n+1} + \theta (u^{n+1} - u^n),
\end{align*}
\]

where \( K^* \) is the adjoint operator of \( K \). The algorithm thus consists of alternating a gradient ascend in the dual variable and a gradient descend in the primal variable. The step-size parameters \( \tau > 0 \) and \( \eta > 0 \) are chosen such that \( \tau \eta L^2 < 1 \), where \( L = \| K \| \). Finally, parameter \( \theta \) controls the over-relaxation in \( u \), \( \theta \in [0, 1] \).

Now, we adapt the algorithm of Chambolle and Pock in order to solve (18). We first emphasize that \( \psi(\omega) \) in (15) is not involved in (18) and, thus, our primal
problem is defined as
\[ \min_{u \in X} \left\{ \sum_{1 \leq i,j \leq N} \omega_{i,j} |(\nabla u)_{i,j}| + \frac{\lambda}{2} \|Au - f\|_X^2 \right\}. \]

Let \( K \) be given by \( Ku = \omega \nabla u \), so that its adjoint operator is defined as \( K^*(p) = -\text{div} (\omega p) \) since \( (Ku, p)_Y = (\nabla u, \omega p)_Y = \langle u, -\text{div} (\omega p) \rangle_X \) for all \( u \in X \) and \( p \in Y \).

It is easy to show that \( \|K\|_2 \leq 8 \).

For casting (18) in the form (21), it remains to identify \( F \) and \( G \), and to detail the proximity operators in (22). For a better understanding, we exhibit first the case \( A = I \).

**Identity operator.** If \( A = I \), define \( F(Ku) = \|Ku\|_1 \) and \( G(u) = \frac{\lambda}{2} \|u - f\|_X^2 \).

Therefore, the associated saddle-point problem is
\[ \min_{u \in X} \max_{p \in Y} \left\{ \langle u, -\text{div} (\omega p) \rangle_X + \frac{\lambda}{2} \|u - f\|_X^2 - \delta_P(p) \right\}, \]
where \( P = \{ p \in Y : |p_{i,j}| \leq 1, \forall i, j \in \{1, \ldots, N\} \} \) and \( \delta_P \) is the indicator function of \( P \) defined as
\[ \delta_P(p) = \begin{cases} 0 & \text{if } p \in P; \\ +\infty & \text{if } p \notin P. \end{cases} \]

The resolvent operator \((I + \eta \partial F^*)^{-1}\) reduces to point-wise Euclidean projectors onto \( L^2 \) balls:
\[ \hat{p}_{i,j} = \frac{p_{i,j}}{\max_{i,j} \{1, |p_{i,j}|\}}, \quad \forall i, j \in \{1, \ldots, N\}. \]

Furthermore, \((I + \tau \partial G)^{-1}\) can be computed point-wise using the Euler-Lagrange equation:
\[ \hat{u}_{i,j} = \frac{u_{i,j} + \tau \lambda f_{i,j}}{1 + \tau \lambda}, \quad \forall i, j \in \{1, \ldots, N\}. \]

The iterative scheme for \( A = I \) is outlined in Algorithm 2.

**Algorithm 2** Detailed algorithm for \( A = I \)
\[
\begin{align*}
&u^0 = f \\
&\text{repeat} \\
&\quad \omega^{n+1} = \phi'(|\nabla u^n|), \quad p^0 = 0, \quad u^0 = u^n, \quad \bar{v}^0 = v^0 \\
&\quad \text{repeat} \\
&\quad \quad p^{k+1} = \frac{p^k + \eta \omega^{n+1} \nabla v^k}{\max \{1, \|p^k + \eta \omega^{n+1} \nabla \theta^k\|\}} \\
&\quad \quad v^{k+1} = \frac{v^k + \tau \text{div} (\omega p^{k+1}) + \tau \lambda f}{1 + \tau \lambda} \\
&\quad \quad \bar{v}^{k+1} = (1 + \theta) v^{k+1} - \theta v^k \\
&\quad \text{until convergence} \\
&\quad u^{n+1} = \bar{v} \\
&\text{until convergence}
\end{align*}
\]

**General case.** For a general linear operator, a convenient reformulation can be obtained by dualizing the norm in the data-fidelity term. This introduces an additional dual variable \( q \in X \), yielding
\[ \min_{u \in X} \max_{p \in Y, q \in X} \left\{ \langle u, -\text{div} (\omega p) \rangle_X + \langle Au - f, q \rangle_X - \delta_P(p) - \frac{1}{2\lambda} \|q\|_X^2 \right\}. \]
We have $F^*(p,q) = \delta_P(p) + \frac{1}{2\lambda}||q||_X^2$ and $G(u) = 0$. The proximity operator for $F^*$ can be computed separately for each variable. For $p \in Y$ we get the same result as in (23) and for $q \in X$ we obtain

$$\hat{q}_{i,j} = q_{i,j} + \frac{\eta \lambda}{1 + \eta \lambda}, \quad \forall i,j \in \{1, \ldots, N\}.$$ 

Finally, since $G = 0$ and $u$ is unconstrained, the proximity operator for $G$ is the identity. The general iterative scheme is outlined in Algorithm 3.

**Algorithm 3** Detailed algorithm for a general linear operator $A$

\[\begin{align*}
&u^0 = f \\
&\text{repeat} \quad \omega^{n+1} = \phi'(|\nabla u^n|), \quad p^0 = 0, \quad q^0 = 0, \quad v^0 = u^n, \quad \bar{v}^0 = v^0 \\
&\quad \text{repeat} \quad p^{k+1} = \frac{p^k + \eta \omega^{n+1} \nabla \bar{v}^k}{\max \{1, |p^k + \eta \omega^{n+1} \nabla \bar{v}^k|\}} \\
&\quad \quad q^{k+1} = \frac{q^k + \eta (A \bar{v}^k - f)}{1 + \eta \lambda} \\
&\quad \quad v^{k+1} = v^k + \tau (\text{div} (\omega p^{k+1}) - A^* q^{k+1}) \\
&\quad \quad \bar{v}^{k+1} = (1 + \theta) v^{k+1} - \theta v^k \\
&\quad \text{until convergence} \quad v^{n+1} = v \\
&\text{until convergence}
\end{align*}\]

5. Experimental results. In this section, we present a set of experimental results illustrating the performance of the new-proposed half-linear regularization. The aim is to show that, whatever may be the perturbations to which an image is subjected, the proposed algorithm provides piecewise restored images with neat edges.

We left Algorithm 1 proceed until $\|u^{n+1} - u^n\|^2/N^2$ became smaller than $10^{-6}$. The initial weight was set to $\omega^0 = \phi'(|\nabla f|)$. The trade-off parameter $\lambda$ and the scaling parameter $\alpha$ were estimated empirically throughout the experimental section. Moreover, the tolerance for Chambolle-Pock’s scheme in Algorithms 2-3 was $10^{-4}$, and the parameters there were chosen so that $L = \sqrt{8}$, $\tau = \eta = 1/L$ and $\theta = 1$ in all cases. All standard deviations given in this section are relative to the intensity range $[0, 255]$.

