J Math Imaging Vis (2009) 34: 307–327
DOI 10.1007/s10851-009-0149-y

Some First-Order Algorithms for Total Variation Based Image Restoration

Jean-François Aujol

Published online: 20 March 2009

Abstract This paper deals with first-order numerical schemes for image restoration. These schemes rely on a duality-based algorithm proposed in 1979 by Bermùdez and Moreno. This is an old and forgotten algorithm that is revealed wider than recent schemes (such as the Chambolle projection algorithm) and able to improve contemporary schemes. Total variation regularization and smoothed total variation regularization are investigated. Algorithms are presented for such regularizations in image restoration. We prove the convergence of all the proposed schemes. We illustrate our study with numerous numerical examples. We make some comparisons with a class of efficient algorithms (proved to be optimal among first-order numerical schemes) recently introduced by Y. Nesterov.

Keywords Algorithms · Duality · Total variation regularization · Image restoration

1 Introduction

During the last 15 years, total variation regularization has known a great success in image processing [4, 5, 20, 47]. It has been used in many applications such as image restoration, image deblurring, image zooming, image inpainting, ... (see [5, 20] and references therein). In all these approaches, a total variation term $\int |Du|$ is to be minimized in some way. The typical problem is the case of image restoration [47] with the minimization of a functional of the type:

$$\int_\Omega |Du| + \frac{1}{2\mu}\|f-u\|^2 \tag{1}$$

$\int |Du|$ stands for the total variation of u [3], and if u is regular it is simply $\int_\Omega |\nabla u|\,dx$. Ω is the image domain, a convex Lipschitz open set in $\mathbb{R}^2$. f is the degraded image to restore. The minimizer u of functional (1) is the restored image we want to compute (see for instance [18] for a thorough mathematical analysis of this problem). μ is a weighting parameter which controls the amount of denoising. In the case of zero mean Gaussian noise, μ can be related to the standard deviation of the noise.

From a numerical point of view, total variation is not straightforward to minimize, since it is not differentiable in zero. A first approach is to regularize it, and instead to consider a term as $\int \sqrt{\beta^2+|\nabla u|^2}\,dx$. We will refer to this choice as smoothed total variation regularization:

$$\int_\Omega \sqrt{\beta^2+|\nabla u|^2}\,dx + \frac{1}{2\mu}\|f-u\|^2 \tag{2}$$

The classical approach is then to use the associated Euler-Lagrange equation to compute the solution. Fixed step gradient descent [47], or later quasi-Newton methods [1, 18, 22, 28, 41, 42] have been proposed for instance (see [5, 20] and references therein). Iterative methods have proved successful [9, 11, 27]. A projected-subgradient method can be found in [24].

Ideas from duality have also been proposed: first by Chan and Golub [21], later by A. Chambolle in [16, 17], and then generalized in [25]. Chambolle's projection algorithm [16] has grown very popular, since it is the first algorithm solving exactly problem (1) and not an approximation like (2), with a complete proof of convergence. Moreover, it is straightforward to implement it. In [53], a very interesting combination of the primal and dual problems has been introduced. Second order cone programming ideas and interior point methods have proved interesting approaches [34, 35]. Recently, it

J.-F. Aujol (✉)
CMLA, ENS Cachan, CNRS, UniverSud, Cachan, France
e-mail: Jean-Francois.Aujol@cmla.ens-cachan.fr

has been shown that graph cuts based algorithms could also be used [17, 26]. Finally, let us notice that it is shown in [51] that Nesterov's schemes [39] provides fast algorithms both for minimizing functional (1) and (2).

In this paper, we revisit Chambolle's projection algorithm. We show that a modification of Chambolle projection algorithm, recently suggested in [17], can be seen as a particular instance of a more general algorithm proposed almost 30 years ago by Bermùdez and Moreno [10]. It is in fact an adaptation of Uzawa algorithm [23] to problem (1). This is the first main contribution of the paper: shedding some new light on these projection based type algorithms. We then apply the approach of Bermùdez and Moreno to smoothed total variation regularization: this gives a new fast algorithm to minimize functionals such as (2). This is the second main contribution of the paper. We also prove the convergence of this new scheme. Notice that Bermùdez and Moreno algorithm has already been used for smoothed total variation based restoration in [2], but with a different numerical scheme. To test the efficiency of these algorithms, inspired by [51], we have decided to make some comparisons with a general class of efficient minimization algorithms introduced by Y. Nesterov in [39]. It has been proved in [51] that they are indeed very efficient for image restoration. We chose to use these type of algorithms, because as in the case of Bermùdez and Moreno approach, they consist in first order schemes and it is proved in [39] that they are optimal (in the sense that no algorithms, using only the values and gradients of the functional to minimize, has a better rate of convergence [38]). We also explain how a recent improvement of these algorithms in [40] can be applied for image restoration. We give some numerical examples of all the schemes introduced in this paper: this is the third main contribution of the paper. Our experiments are in favor of Bermùdez-Moreno approach to get a fast approximation for smoothed total variation regularization, whereas Nesterov schemes seem to perform better for total variation regularization. Notice that to get a highly accurate solution for smoothed total variation regularization, Nesterov's schemes seem also to be the best choice. However, such an accuracy is not necessary for image restoration.