5.1. The role of the dual variable as edge detector. Section 3 revealed that requirements (A3)-(A5) strengthen the role of the dual variable as indicator of edges. This fact was experimentally checked by applying the proposed dual algorithm for $A = I$ (Algorithm 2 was thus used for solving the associated convex optimization problem).

A first example is given in Figure 2. We display the weight associated to each potential function in Table 1 after applying our algorithm to a degraded image with noise of standard deviation $\sigma = 10$. For a better visualization, the dual variable is plotted in grayscale, so that black pixels correspond to negligible weights (enhancement) whereas white pixels stand for the largest ones (diffusion). Despite the noise, $\omega$ perfectly detects the edges and other image features.

Figure 3 shows the restoration of a non-constant image. As it has been theoretically justified, the solutions are piecewise constant images with neat edges.
The noisy image contains Gaussian noise of standard deviation \( \sigma = 10 \). The parameters were fixed to \( \lambda = 0.1 \) and \( \alpha = 30 \). As expected, the dual variable perfectly detects image features and it provides neat edges.

We also observe that the potential function used in the experimentation influences the result. Indeed, some image features are emphasized by some \( \phi \)'s whereas they are omitted by some others. Consequently, the weight assigned to each pixel in the image changes depending on the potential function being used. Note that, in general, image regions with slight intensity variations are more regularized with \( \phi_1 \) than with the other potentials. It can be thought that this fact is related to the property \( \lim_{t \to 0^+} \phi''_1(t) = 0 \) in (8) and, as a consequence, small gradients are not enhanced. Anyway, the behavior of the dual variable can be supervised by the scaling parameter \( \alpha \), which controls the gradient level that one wants to preserve. In this sense, Figure 4 displays the weights for several values of \( \alpha \). As the scaling parameter increases, more singularities are penalized.

5.2. Applications in image restoration using different potential functions.
Many nonconvex potential functions have been proposed in the literature as regularization terms. We are interested in those listed in Table 1 satisfying conditions (A1)-(A6). In the following, the optimal values for the trade-off parameter \( \lambda \) and the scaling parameter \( \alpha \) were chosen manually to maximize the so-called peak signal-to-noise ratio\(^4\) (PSNR) for each potential function.

**Image denoising.** Denoising is the problem of removing the inherent noise from an image. The standard noise model is an additive white Gaussian noise, which is supposed to be at each point a zero-mean i.i.d. Gaussian random variable. Therefore, the half-linear regularization applies to solve the minimization problem (3)-(4) for \( A = I \).

Figure 5 displays an example on piecewise constant data corrupted with noise of standard deviation \( \sigma = 25 \). Note that all potential functions behave similarly except \( \phi_4 \), which performs worse than the others. The numerical results are quite impressive, since the RMSE values have been reduced by one third whereas the PSNR values have been considerably improved. Furthermore, the denoised images are visually close to the ground truth in all cases. Therefore, the proposed model is

---

\( ^4 \)The peak signal-to-noise ratio (PSNR) is the ratio between the reference signal and the distortion signal in an image, given in decibels. The higher the PSNR, the closer the distorted image is to the original. Its value is defined as \( \text{PSNR} = 10 \log_{10} \left( \frac{\text{MAX}^2}{\text{MSE}} \right) \), where MSE is the mean-squared error between the original and the distorted image, and \( \text{MAX} \) is the maximum possible pixel value of the image. In our case, since the pixels are represented using 8 bits per sample, \( \text{MAX} = 255 \).
Figure 3. Restored images (first row) and its associated weights (second row) obtained after applying the proposed algorithm for each potential function. The observed data was a noise-free image, and the parameters were fixed to $\lambda = 0.1$ and $\alpha = 10$. Note that pixels have different weights depending on the potential function being used. On the other hand, small gradients are more regularized with $\phi_1$ than with the other potential functions.

Figure 4. Weights associated to the results obtained after applying the proposed algorithm with $\phi_2$ and for several values of $\alpha$. The dual variable can be efficiently modeled by $\alpha$, since it tunes the value of the gradient above which a discontinuity is detected. As $\alpha$ increases, more gradients are penalized.

able to remove the noise from the data and it provides denoised images with closed constant regions surrounded by neat edges.

Once we have confirmed that the half-linear regularization works well on the class of images related to the underlying image model, we want to check its performance on textured images. Figure 6 shows that, as a consequence of Theorem 3.2, the algorithm provides segmented-denoised images where high-frequency details such as texture features are removed. In fact, TV-based denoising algorithms are not able to distinguish between high-frequency signals generated by texture from noise.
Figure 5. Denoising of a piecewise constant image corrupted with Gaussian noise of standard deviation $\sigma = 25$. The results are close to each other, except for $\phi_4$ that performs worse. Despite the high noise level, the denoised images are piecewise constant with neat edges and closed regions.

**Image deconvolution.** Deblurring is the problem of restoring an image that has been blurred and possibly corrupted with noise. Deconvolution refers to the case where the blur to be removed is linear and shift-invariant, so it may be expressed as a convolution with a point spread function. We suppose that the linear operator is given by $Au = \varphi * u$, where $\varphi$ is a Gaussian convolution kernel. In this case, we used Algorithm 3 to solve (18), which requires convolving twice with Gaussian functions. For large values of the standard deviation, this operation in the spatial domain is extremely time consuming. This can be solved by working in the Fourier domain where the convolution becomes a mere multiplication. In fact, the convolution with a normalized Gaussian can be implemented exactly in the discrete Fourier domain [31].

We display first an example of deconvolution for a piecewise constant image. The observed data was simulated by convolving the ground truth with a Gaussian kernel of standard deviation $\sigma = 5$. Figure 7 shows the restored images for each potential function. Although $\phi_1$ provides the best result in terms of RMSE and PSNR values, all restored images are of high quality since the blur has been completely removed.
Figure 6. Denoising of a textured image corrupted with Gaussian noise of standard deviation $\sigma = 10$. As expected, the algorithm provides segmented-denoised images with neat edges, but where high-frequency details are not preserved. This is the case of the house bricks texture.

while edges have been enhanced. Therefore, images look clear and are visually close to the ground truth.

Figure 8 illustrates the restoration of a non-constant image that was convolved with a Gaussian kernel of standard deviation $\sigma = 3$. In this case, our algorithm provides segmented images from blurred data and, thus, the problems of segmentation and deconvolution are solved jointly. For this reason, numerical differences between the segmented images and the ground truth cannot be impressive.

5.3. Comparison with half-quadratic regularization. We shall compare in the next subsection our algorithm with some state-of-the-art TV-based restoration methods solving problem (3)-(4), the half-quadratic approach among themselves. However, let us first exhibit numerically benefits, limitations and complementarities of both regularizers.

We implemented in C/C++ the ARTUR algorithm proposed in [14] with the potential function $\phi_{GM} = t^2/(1 + t^2)$, which satisfies all edge-preserving requirements for half-quadratic regularization. Note that $\phi_{GM}$ is nonconvex but smooth. For our algorithm, we used $\phi_1(t) = \sqrt{\phi_{GM}(t)}$.
First, we tested both algorithms on a piecewise constant image corrupted with Gaussian noise of standard deviation \( \sigma = 20 \). The parameters \( \lambda \) and \( \alpha \) were chosen manually to maximize the PSNR value for each method. Figure 9 reveals that the result provided by ARTUR becomes cartoon-like due to over-smoothing and, thus, some important details have been blurred. This phenomenon does not happen with our algorithm, since image features are highly preserved and edges look clear. These assertions are confirmed by the RMSE and PSNR values, for which our model has better performance.

Both regularizers were also used to denoise a textured image corrupted with Gaussian noise of standard deviation \( \sigma = 10 \). Figure 10 displays the best results for each algorithm. In general terms, our method provides a piecewise constant denoised image, whereas the result provided by ARTUR is piecewise smooth. Note also that important artifacts in the sky appear in the latter case. The proposed dual algorithm also outperforms the other in terms of the error. Anyway, both regularizers remove high-frequency details such as texture features.

In practice, the previous examples suggest that our model is more robust, specially at edges. To show the robustness, the trade-off parameter was voluntary...
Figure 8. Deconvolution of a non-constant image convolved with Gaussian kernel of standard deviation \( \sigma = 3 \). The results are segmented images where the blur has been completely removed. Since the ground truth is not piecewise constant, then it stands to reason that the numerical results are not impressive.

bad adjusted, \( \lambda = 10^{-4} \), and both algorithms were run for the same number of iterations. One sees in Figure 11 that the restored noise-free image has been more regularized with ARTUR than with the proposed algorithm. It is clear that the asymptotic behavior of the half-linear regularizer is a piecewise constant image. On the contrary, the asymptotic behavior of the half-quadratic regularizer is an isotropic smoothed image. Consequently, the half-quadratic criterion yields to an over-smoothing process in such a way that image features are partial or completely removed. Our regularization approach is able to preserve the edges, even when the role of the fidelity-term in the energy functional is almost insignificant.

5.4. Comparison with TV-based image denoising algorithms. The image denoising problem has been intensively studied during the last two decades and numerous algorithms have been proposed and lead to brilliant success. Here we are interested in those arranged in the class of variational methods based on the widely mentioned Rudin-Osher-Fatemi model \([43]\). A detailed comparison between Algorithm 1 and some TV-based denoising algorithms was performed.
Figure 9. Denoising of a piecewise constant image corrupted with Gaussian noise of standard deviation $\sigma = 20$. The values of the parameters were fixed to $\lambda = 0.09$ and $\alpha = 20$ for the half-quadratic model, and to $\lambda = 0.03$ and $\alpha = 20$ for the proposed one. Note that our result better approaches the ground truth since edges are sharp and the overall image looks clear. The denoised image provided by the half-quadratic model suffers from over-smoothing and, as a consequence, image features are blurred.

We compared our algorithm with the Chambolle’s projection algorithm [12], the Split Bregman method [45], and the ARTUR algorithm [14]. Reliable software and online demos on the two first approaches can be found at IPOL webpage [18, 22]. In all cases, the trade-off parameter $\lambda$ was selected manually to maximize the PSNR values. Other algorithm-specific parameters were fixed as given in the corresponding papers. The proposed dual algorithm used $\phi_3$ as potential function. The study was led on the noise-free images displayed in Figure 12. All algorithms were processed
Figure 10. Denoising of a textured image corrupted with Gaussian noise of standard deviation $\sigma = 10$. The parameters were fixed to $\lambda = 0.9$ and $\alpha = 25$ for the half-quadratic model, and to $\lambda = 0.2$ and $\alpha = 45$ for the proposed one. The result provided by the half-quadratic criterion presents artifacts in the sky. On the other hand, both algorithms remove high-frequency details such as the house texture.

Tables 2-4 display the RMSE and the PSNR values of the noisy images and of the denoised images obtained after applying all compared techniques. In all cases, the proposed algorithm provides the best result, which is more than evident as the noise level increases. It is important to remark the superiority of the half-quadratic and the half-linear regularizations on piecewise constant images such as Phantom, being
Figure 11. Robust restoration of a noise-free image. Both algorithms were run for the same number of iterations and the trade-off parameter was fixed to be voluntarily small, $\lambda = 10^{-4}$. Note that the half-quadratic algorithm annihilates most of the edges and image features. On the contrary, the asymptotic behavior of the half-linear regularizer yields a piecewise constant image with neat edges.

Figure 12. Set of grayscale images used for the comparison between the proposed dual algorithm and some state-of-the-art TV-based image denoising methods.

our result the best one also in these cases. We also note that the performance of all methods on images with high-frequency regions such as Building is less impressive. This can be easily justified by taking into account that TV-regularizers cannot distinguish between high-frequencies due to texture from those due to noise. In these cases it is more appropriate to use a patch-based denoising method. However, these algorithms have much greater computational complexity and are usually much more slower than most TV-based denoising techniques.
Table 2. From up to down, the RMSE and PSNR averages over
\( \sigma \in \{5, 10, 15, 20, 25, 30\} \) for each restored image are displayed.

| Noisy image | Chambolle | Split Bregman | Half-quadratic | Half-linear |
|-------------|-----------|---------------|----------------|-------------|
| RMSE | PSNR | RMSE | PSNR | RMSE | PSNR | RMSE | PSNR | RMSE | PSNR |
| 17.38 | 24.65 | 7.21 | 31.50 | 8.09 | 30.44 | 7.76 | 31.06 | 6.99 | 31.88 |
| 17.03 | 24.82 | 10.50 | 28.38 | 11.23 | 27.79 | 10.34 | 28.73 | 9.57 | 29.28 |
| 16.81 | 24.91 | 9.00 | 29.64 | 10.04 | 28.68 | 8.92 | 29.91 | 8.26 | 30.46 |
| 16.47 | 25.03 | 8.80 | 29.98 | 9.95 | 28.93 | 8.62 | 30.27 | 7.88 | 30.96 |
| 14.24 | 26.41 | 7.20 | 32.33 | 9.22 | 30.03 | 5.77 | 34.21 | 5.29 | 35.31 |
| 16.89 | 24.85 | 9.80 | 29.26 | 10.85 | 28.30 | 9.31 | 29.68 | 8.27 | 30.62 |