Before presenting the plan of the paper, let us emphasize once more the main contributions of the paper:

- Shedding some new light on the Chambolle projection algorithm [16], by seeing how it can be related to a particular instance of Bermùdez and Moreno algorithm [10].
- Introducing a new and efficient scheme for smoothed total variation based image restoration.
- Presenting numerous numerical comparisons with a general class of algorithms recently introduced by Nesterov [39].

The organization of the paper is the following. In Sect. 2, we recall Bermùdez-Moreno algorithm [10]. We show how it can be applied to total variation regularization in Sect. 3. We also explain the relations between this scheme and Chambolle's projection algorithm [16], and we give some numerical examples. In Sect. 4, we detail how Bermùdez-Moreno algorithm can be applied to smoothed total variation based image restoration, providing a new algorithm to solve this type of problem. We then explain in Sect. 5 how these schemes can be used for image deblurring. Whereas Sects. 2 to 5 are related to applications of Bermùdez and Moreno framework, Sect. 6 concerns a different type of approach (and thus this section can be read independently). In Sect. 6, we recall a general class of minimization algorithms introduced by Y. Nesterov in [39]. These algorithms have proved very efficient in [51] for solving image processing problems. We then explain how a recent improvement of these algorithms in [40] can be applied for image restoration. In Sect. 7 we make some comparisons between the different schemes presented in this paper. Appendix details the proof of convergence of Bermùdez-Moreno algorithm.

2 Bermùdez-Moreno Algorithm

In this section, we present the algorithm proposed by Bermùdez and Moreno in [10]. This is a general minimization algorithm. Surprisingly, this approach seems to have been ignored by the image processing community, although it provides efficient algorithms for solving classical image processing problems as we will see in the next sections of the paper. In particular, it gives an algorithm to solve problem (1) without resorting to some smooth approximation like problem (2). Notice that A. Chambolle's paper [16] with its projection algorithm to solve problem (1) (the first algorithm to solve exactly (1)) was published 12 years after the seminal work of Rudin at al. [47]. And yet, [10], which was published 11 years before [47], already provided a similar algorithm with a proof of convergence. We follow here the presentation of [10] and [31] (Chap. II.3). The general minimization problem considered is the following (and we will see that this framework can be used in many image processing problems):

$$\inf_{z\in V}\left\{\frac{1}{2}\langle Az,z\rangle-\langle g,z\rangle+\psi(z)\right\} \tag{3}$$

with V Hilbert space, ψ a proper convex lower semi continuous (l.s.c.) function defined on V:

$$\psi=\phi\, o\, B^* \tag{4}$$

where E is a Hilbert space, B a bounded linear operator, $B:E\to V$, $B^*:V\to E$, $\phi:E\to\mathbb{R}$. We recall that if H is a convex function, we say that it is proper if $H(x)>-\infty$ for all x, and if there exists x^0 such that $H(x^0)<+\infty$. We denote by dom H the set on which $H(x)<+\infty$ [15, 31].

Assumptions on A In all the paper, we will make the following assumptions on A: A is assumed to be a linear symmetric coercive operator, i.e. there exists $\alpha > 0$ such that for all z in V:

$$\langle Az, z\rangle_V \geq \alpha \|z\|_V^2 \tag{5}$$

Notice that it implies in particular that A is a monotone operator, i.e. $\langle Ay - Az, y - z\rangle \geq 0$ for all $y, z \in V$.

We will also make the following assumptions on A:

$$\begin{cases} A \text{ is continuous on the finite dimensional} \\ \quad \text{subspaces of } V. \\ \text{There exists } z^0 \text{ in dom } \psi \text{ such that:} \\ \quad \frac{\langle Az, z-z^0\rangle + \psi(z)}{\|z\|} \to +\infty \text{ if } \|z\| \to +\infty \end{cases} \tag{6}$$

Notice that in the next sections, all these assumptions will indeed be satisfied. In particular, since we will only consider operators A of the type $A = \gamma I$ for some $\gamma > 0$, the technical assumption (6) will be trivially verified.

Notations We use the following notations [44]. If H is a maximal monotone operator, we denote by H_λ its Yosida approximation (L_λ is the resolvent of λH):

$$H_\lambda = \frac{I - L_\lambda}{\lambda} \quad \text{where } L_\lambda = (I + \lambda H)^{-1} \tag{7}$$

Bermùdez and Moreno derive their results for $H = \partial\phi - \omega I$. Here we choose $\omega = 0$, and we take the operator H as (ϕ being defined in (4)):

$$H = \partial\phi \tag{8}$$

Notice that since ϕ is assumed to be a convex proper lower semi continuous function, its subdifferential $\partial\phi$ is a maximal monotone operator [4, 14, 15, 44]. We had to recall the notions of subdifferential of a convex function, maximal monotone operator, and Yosida approximations, because Bermùdez and Moreno approach is based on convex analysis. The algorithm they propose to compute the solution of (3) relies on the associated Euler-Lagrange equation (which in this case happens to be a subdifferential inclusion). See Appendix for further details.

Algorithm In [10], Bermùdez and Moreno propose to use the following algorithm to minimize (3). y^0 being arbitrary, consider the iterative scheme:

$$\begin{cases} u^m = A^{-1}(g - By^m) \\ y^{m+1} = H_\lambda(B^* u^m + \lambda y^m) \end{cases} \tag{9}$$

They prove the following convergence result (Proposition 3.1 in [10]):

Theorem 1 *Assume that A is a linear symmetric coercive operator satisfying (5) and (6), and that ϕ is a convex proper lower semi continuous function. Assume furthermore that:*

$$0 < \frac{1}{\lambda} < \frac{2\alpha}{\|B^*\|^2} \tag{10}$$

Then the sequence (u^m) defined by (9) is such that: $\lim_{m\to+\infty} u^m = u$ (for the strong topology of V) with u solution of: $g - Au \in B\partial\phi(B^*u)$, i.e. u unique solution of (3). Moreover, $y^m \rightharpoonup y$ in E weak, with: $y \in \partial\phi(B^*u)$.

The proof of Theorem 1 is detailed in Appendix.

Relation with Forward-Backward Splitting It was pointed out to the author by one of the anonymous reviewer that Bermùdez-Moreno algorithm can be seen as a particular instance of Forward-Backward Splitting applied to the dual problem of (3); see also Remark 3.2 in [10]. Theorem 1 can then be deduced as a consequence of results from e.g. [25].

One of the main interest of Theorem 1 is that it is not restricted to the case when ϕ is a support function [31]. However, due to the importance of total variation regularization in image processing, we first consider the case of problem (1) in Sects. 3 and 3.4. We will consider the case of problem (2) in Sect. 4, where ϕ is no longer a support function.

3 Application to Total Variation Regularization

In this section, we show how Bermùdez-Moreno algorithm (9) can be used for total variation regularization. In Sect. 3.1, we first consider the continuous setting to derive the link with Bermùdez-Moreno's work. In Sect. 3.2, we then consider the discrete case and we show that Bermùdez and Moreno algorithm consists in solving the dual problem of (1) with a projected gradient algorithm, whose convergence is guaranteed thanks to Theorem 1. We give some numerical examples in Sect. 3.3. In Sect. 3.4, we explain the connection between Bermùdez-Moreno framework and other existing approaches.

3.1 Continuous Setting

Let us consider the celebrated ROF model [47]:

$$\inf_{u\in L^2(\Omega)} J(u) + \frac{1}{2\mu}\|f - u\|_{L^2(\Omega)}^2 \tag{11}$$

Here $J(u)$ is the total variation of u extended to $L^2(\Omega)$ (since in dimension 2, we have $BV(\Omega) \subset L^2(\Omega)$ [3]):

$$J(u) = \begin{cases} \int_\Omega |Du| & \text{if } u \in BV(\Omega) \\ +\infty & \text{otherwise} \end{cases} \tag{12}$$

In fact, (11) is a particular case of (3). Indeed, take $V = L^2(\Omega)$, $E = (L^2(\Omega))^2$, $A = \frac{1}{\mu} I$, $g = \frac{1}{\mu} f$. A is of course coercive with coercivity constant $\alpha = \frac{1}{\mu}$. $J(u) = \psi(u) = \phi(B^*(u))$, and

$$J(u) = \sup_{v \in K} \langle u, \operatorname{div} v \rangle \tag{13}$$

Hence ϕ is the support function of K (closed convex set in $(L^2(\Omega))^2$):

$$K = \left\{ v \in (L^2(\Omega))^2 \ / \ \operatorname{div} v \in L^2(\Omega), \right. \\ \left. \|v\|_\infty \leq 1 \text{ with } |v| = \sqrt{v_1^2 + v_2^2} \right\} \tag{14}$$

We have:

$$\phi(w) = \sup_{v \in K} \langle w, v \rangle_{(L^2(\Omega))^2}, \\ B = -\operatorname{div} = \nabla^*, \quad \text{and} \quad B^* = \nabla \tag{15}$$

where we have used the fact that K is symmetric to 0. We recall that $\langle w, v \rangle_{(L^2(\Omega))^2} = \langle w_1, v_1 \rangle_{L^2(\Omega)} + \langle w_2, v_2 \rangle_{L^2(\Omega)}$.

Moreover, since ϕ is the support function of K, then $H_\lambda(v)$ is the orthogonal projection of $\frac{v}{\lambda}$ onto K [10, 44], i.e.: $H_\lambda(v) = P_K(\frac{v}{\lambda})$, where if $x = (x_1, x_2)$,

$$P_K(x) = \left(\frac{x_1}{\max\{1, |x|\}}, \frac{x_2}{\max\{1, |x|\}} \right) \tag{16}$$

Bermùdez-Moreno algorithm (9) in this case is: u^0 arbitrary, and:

$$\begin{cases} u^m = f + \mu \operatorname{div} y^m \\ y^{m+1} = P_K(y^m + \frac{1}{\lambda} \nabla u^m) \end{cases} \tag{17}$$

Applying Theorem 1, we get the following result:

Proposition 1 *If $\lambda > \frac{\mu}{2} \|B^*\|^2$, then the sequence (u^m, y^m) defined by scheme (17) is such that $u^m \to u$ (in $L^2(\Omega)$ strong) and $y^m \rightharpoonup y$ (in $L^2(\Omega) \times L^2(\Omega)$ weak), with u solution of (11).*

3.2 Discrete Setting

From now on, and until the end of the paper, we will restrict our attention to the discrete setting. We take here the same notations as in [16]. The image is a two dimension vector of size $N \times N$. We denote by X the Euclidean space $\mathbb{R}^{N \times N}$, and $Y = X \times X$. The space X will be endowed with the inner product $(u, v) = \sum_{1 \leq i,j \leq N} u_{i,j} v_{i,j}$ and the norm $\|u\| = \sqrt{(u, u)}$. To define a discrete total variation, we introduce a discrete version of the gradient operator. If $u \in X$, the gradient ∇u is a vector in Y given by: $(\nabla u)_{i,j} = ((\nabla u)^1_{i,j}, (\nabla u)^2_{i,j})$, with

$$(\nabla u)^1_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j} & \text{if } i < N \\ 0 & \text{if } i = N \end{cases} \quad \text{and}$$

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

The discrete total variation of u is then defined by:

$$J(u) = \sum_{1 \leq i,j \leq N} |(\nabla u)_{i,j}| \tag{18}$$

We also introduce a discrete version of the divergence operator. We define it by analogy with the continuous setting by $\operatorname{div} = -\nabla^*$ where ∇^* is the adjoint of ∇: that is, for every $p \in Y$ and $u \in X$, $(-\operatorname{div} p, u)_X = (p, \nabla u)_Y$. It is easy to check that:

$$(\operatorname{div}(p))_{i,j} = \begin{cases} p^1_{i,j} - p^1_{i-1,j} & \text{if } 1 < i < N \\ p^1_{i,j} & \text{if } i = 1 \\ -p^1_{i-1,j} & \text{if } i = N \end{cases} \\ + \begin{cases} p^2_{i,j} - p^2_{i,j-1} & \text{if } 1 < j < N \\ p^2_{i,j} & \text{if } j = 1 \\ -p^2_{i,j-1} & \text{if } j = N \end{cases} \tag{19}$$

From now on, we will use these discrete operators. *Notice that in all the rest of the paper (except in the appendix), we place ourself in the discrete setting. We will sometimes use continuous notations; however, the reader has to keep in mind that only the discrete case is considered.*

We will use Meyer G space for oscillating patterns [8, 37]:

$$G = \{v \in X / \exists g \in Y \text{ such that } v = \operatorname{div}(g)\} \tag{20}$$

and if $v \in G$:

$$\|v\|_G = \inf \left\{ \|g\|_\infty / v = \operatorname{div}(g), \ g = (g^1, g^2) \in Y, \right. \\ \left. |g_{i,j}| = \sqrt{(g^1_{i,j})^2 + (g^2_{i,j})^2} \right\} \tag{21}$$

where $\|g\|_\infty = \max_{i,j} |g_{i,j}|$. Moreover, we will use the notation:

$$G_\mu = \{v \in G / \|v\|_G \leq \mu\} \tag{22}$$

With these classical finite differences, we have: $\|\nabla u\|^2 \leq 8\|u\|^2$. Hence $\|\nabla\|^2 = \|\nabla^*\|^2 \leq 8$. And in fact it is possible to show [16] that $\|\nabla\|^2 = \|\nabla^*\|^2 = 8$.

Let us consider new variables:

$$v^m = \frac{u^m}{\mu}, \qquad p^m = y^m, \qquad \tau = \frac{\mu}{\lambda} \tag{23}$$

Then we can rewrite (17) into: p^0 arbitrary, and

$$\begin{cases} v^m = \frac{f}{\mu} + \operatorname{div} p^m \\ p^{m+1} = P_K(p^m + \tau \nabla v^m) \end{cases} \tag{24}$$

Applying Theorem 1, we get the following result:

Proposition 2 *Let X the Euclidean space $\mathbb{R}^{N\times N}$, and $Y = X \times X$. If $\tau < \frac{1}{4}$, then the sequence (v^m, p^m) defined by scheme (24) is such that $v^m \to v$ in X and $p^m \to p$ in Y with μv solution of (11).*

Notice that (24) can be written in a more compact way:

$$p^{m+1} = P_K\left(p^m + \tau \nabla\left(\frac{f}{\mu} + \operatorname{div} p^m\right)\right) \tag{25}$$

3.3 Numerical Examples

We show here some numerical experiments with scheme (24). We will make some comparisons with other existing algorithms in Sect. 7 and study their numerical accuracy. On Fig. 1, we display the classical images Lenna and cameraman that we use in this paper to illustrate our study. We also show their noisy versions (degraded by additive zero mean Gaussian noise with standard deviation $\sigma = 20$, the dynamic range of the gray values of the image being $[0, 255]$). On Fig. 2, we show the restoration we get with (24). These results have the classical behavior of total variation based image restoration.