Table 3. From up to down, the RMSE and PSNR averages over
all restored images for each \( \sigma \in \{5, 10, 15, 20, 25, 30\} \) are displayed.

| Noisy image | Chambolle | Split Bregman | Half-quadratic | Half-linear |
|-------------|-----------|---------------|----------------|-------------|
| RMSE | PSNR | RMSE | PSNR | RMSE | PSNR | RMSE | PSNR | RMSE | PSNR |
| 4.85 | 34.45 | 3.59 | 37.27 | 4.25 | 35.66 | 3.36 | 37.91 | 3.11 | 38.80 |
| 9.60 | 28.51 | 5.94 | 32.79 | 6.69 | 31.69 | 5.48 | 33.56 | 5.13 | 34.22 |
| 14.31 | 25.04 | 8.01 | 30.16 | 8.91 | 29.18 | 7.43 | 30.90 | 6.98 | 31.43 |
| 18.88 | 22.63 | 9.90 | 28.31 | 11.06 | 27.31 | 9.39 | 28.81 | 8.71 | 29.47 |
| 23.40 | 20.77 | 11.65 | 26.90 | 13.14 | 25.83 | 11.45 | 27.08 | 10.36 | 27.93 |
| 27.77 | 19.28 | 13.42 | 25.67 | 15.33 | 24.50 | 13.59 | 25.59 | 11.97 | 26.66 |

Table 4. Global averages of the RMSE and PSNR values over all images and all standard deviations.

| Noisy image | Chambolle | Split Bregman | Half-quadratic | Half-linear |
|-------------|-----------|---------------|----------------|-------------|
| RMSE | PSNR | RMSE | PSNR | RMSE | PSNR | RMSE | PSNR | RMSE | PSNR |
| 16.47 | 25.11 | 8.75 | 30.18 | 9.90 | 29.03 | 8.45 | 30.04 | 7.71 | 31.42 |

In general, a low performance in the RMSE and PSNR values also entails a rejection by a human visual inspection. In spite of this, any numerical criterion cannot fully replace human evaluation, which still is an important criterion for judging the performance of denoising algorithms. A first example on the piecewise constant Phantom image is displayed in Figure 13, where the observed data contains Gaussian noise of standard deviation \( \sigma = 30 \). Note that the denoised images provided by all techniques except ours are over-smoothed and, thus, edges look blurred. Furthermore, Chambolle’s and Split Bregman methods suppress important image features. On the contrary, our algorithm outperforms all other techniques and the result best approaches the ground truth. We achieve to completely remove the high-level noise from the input data while sharpening the edges. Accordingly, the overall denoised image looks clear. Figure 14 displays a second example on the Building image corrupted with Gaussian noise of standard deviation \( \sigma = 15 \). All denoised images are visually similar and of high quality, although the result provided by the ARTUR algorithm presents several artifacts on the sky. However, all methods fail to recover texture features such as the trees in front of the building entrance.

5.5. Extension to color images. For color images, the standard dual definition of the vectorial total variation introduced by Blomgren and Chan [7] is used in place.
Figure 13. Denoising of Phantom image corrupted with Gaussian noise of standard deviation $\sigma = 30$. All results except ours suffer from over-smoothing and, thus, the denoised images become cartoon-like. Furthermore, some image features have disappeared in the results provided by Chambolle’s and Split Bregman methods. On the other hand, our result best approaches the original image because the noise has been completely removed and the overall image looks clear.

of the total variation:

$$VTV(u) = \sum_{1 \leq i,j \leq N} \sqrt{\sum_{1 \leq m \leq M} |(\nabla u_m)_{i,j}|^2},$$

where $u = (u_1, \ldots, u_M)$ and $M$ is the number of color channels. The dual energy (15) is now

$$J^\ast(u, \omega) = \sum_{1 \leq i,j \leq N} \left( \omega_{i,j} \sqrt{\sum_{1 \leq m \leq M} |(\nabla u_m)_{i,j}|^2 + \psi(\omega_{i,j})} \right) + \frac{\lambda}{2} \| A\omega - f \|_X^2,$$

where $f = (f_1, \ldots, f_M)$ is the vector-valued data. Furthermore, the weight function is given by $\omega_{i,j} = \phi' \left( \sqrt{\sum_m |(\nabla u_m)_{i,j}|^2} \right)$. It is straightforward to adapt Algorithms 2 and 3 for solving the vectorial minimization problem (18).

In order to illustrate the behavior of the half-linear regularization, Figure 15 displays an example of color denoising for an image corrupted with Gaussian noise of standard deviation $\sigma = 15$. We used $\phi_3$ as potential function for our algorithm, and we compared it with Chambolle’s projection method. As in the grayscale case, the denoised image provided by Chambolle’s algorithm suffers from over-smoothing.
since some details are blurred. Indeed, note the loss of information on the letters from the cover book. On the contrary, our result looks clear and image features are better preserved.

6. **Conclusions.** In this paper, we have proposed the *half-linear regularization* to minimize objective functionals composed of a quadratic data fidelity-term involving an arbitrary linear operator, and a nonsmooth nonconvex regularization term that depends on the norm of the gradient of the image to be recovered. By imposing some hypotheses on the potential function, we have proved the existence of a solution for the minimization problem. Furthermore, the proposed approach introduces a closed-form auxiliary variable that correctly detects edges and, as the experimental results have exhibited, it preserves them from smoothing. Accordingly, we have presented a dual optimization algorithm that converts the primal problem into a sequence of convex optimization problems. Because of the properties of the minimizers, the solutions are piecewise constant images with neat edges, where the number of the regions and their values are not fixed in advance from noisy and blurred data. We have validated the proposed strategy with theoretical results and we established convergence properties for our algorithm.
Another important issue that should be kept in mind is that our approach limits the eligible potential functions to concave ones. This may be seen as both a limitation and an important contribution. Indeed, the concavity, which cannot be treatable by the half-quadratic regularization, provides remarkable benefits in the image restoration field. Concave potential functions allow to solve jointly the segmentation and the restoration of images from noisy and blurred data in the context of linear inverse problems.

Experiments have convincingly shown that the new-proposed half-linear regularization gives state-of-the-art results in several image restoration tasks such as denoising or deconvolution. Furthermore, it outperforms TV-based image restoration algorithms for restoring piecewise constant images. The results of this paper could be extended to image problems involving TV-regularization. Furthermore, it would be interesting to test a continuation-like strategy. These issues could be a part of a future work.

Appendix A. Proof of Theorem 3.2. Let us denote the set of all pixels by $I = \{1, \ldots, N^2\}$. For each $k \in I$ we rewrite the differential operator $D_k : X \rightarrow \mathbb{R}^2$...
as $D_k u = \left[ (\nabla u)_{i,j}^1, (\nabla u)_{i,j}^2 \right]_T$, where $k = j + (i-1)N$, and its standard norm as

$$\|D_k u\| = \sqrt{(\nabla u)_{i,j}^1)^2 + (\nabla u)_{i,j}^2)^2}.$$ 

With $\tilde{u}$ we associate the subsets $\tilde{I}_0 = \{ i \in I : \|D_i \tilde{u}\| = 0 \}$ and $\tilde{I}_1 = I \setminus \tilde{I}_0 = \{ i \in I : \|D_i \tilde{u}\| > 0 \}$, as well as the vector subspace $V(\tilde{I}_0) = \{ v \in X : \|D_i v\| = 0, \forall i \in \tilde{I}_0 \}$. Henceforth we suppose $\tilde{I}_1 \neq \emptyset$, otherwise $\|D_i \tilde{u}\| = 0$ for all $i \in I$ and, thus, (16) is trivially satisfied.

For $\rho = \frac{\min_{i \in \tilde{I}_1} \|D_i \tilde{u}\|}{\max_{i \in \tilde{I}_1} \|D_i \tilde{u}\|} > 0$, where $\|D_i\|$ denotes the difference operator norm, let $B_\rho(0)$ be the open ball of radius $\rho$ centered at 0. For every $v \in B_\rho(0)$, we have

$$\|D_i (\tilde{u} + v)\| \geq \|D_i \tilde{u}\| - \|D_i v\| \geq \min_{i \in \tilde{I}_1} \|D_i \tilde{u}\| - \max_{i \in \tilde{I}_1} \|D_i\| \|v\|_X$$

$$= (\rho - \|v\|_X) \max_{i \in \tilde{I}_1} \|D_i\| > 0, \forall i \in \tilde{I}_1,$$

which implies that $\phi(\|D_i (\tilde{u} + v)\|) > 0$ at all $i \in \tilde{I}_1$ due to (A1). On the other hand, by noticing that $\tilde{u} \in V(\tilde{I}_0)$ and $\|D_i (\tilde{u} + v)\| = 0$ at $i \in \tilde{I}_0$ for any $v \in V(\tilde{I}_0)$ and, consequently, $\phi(\|D_i (\tilde{u} + v)\|) = 0$ in these cases. Accordingly, the functional $J : V(\tilde{I}_0) \cap B_\rho(0) \rightarrow \mathbb{R}$ defined by

$$J(u) = \sum_{i \in \tilde{I}_1} \phi(\|D_i u\|) + \frac{\lambda}{2} \|Au - f\|_X^2$$

satisfies $J(\tilde{u} + v) = J(\tilde{u} + v)$ for any $v \in V(\tilde{I}_0) \cap B_\rho(0)$, so that $J$ is $C^2$-continuous on a neighbourhood of $\tilde{u}$ and it has a local minimizer at $\tilde{u}$. Hence, $\nabla J(\tilde{u})$ and $\nabla^2 J(\tilde{u})$ are well-defined in the usual sense, and the necessary second-order condition for a minimizer must also hold, that is,

$$\langle \nabla^2 J(\tilde{u}) v, v \rangle \geq 0, \forall v \in V(\tilde{I}_0).$$

We develop now last inequality. For any $i \in \tilde{I}_1$, we have that

$$\langle \nabla^2 \phi(\|D_i \tilde{u}\|) v, v \rangle = \phi''(\|D_i \tilde{u}\|) \frac{\langle D_i \tilde{u}, D_i v \rangle^2}{\|D_i \tilde{u}\|}$$

$$+ \phi'(\|D_i \tilde{u}\|) \frac{\|D_i \tilde{u}\|^2 \|D_i v\|^2 - \langle D_i v, D_i \tilde{u}\rangle^2}{\|D_i \tilde{u}\|^3}$$

and also $\langle \nabla^2 \|A \tilde{u} - f\|_X^2 v, v \rangle = 2 \|Av\|_X^2$, from where (24) writes as

$$\sum_{i \in \tilde{I}_1} \phi''(\|D_i \tilde{u}\|) \frac{\langle D_i \tilde{u}, D_i v \rangle^2}{\|D_i \tilde{u}\|^2} + \sum_{i \in \tilde{I}_1} \phi'(\|D_i \tilde{u}\|) \frac{\|D_i \tilde{u}\|^2 \|D_i v\|^2 - \langle D_i v, D_i \tilde{u}\rangle^2}{\|D_i \tilde{u}\|^3}$$

$$+ \lambda \|Av\|_X^2 \geq 0, \forall v \in V(\tilde{I}_0).$$

On the other hand, let $j \in \tilde{I}_1$ be such that

$$\|D_j \tilde{u}\| \leq \|D_i \tilde{u}\|, \forall i \in \tilde{I}_1,$$

and set $\kappa = \|D_j \tilde{u}\|$. Note that $\kappa > 0$ by definition of $\tilde{I}_1$ and it is finite due to the finiteness of $\tilde{I}_1$. Define $\hat{\tilde{u}} = \kappa^{-2} \tilde{u}$, which clearly satisfies $\hat{\tilde{u}} \in V(\tilde{I}_0)$. We shall show that there exists $\theta > 0$ such that if $0 \leq \|D_j \hat{\tilde{u}}\| < \theta$, then $\langle \nabla^2 J(\hat{\tilde{u}}) \hat{v}, \hat{v} \rangle < 0$, which is a contradiction with (24). First, we have

$$\langle D_i \hat{\tilde{u}}, D_i \hat{v} \rangle = \kappa^{-2} \|D_i \hat{\tilde{u}}\|^2 = \|D_i \tilde{u}\| \|D_i \hat{\tilde{u}}\|, \forall i \in \tilde{I}_1,$$
so that expression (25) becomes
\[ \sum_{i \in \tilde{I}_1} \phi''(\|D_i \hat{u}\|) \frac{\langle D_i \hat{u}, D_i \hat{v} \rangle}{\|D_i \hat{u}\|^2} + \lambda \|A \hat{v}\|^2_X \geq 0. \]

From (26) and (27) we deduce that \( \langle D_i \hat{u}, D_i \hat{v} \rangle^2 = \kappa^{-4}\|D_i \hat{u}\|^4 \geq 1 \). Let us also define \( \mu = \max\{1, \kappa^{-4}\} \). Since \( \phi \) strictly increases from (A1), then [35, Theorem 2.3] states that \( \|A \hat{u}\|_X \leq \|f\|_X \) and, thus, \( \|A \hat{u}\|^2_X = \kappa^{-4}\|A \hat{u}\|^2_X \leq \kappa^{-4}\|f\|^2_X \). Using it, as well as \( \phi''(\|D_i \hat{u}\|) \leq 0 \) at \( i \in \tilde{I}_1 \) by (A3), we get
\[ (28) \quad \langle \nabla^2 J(\hat{u}) \hat{v}, \hat{v} \rangle \leq \sum_{i \in \tilde{I}_1} \phi''(\|D_i \hat{u}\|) \frac{\|D_i \hat{u}\|^2}{\|D_i \hat{u}\|^2} + \lambda \kappa^{-4}\|A \hat{v}\|^2_X \leq \phi''(\|D_i \hat{u}\|) \frac{\|D_i \hat{u}\|^2}{\|D_i \hat{u}\|^2} + \lambda \mu \|f\|^2_X. \]