3.4 Relation with Chambolle Projection Algorithm

In [16], A. Chambolle proposes a nonlinear projection algorithm to minimize the ROF model (11). This algorithm is based on the remark that the solution of (11) is given by $u = f - P_{G_\mu}(f)$, where P_{G_μ} is the orthogonal projector onto G_μ (defined by (22)). [16] gives an algorithm to compute $P_{G_\mu}(f)$. It indeed amounts to finding:

$$\min\{\|\mu \operatorname{div}(p) - f\|_X^2 : p/|p_{i,j}| \le 1\ \forall i, j = 1, \dots, N\} \tag{26}$$

This problem can be solved by a fixed point method: $v^0 = 0$, $p^0 = 0$, and

$$\begin{cases} v^m = \frac{f}{\mu} + \operatorname{div} p^m \\ p^{m+1}_{i,j} = \frac{p^m_{i,j} + \tau(\nabla v^m)_{i,j}}{1 + \tau|(\nabla v^m)_{i,j}|} \end{cases} \tag{27}$$

It is shown in [16] that if $\tau < 1/8$ in (27), then μv^m converges to the solution of (11). In practice, convergence of (27) is generally observed as long as $\tau < 1/4$. An extension of this algorithm to color images has been proposed in [13]. The case of more general Hilbert space has been considered in [7].

In [17], A. Chambolle has proposed a modification of his projection algorithm. Instead of using (27), he suggests in [17] to use a simple projected gradient method to compute the projection P_{G_μ}:

$$\begin{cases} v^m = \frac{f}{\mu} + \operatorname{div} p^m \\ p^{m+1}_{i,j} = \frac{p^m_{i,j} + \tau(\nabla v^m)_{i,j}}{\max\{1, |p^n_{i,j} + \tau(\nabla v^m)_{i,j}|\}} \end{cases} \tag{28}$$

And this last equation is exactly scheme (24). In [17], A. Chambolle has proved the stability of (28). However, since the functional is not elliptic [45], the convergence of the projection algorithm is not straightforward. In this paper, the convergence of (28) provided $\tau < 1/4$ is a consequence of Proposition 2. See also [30] where a direct proof of convergence (inspired by this work)) of a similar projection algorithm is proposed. Notice that a partial proof of convergence of the projection algorithm has independently been proposed in [54]: the authors get the same type of result as the one of Proposition 2 here, but with only the convergence of v^m in (24) (in their result, the sequence p^m is not guaranteed to converge). Moreover, as we will see in the next section, the general algorithm (9) proposed by Bermùdez and Moreno [10] can be of interest to other image restoration problems, such as smoothed total variation regularization based ones (2). Numerical comparisons of all these schemes ((24), (27)) will be discussed in Sect. 7.2.

The fact that in the case of the ROF problem (11), Bermùdez-Moreno algorithm (scheme (24) is just a projected gradient algorithm on the dual problem has many implications:

1. Let us notice that problem (11) is of the type:

$$\inf_{u \in Q} E(u) \tag{29}$$

   where E is a convex Lipschitz non differentiable function, and Q a convex closed set. For this type of problem, it can be shown [38] (Theorem 3.2.1, p. 138) that no algorithm (only using the values and gradients of E) has a better rate of convergence than $O(\frac{1}{\sqrt{k}})$ (in term of objective function) uniformly on all problems of the form (29), with k the number of iterations of the algorithm. Nevertheless, it is also proved in [40] (see also [50, Theorem 3.12, p. 36]) that the projected gradient method for minimizing a convex Lipschitz differentiable functional on a closed convex set is of order $O(\frac{1}{k})$. Scheme (24) is therefore an algorithm of order $O(\frac{1}{k})$ for solving (11).
2. It is well-known that the projected gradient algorithm is a particular instance of the *proximal forward-backward algorithm* [25]. This provides a general framework for minimizing the sum of two convex functions. The convergence result of Proposition 2 could also be derived from [25]. This approach has been used for instance in