From (A2)-(A4) it follows that \( \phi''(t) \) is continuous on \((0, +\infty)\) and \( \lim_{t \to +\infty} \frac{\phi''(t)}{t^2} = 0 \). Combining it with (A6), we get that for any \( c \in (0, +\infty) \) there exists an unique \( \theta_c \in (0, +\infty) \) such that \( \phi''(\theta_c) + c = 0 \). If we take \( c = \lambda \mu \|f\|^2_X \in (0, +\infty) \), then there exists an unique \( \theta \in (0, +\infty) \) such that
\[ \frac{\phi''(\theta)}{\theta^2} = -\lambda \mu \|f\|^2_X. \]

If \( 0 < \|D_j \hat{u}\| = \kappa < \theta \) then (A6) yields \(-\infty < \frac{\phi''(\|D_i \hat{u}\|)}{\|D_i \hat{u}\|^2} < \frac{\phi''(\theta)}{\theta^2} < 0 \). Replacing it into (28) implies
\[ \langle \nabla^2 J(\hat{u}) \hat{v}, \hat{v} \rangle < \frac{\phi''(\theta)}{\theta^2} \|D_j \hat{u}\|^2 + \lambda \mu \|f\|^2_X = 0, \]
which contradicts (24). It follows that \( \|D_i \hat{u}\| \geq \|D_j \hat{u}\| \geq \theta \) for all \( i \in \tilde{I}_1 \) and, thus, (16) holds true.

**Appendix B. Proof of Proposition 1.** In the following we use interchangeably \( \partial F \) to denote the sub-differential of a convex function and the upper-differential of a concave function. In the latter case, the supper-differential has to be understood as an equivalent of sub-differential of convex functions, that is, \(-\partial (-F)\).

By definition of the mapping \( M \) in (19), \( u \in S \) if and only if
\[ u = \arg \min_{v \in X} J^*(v, \omega_u), \]
with \( \omega_u = \phi'(\|\nabla u\|) \). Let \( \partial_u \) denote the sub-differential of \( J^* \) with respect to its first variable. The result of the proposition follows from the following equivalences:
\[ \begin{align*}
&u \in S \iff 0 \in \partial_u J^* (u, \omega_u) \\
&\iff 0 \in \partial_u \left\{ \sum_{1 \leq i, j \leq N} \left( (\omega_u)_{i,j} | (\nabla u)_{i,j} | + \psi ((\omega_u)_{i,j}) \right) + \lambda A^* (Au - f) \right\} \iff 0 \in \partial J(u),
\end{align*} \]

which means that \( u \) is a stationary point of the primal functional \( J \). Note that in the second equivalence we use the expression of \( J^* \), given in (15), and in the first transition from the second row we use Theorem 3.1 that writes \( \phi(\|\nabla u\|_{i,j}) = (\omega_u)_{i,j} | (\nabla u)_{i,j} | + \psi ((\omega_u)_{i,j}) \) since \( \omega_u = \phi'(\|\nabla u\|) \).
Appendix C. Proof of Theorem 4.1. We first give some definitions and theoretical results related to the properties of point-to-set maps that will be useful for the proof.

Definition C.1. A point-to-set map $M : X \rightarrow \mathcal{P}(X)$ is said to be:

- uniformly compact on $X$ if there exists a compact set $\Theta$, independent of $u$, such that $M(u) \subseteq \Theta$ for all $u \in X$;
- strictly monotonic with respect to a continuous function $\varphi : X \rightarrow \mathbb{R}$ at a point $u \in X$ if $v \in M(u)$ implies $\varphi(v) < \varphi(u)$, whenever $u$ is not a fixed point of $M$;
- upper semi-continuous at a point $u \in X$ if $u^n \in M(v^n)$, $u^n \to \hat{u}$ and $v^n \to \hat{v}$ imply $\hat{u} \in M(\hat{v})$.

Theorem C.2 (Meyer, [29]). Let $M : X \rightarrow \mathcal{P}(X)$ be a point-to-set map such that it is uniformly compact, upper semi-continuous and strictly monotonic, with respect to a continuous function $\varphi : X \rightarrow \mathbb{R}$. If $\{u^n\}_{n \geq 0}$ is a sequence generated by the algorithm corresponding to $M$, then all its accumulation points are fixed points, $\varphi(u^n) \to \varphi(\hat{u})$ with $\hat{u}$ being a fixed point, $\|u^{n+1} - u^n\|_X \to 0$, and either $\{u^n\}_{n \geq 0}$ converges or its accumulation points form a continuum.

We show now that conditions of Theorem C.2 are met in our case. Using the same notation to identify both the algorithm and the map, let $\{u^n\}_{n \geq 0}$ be a sequence generated via the relation (20). Furthermore, define $\varphi : X \rightarrow \mathbb{R}$ as $\varphi(u) = J^*(u, \omega_u)$, with $\omega_u = \phi'(|\nabla u|)$.

- $M$ is uniformly compact.

Without loss of generality, we can assume that there exists a sufficiently large constant $\nu < +\infty$ such that any minimizer of $J^*$ with respect to its first variable satisfies $\|u\|_\infty \leq \nu$. Since $M(u) = \arg\min_u J^*(v, \omega_u)$, then $\|M(u)\|_\infty \leq \nu$, that is, $M(u)$ lies in the compact set $\Theta = [-\nu, \nu]^{N \times N}$.

- $M$ is strictly monotonic with respect to $\varphi$.

If $\hat{u} \in M(u)$, then $J^*(\hat{u}, \omega_u) \leq J^*(v, \omega_u)$ for any $v \in X$. In particular,

$$J^*(\hat{u}, \omega_u) \leq J^*(u, \omega_u), \tag{29}$$

and the equality holds only when $u$ is also a minimizer of $J^*(\cdot, \omega_u)$, that is, $u \in S$. As a consequence of Theorem 3.1, one has that $J^*(\hat{u}, \omega_u) \leq J^*(\hat{u}, \omega_u)$, so that combining with (29) implies that $J^*(\hat{u}, \omega_u) \leq J^*(u, \omega_u)$, with equality only if $u \in \mathcal{S}$. This ensures the strictly monotonicity of $M$.

- $M$ is upper semi-continuous.

Let $u^n \in M(v^n)$ and suppose that $u^n \to \hat{u}$ and $v^n \to \hat{v}$. Therefore, the discrete gradients satisfy $|\nabla u^n|_{i,j} \to |\nabla \hat{u}|_{i,j}$ and $|\nabla v^n|_{i,j} \to |\nabla \hat{v}|_{i,j}$ for any $i,j \in \{1, \ldots, N\}$. Moreover, since $\phi'$ is continuous according to (A2) and Remark 1, then $\phi'(|\nabla u^n|_{i,j}) \to \phi'(|\nabla \hat{u}|_{i,j})$. For a better readability, let us denote $\omega^n = \phi'(|\nabla v^n|)$ for each $n \geq 0$ and $\bar{\omega} = \phi'(|\nabla \hat{v}|)$. From $u^n \in M(v^n)$, we deduce that $J^*(u^n, \omega^n) \leq J^*(u, \omega^n)$ for any $u \in X$. Using (15), we obtain

$$\sum_{1 \leq i,j \leq N} \left( \omega^n_{i,j} |(\nabla u^n)_{i,j}| + \psi(\omega^n_{i,j}) \right) + \frac{\lambda}{2} \|Au^n - f\|_X^2$$

$$\leq \sum_{1 \leq i,j \leq N} \left( \omega^n_{i,j} |(\nabla u)_{i,j}| + \psi(\omega^n_{i,j}) \right) + \frac{\lambda}{2} \|Au - f\|_X^2, \quad \forall u \in X.$$

Inverse Problems and Imaging Volume 9, No. 2 (2015), 337–370
Because all functions and operators in (30) are continuous, taking the limit yields
\[
\sum_{1 \leq i,j \leq N} \left( \hat{\omega}_{i,j} \right| (\nabla \hat{u})_{i,j} | + \psi (\hat{\omega}_{i,j}) ) + \frac{\lambda}{2} \| A \hat{u} - f \|_X^2 \\
\leq \sum_{1 \leq i,j \leq N} \left( \hat{\omega}_{i,j} \right| (\nabla u)_{i,j} | + \psi (\hat{\omega}_{i,j}) ) + \frac{\lambda}{2} \| A u - f \|_X^2, \quad \forall u \in X,
\]
whence one deduces that \( J^*(\hat{u},\hat{\omega}) \leq J^*(u,\hat{\omega}) \) for any \( u \in X \). Consequently, \( \hat{u} \in \mathcal{M}(\tilde{v}) \).

Having shown that Theorem C.2 applies, then all accumulation points of \( \{ u^n \} \) are fixed points, \( J^*(u^n,\omega_n) \) converges to \( J^*(\hat{u},\hat{\omega}) \) with \( \hat{u} \) being a fixed point, \( \| u^{n+1} - u^n \|_X \to 0 \), and either \( \{ u^n \} \) converges or its accumulation points form a continuum. Now, we are able to demonstrate Theorem 4.1:

i) Since \( J(u^n) = J^*(u^n,\phi'(|\nabla u^n|)) \) and \( \phi'(|\nabla u^n|) = \omega_{u^n} \) for every \( n \geq 0 \), then the monotonicity of \( \mathcal{M} \) guarantees that \( \{ J(u^n) \} \) decreases. Furthermore, \( J^*(u^n,\omega_{u^n}) \to J^*(\hat{u},\hat{\omega}) \) implies that the sequence of energies converges to \( J(\hat{u}) = J^*(\hat{u},\hat{\omega}) \), where \( \hat{u} \) is a fixed point of \( \mathcal{M} \) and, by Proposition 1, it must be a stationary point of \( J \).

ii) By compactness of \( \mathcal{M} \), any sequence has convergent subsequences. Their accumulation points are fixed points according to Theorem C.2 and, by Proposition 1, they are stationary points of \( J \).

iii) and iv) follow directly from Theorem C.2.

v) By (ii) the accumulation points of \( \{ u^n \} \) are stationary points of \( J \) and, since \( J \) has only isolated stationary points, they cannot form a continuum in \( S \). Consequently, the sequence \( \{ u^n \} \) must converge to a stationary point of \( J \) due to (iv).

vi) Suppose that there exists \( \epsilon > 0 \) and a subsequence \( \{ u^{n_k} \}_{k \geq 0} \) of \( \{ u^n \}_{n \geq 0} \) such that
\[
\| u^{n_k} - \hat{u} \|_X \geq \epsilon, \quad \forall k \geq 0, \quad \forall \hat{u} \in S.
\]
By compactness of \( \mathcal{M} \), there exists a subsequence \( \{ u^{n_{k_j}} \}_{j \geq 0} \) of \( \{ u^{n_k} \}_{k \geq 0} \) such that it converges to some \( u^* \in X \). But, by (ii) we necessarily have that \( u^* \in S \), which contradicts (31).

vii) This statement follows from Proposition 1 and from the following theorem:

**Theorem C.3** (Meyer, [29]). Let \( \mathcal{M} \) satisfy the conditions of Theorem C.2. If \( \hat{u} \) is an isolated fixed point of \( \mathcal{M} \) that is also a strong local minimum of \( J \), then there exists an open set \( B \) such that \( u^0 \in B \cap X \) implies that \( \{ u^n \}_{n \geq 0} \) converges to \( \hat{u} \).

**Acknowledgments.** The authors would like to thank Prof. Ernesto Aranda from University of Castilla - La Mancha for many helpful scientific discussions on the proof of Theorem 2.1. The authors would also like to thank the four anonymous referees for their valuable comments.

**REFERENCES**

[1] H. Attouch, J. Bolte, P. Redont and A. Soubeyran, Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Lojasiewicz inequality, *Math. Oper. Res.*, 35 (2010), 438–457.
[2] G. Aubert, A. El Hamidi, C. Ghannam and M. Ménard, On a class of ill-posed minimization problems in image processing, *J. Math. Anal. Appl.*, **352** (2009), 380–399.

[3] G. Aubert and P. Korporobst, *Mathematical Problems in Image Processing*, Applied Mathematical Sciences, Springer-Verlag, New York, 2006.

[4] A. Beck and M. Teboulle, Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems, *IEEE Trans. Image Process.*, **18** (2009), 2419–2434.

[5] J. Bect, L. Blanc-Féraud, G. Aubert and A. Chambolle, A $\ell^1$–unified variational framework for image restoration, in *Proc. of 8th European Conf. Computer Vision (ECCV)*, Lecture Notes Comput. Sci., 3024, Springer-Verlag, 2004, 1–13.

[6] J. E. Besag, *Digital image processing: Towards Bayesian image analysis*, *J. Appl. Stat.*, **16** (1989), 395–407.

[7] P. Blomgren and T. F. Chan, *Color TV: Total variation methods for restoration of vector-valued images*, *IEEE Trans. Image Process.*, **7** (1998), 304–309.

[8] J. Bolte, S. Sabach and M. Teboulle, Proximal alternating linearized minimization for non-convex and nonsmooth problems, *Math. Program.*, **146** (2014), 459–494.