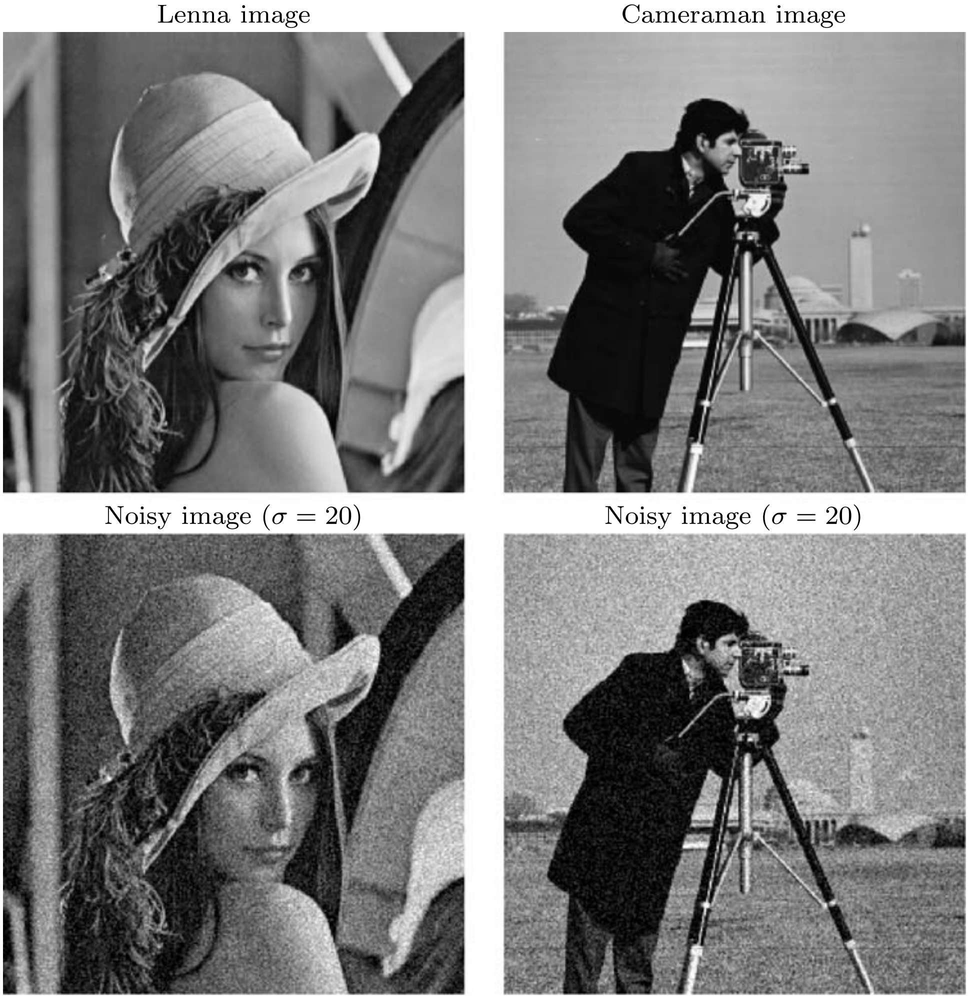


Fig. 1 The classical Lenna and cameraman image, and their noisy version (additive zero mean Gaussian noise with standard deviation $\sigma = 20$)

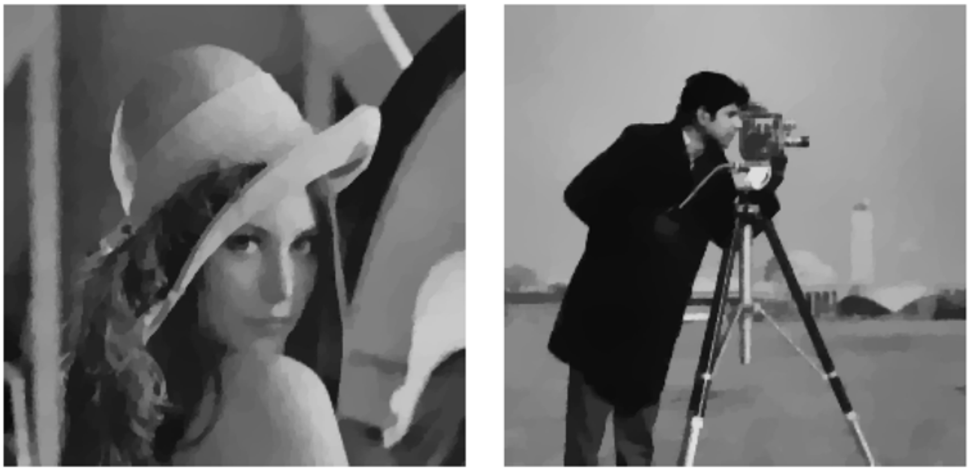


Fig. 2 Total variation restoration of the noisy images presented on the bottom row of Fig. 1 with scheme (24). In both cases, the Lagrange multiplier is $\mu = 30$

[52] to prove the convergence of a similar algorithm to (24). The connection between the projected gradient algorithm and the *proximal forward-backward algorithm* is emphasized in [29, 33]. Of course, as explained at the end of Sect. 2, the relation (in the particular case of total variation regularization) between Bermùdez-Moreno algorithm and the *proximal forward-backward algorithm* comes from the fact that in general Bermùdez-Moreno framework is a particular instance of Forward Backward Splitting. See also [48] where the connection is made between Forward-Backward Splitting and the Split Bregman algorithm recently proposed in [36].

4 Smoothed Total Variation Regularization

In this section, we consider the following problem:

$$\inf_u \int_\Omega \sqrt{\beta^2 + |\nabla u|^2}\, dx + \frac{1}{2\mu}\|f-u\|_2^2 \tag{30}$$

We refer to this problem as the smoothed total variation based regularization problem. For small values of β it can be seen as an approximation of (11). This type of regularization is very common in image processing (see [5, 20] and references therein). Compared to total variation regularization, it has the advantage of being a smooth regularization. And compared to stronger regularization such as $\|\nabla u\|^2$, it has the advantage of not eroding too much the edges of the image.

In Sect. 4.1, we explain how Bermùdez-Moreno algorithm (9) can be used to solve this problem. The new algorithm we propose has a fixed point iteration step. We show the convergence of this fixed point iteration in Sect. 4.3. We will show some numerical examples with this new scheme in Sect. 4.1.

4.1 Presentation of the Scheme

Let us denote by

$$\phi_\beta(\xi) = \int_\Omega \sqrt{\beta^2 + |\xi|^2}\, dx \tag{31}$$

We have

$$\partial\phi_\beta(\xi) = \frac{\xi}{\sqrt{\beta^2+|\xi|^2}} \tag{32}$$

Let us consider the following scheme:

$$\begin{cases} u^m = f + \mu \operatorname{div} y^m \\ y^{m+1} = \frac{I-(I+\lambda\partial\phi_\beta)^{-1}}{\lambda}(\nabla u^m + \lambda y^m) \end{cases} \tag{33}$$

Applying Theorem 1, we get:

Proposition 3 *Let X the Euclidean space $\mathbb{R}^{N\times N}$, and $Y = X\times X$. If $\lambda > 4\mu$, then the sequence (u^m, y^m) defined by scheme (33) is such that $u^m \to u$ in X and $y^m \to y$ in Y with u solution of (30).*

The second equation of (33) implies:

$$\lambda y^{m+1} = \nabla u^m + \lambda y^m - (I+\lambda\partial\phi_\beta)^{-1}(\nabla u^m + \lambda y^m) \tag{34}$$

As in the total variation case, let us set:

$$v^m = \frac{u^m}{\mu} \quad \text{and} \quad \tau = \frac{\mu}{\lambda} \quad \text{and} \quad y^m = p^m \tag{35}$$

(33) becomes:

$$\begin{cases} v^m = \frac{f}{\mu} + \operatorname{div} p^m \\ (I+\lambda\partial\phi_\beta)(\lambda(\tau\nabla v^m + p^m - p^{m+1})) = \lambda(\tau\nabla v^m + p^m) \end{cases} \tag{36}$$

Let us set:

$$w^{m+1} = \tau\nabla v^m + p^m - p^{m+1} \tag{37}$$

From the second line of (36), we get:

$$w^{m+1} + \partial\phi_\beta(\lambda(w^{m+1})) = \tau\nabla v^m + p^m \tag{38}$$

But

$$\partial\phi_\beta(\lambda w^{m+1}) = \frac{\lambda w^{m+1}}{\sqrt{\beta^2 + |\lambda w^{m+1}|^2}} = \frac{w^{m+1}}{\sqrt{\frac{\beta^2}{\lambda^2} + |w^{m+1}|^2}} \tag{39}$$

We thus get from (38)

$$w^{m+1} + \frac{w^{m+1}}{\sqrt{\frac{\beta^2\tau^2}{\mu^2} + |w^{m+1}|^2}} = \tau\nabla v^m + p^m \tag{40}$$

Using the notations $\gamma = \frac{\beta\tau}{\mu}$, and $C^m = \tau\nabla v^m + p^m$, the previous equation becomes:

$$w^{m+1}\left(1 + \frac{1}{\sqrt{\gamma^2 + |w^{m+1}|^2}}\right) = C^m \tag{41}$$

(41) is easily solved with a fixed point iteration. Indeed we have the following result:

Proposition 4 *Let X the Euclidean space $\mathbb{R}^{N\times N}$, and $Y = X\times X$. Consider the sequence $x^0 = w^m$:*

$$x^{k+1} = C^m\left(\frac{\sqrt{\gamma^2+|x^k|^2}}{1+\sqrt{\gamma^2+|x^k|^2}}\right) \tag{42}$$

Then $x^k \to w^{m+1}$ in Y as $k \to +\infty$.

The proof of this result will be detailed in Sect. 4.3. Bermùdez and Moreno algorithm has already been used for smoothed total variation based restoration in [2]. The authors of [2] use a different approach than in this paper. To solve (41), they take the square of both sides of (41), and they use a Newton method to compute $|w^{m+1}|$. They then compute w^{m+1} with (41). But with such an approach, the authors of [2] report poor numerical results. We also tried this approach, and we have seen the same poor results as in [2]. We therefore advocate the use of the fixed point algorithm proposed here to solve (41), which we prove to converge without further assumption (notice that another alternative would be to solve directly (41) with Newton method). The final scheme to solve (30) is thus:

$$\begin{cases} v^m = \frac{f}{\mu} + \operatorname{div} p^m \\ w^{m+1} = (1 + \frac{1}{\sqrt{\frac{\beta^2\tau^2}{\mu^2} + |w^{m+1}|^2}})^{-1}(\tau\nabla v^m + p^m) \\ p^{m+1} = \tau\nabla v^m + p^m - w^{m+1} \end{cases} \tag{43}$$

The second equation is solved with a fixed point iteration (42). We will see that in practice, a single iteration is enough, and thus the second line of (43) reduces to:

$$w^{m+1} = \left(1 + \frac{1}{\sqrt{\frac{\beta^2\tau^2}{\mu^2} + |w^m|^2}}\right)^{-1}(\tau\nabla v^m + p^m) \tag{44}$$

Applying Theorem 1, we have the following convergence result:

Proposition 5 *Let X the Euclidean space $\mathbb{R}^{N\times N}$, and $Y = X \times X$. If $\tau < \frac{1}{4}$, then the sequence (v^m, w^m, p^m) defined by scheme (43) is such that $v^m \to v$ in X, $w^m \to w$ in Y, and $p^m \to p$ in Y with μv solution of (30).*