[9] J. V. Burke, A. S. Lewis and M. L. Overton, A robust gradient sampling algorithm for nonsmooth, nonconvex optimization, *SIAM J. Optim.*, **15** (2005), 751–779.

[10] G. Carlier and M. Comte, On a weighted total variation minimization problem, *J. Funct. Anal.*, **250** (2007), 214–226.

[11] F. Catté, P.-L. Lions, J.-M. Morel and T. Coll, Image selective smoothing and edge detection by nonlinear diffusion, *SIAM J. Numer. Anal.*, **29** (1992), 182–193.

[12] A. Chambolle, An algorithm for total variation minimization and applications, *J. Math. Imaging Vis.*, **20** (2004), 89–97.

[13] A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, *J. Math. Imaging Vis.*, **40** (2011), 120–145.

[14] P. Charbonnier, L. Blanc-Féraud, G. Aubert and M. Barlaud, Deterministic edge-preserving regularization in computed imaging, *IEEE Trans. Image Process.*, **6** (1997), 298–311.

[15] X. Chen and W. Zhou, Smoothing nonlinear conjugate gradient method for image restoration using nonsmooth nonconvex minimization, *SIAM J. Imaging Sci.*, **3** (2010), 765–790.

[16] P. L. Combettes and V. Wajs, Signal recovery by proximal forward-backward splitting, *SIAM J. Multiscale Model. Simul.*, **4** (2005), 1168–1200.

[17] A. H. Delaney and Y. Bresler, Globally convergent edge-preserving regularized reconstruction: An application to limited-angle tomography, *IEEE Trans. Image Process.*, **7** (1998), 204–221.

[18] J. Duran, B. Coll and C. Sbert, Chambolle’s projection algorithm for total variation denoising, *Image Process. On Line*, **3** (2013), 301–321.

[19] D. Geman and G. Reynolds, Constrained restoration and the recovery of discontinuities, *IEEE Trans. Pattern Anal. Mach. Intell.*, **14** (1993), 367–383.

[20] D. Geman and C. Yang, Nonlinear image recovery with half-quadratic regularization, *IEEE Trans. Image Process.*, **4** (1995), 932–946.

[21] S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *Journal of Applied Statistics*, **20** (1993), 25–62.

[22] P. Getreuer, *Rudin-Osher-Fatemi total variation denoising using Split Bregman*, *Image Process. On Line*, **2** (2012), 74–95.

[23] T. Goldstein and S. Osher, The Split Bregman method for $\ell^1$–regularized problems, *SIAM J. Imaging Sci.*, **2** (2009), 323–343.

[24] A. Hamidi, M. Ménard, M. Lugiez and C. Ghannam, Weighted and extended total variation for image restoration and decomposition, *Pattern Recogn.*, **43** (2010), 1564–1576.

[25] M. Hintermuller and T. Wu, Nonconvex TV$^q$ models in image restoration: Analysis and a trust-region regularization-based superlinearly convergent solver, *SIAM J. Imaging Sci.*, **6** (2013), 1385–1415.

[26] K. Ivanov, Conditions for well-posedness in the Hadamard sense in spaces of generalized functions, *Siberian Math. J.*, **28** (1987), 906–911.

[27] S. Kindermann, S. Osher and P. W. Jones, Deblurring and denoising of images by nonlocal functionals, *SIAM J. Multiscale Model. Simul.*, **4** (2005), 1091–1115.

[28] A. Marquina and S. Osher, Explicit algorithms for a new time dependent model based on level set motion for nonlinear deblurring and noise removal, *SIAM J. Sci. Comput.*, **22** (2000), 387–405.

[29] R. R. Meyer, Sufficient conditions for the convergence of monotonic mathematical programming algorithms, *J. Comput. Syst. Sci.*, **12** (1976), 108–121.
[30] J.-J. Moreau, Fonctions convexes duales et points proximaux dans un espace h"{e}lbertien, C. R. Acad. Sci. Paris S"{e}r. A Math., 255 (1962), 2897–2899.
[31] J.-M. Morel and G. Yu, Is SIFT scale invariant?, Inverse Probl. Imag., 5 (2011), 115–136.
[32] D. Mumford and J. Shah, Optimal approximations by piecewise smooth functions and associated variational problems, Comm. Pure and Appl. Math., 42 (1989), 577–685.
[33] Y. Nesterov, Smooth minimization of non-smooth functions, Math. Program., 103 (2005), 127–152.
[34] M. Nikolova, Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers, SIAM J. Numer. Anal., 40 (2002), 965–994.
[35] M. Nikolova, Analytical bounds on the minimizers of (nonconvex) regularized least-squares, Inverse Probl. Imag., 2 (2008), 133–149.
[36] M. Nikolova, M. K. Ng and C.-P. Tam, Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction, IEEE Trans. Image Process., 19 (2010), 3073–3088.
[37] M. Nikolova, M. K. Ng, S. Zhang and W.-K. Ching, Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization, SIAM J. Imaging Sci., 1 (2008), 2–25.
[38] P. Ochs, A. Dosovitskiy, T. Brox and T. Pock, An iterated $\ell^1$ algorithm for non-smooth nonconvex optimization in computer vision, in Proc. of IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR), IEEE, 2013, 1759–1766.
[39] P. Perona and J. Malik, Scale-space and edge detection using anisotropic diffusion, IEEE Trans. Pattern Anal. Mach. Intell., 12 (1990), 629–639.
[40] T. Pock and A. Chambolle, Diagonal preconditioning for first order primal-dual algorithms in convex optimization, in Proc. of IEEE Int. Conf. Computer Vision (ICCV), IEEE, 2011, 1762–1769.
[41] M. Robini, A. Lachal and I. Magnin, A stochastic continuation approach to piecewise constant reconstruction, IEEE Trans. Image Process., 16 (2007), 2576–2589.
[42] M. Robini, T. Rastello and I. Magnin, Simulated annealing, acceleration techniques, and image restoration, IEEE Trans. Image Process., 8 (1999), 1374–1387.
[43] L. I. Rudin, S. J. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D, 60 (1992), 259–268.
[44] J. Weickert, Anisotropic Diffusion in Image Processing, ECMI Series, Teubner-Verlag, Stuttgart, 1998.
[45] W. Yin, S. Osher, D. Goldfarb and J. Darbon, Bregman iterative algorithms for $\ell^1$–minimization with applications to compressed sensing, SIAM J. Imaging Sci., 1 (2008), 143–168.
[46] X. Zhang, M. Burger and S. Osher, A unified primal-dual algorithm framework based on Bregman iteration, J. Sci. Comput., 46 (2011), 20–46.

Received October 2013; revised April 2014.

E-mail address: tomeu.coll@uib.es
E-mail address: joan.duran@uib.es
E-mail address: catalina.sbert@uib.es