4.2 Interpretation of Scheme (43)

One first needs to remember that we are interested in solving problem (30). Using the change of notation $v = u/\mu$, solving (30) is equivalent to solving:

$$\inf_v \int \sqrt{\frac{\beta^2}{\mu^2} + |\nabla v|^2}\, dx + \frac{1}{2}\left\|\frac{f}{\mu} - v\right\|^2 \tag{45}$$

The associated Euler-equation is:

$$0 = v - \frac{f}{\mu} - \operatorname{div}\left(\frac{\nabla v}{\sqrt{\frac{\beta^2}{\mu^2} + |\nabla v|^2}}\right) \tag{46}$$

The most classical methods to solve this equation are the fixed step gradient descent as in [47], and the quasi-Newton method (which can be seen also as semi-quadratic regularization) as for instance in [1, 18, 19, 22, 28, 42]. The idea of the quasi-Newton method is to linearize the non-linear term in the above equation, and to consider an iterative scheme of the type:

$$0 = v^{m+1} - \frac{f}{\mu} - \operatorname{div}\left(\frac{\nabla v^{m+1}}{\sqrt{\frac{\beta^2}{\mu^2} + |\nabla v^m|^2}}\right) \tag{47}$$

Here, we propose a different iterative scheme to solve (46)

$$0 = v^m - \frac{f}{\mu} - \operatorname{div} p^m \tag{48}$$

with

$$p^m = \frac{z^m}{\sqrt{\frac{\beta^2}{\mu^2} + |z^m|^2}} \tag{49}$$

In the limit, we would like to have $z^m \to \nabla v$. To update p^m, we use the following equation:

$$p^{m+1} = p^m + \tau(\nabla v^m - z^{m+1}) \tag{50}$$

If (p^m) converges, then $v^m \to v$ with (48), and $z^m \to \nabla v$ with (50) as $m \to +\infty$. The system of equations (48)–(50) can be rewritten into:

$$\begin{cases} v^m = \frac{f}{\mu} + \operatorname{div} p^m \\ z^{m+1}(\tau + \frac{1}{\sqrt{\frac{\beta^2}{\mu^2} + |z^{m+1}|^2}}) = \tau\nabla v^m + p^m \\ p^{m+1} = p^m + \tau(\nabla v^m - z^{m+1}) \end{cases} \tag{51}$$

If we make the change of variable $z^m = w^m/\tau$, then scheme (51) is exactly (43), i.e. Bermùdez-Moreno algorithm for solving problem (30).

4.3 Convergence of the Fixed Point Iteration

In this section, we detail the proof of Proposition 4. The proof relies on Weizfeld method [19, 32, 49]. We adopt here the presentation of [19] for Weizfeld method. Let us first introduce some notations. We consider the following functional:

$$F(u) = \frac{1}{2}\|u - C\|^2 + \|(\gamma^2 + |u|^2)^{1/4}\|^2 \tag{52}$$

We have:

$$\nabla F(u) = u - C + \frac{u}{\sqrt{\gamma^2 + |u|^2}} \tag{53}$$

Let us define:

$$\mathcal{A}(u) = I + \frac{I}{\sqrt{\gamma^2 + |u|^2}} \tag{54}$$

Notice that $u \to \mathcal{A}(u)$ is continuous, and that $\lambda_{\min}(\mathcal{A}(u)) \geq 1$, where $\lambda_{\min}(M)$ stands for the smallest eigenvalue of M. Let us finally define:

$$G(v, u) = F(u) + \langle v - u, \nabla F(u)\rangle + \frac{1}{2}\langle v - u, \mathcal{A}(u)(v - u)\rangle \tag{55}$$

Notice that G consists in a linearization of F. In fact, G defines a general Weizfeld method for the problem:

$$\inf_u F(u) \tag{56}$$

Notice that since F is strictly convex and coercive, there exists a unique u solution of (56), and u is the solution of:

$$\nabla F(u) = u\left(1 + \frac{1}{\sqrt{\gamma^2 + |u|^2}}\right) - C = 0 \qquad (57)$$

We now define the iteration of Weizfeld method:

$$u^{m+1} = \operatorname{argmin}_v G(v, u^m) \qquad (58)$$

Since G is strictly convex and coercive, there exists a unique u^{m+1} solution of (58). It satisfies the Euler-Lagrange equation:

$$\nabla F(u^m) + \langle \mathcal{A}(u^m)(u^{m+1} - u^m)\rangle = 0 \qquad (59)$$

i.e.:

$$u^{m+1}\left(1 + \frac{1}{\sqrt{\gamma^2 + |u^m|^2}}\right) = C \qquad (60)$$

which is precisely iteration (42).

Proposition 6 *If u is fixed, then for all v we have*: $G(v,u) - F(v) \geq 0$.

Proof A standard computation leads to:

$$\begin{aligned} G(v,u) - F(v) &= \left\langle u - v, \frac{-1}{2}\frac{u+v}{\sqrt{\gamma^2 + |u|^2}}\right\rangle \\ &\quad + \int\left(\sqrt{\gamma^2 + |u|^2} - \sqrt{\gamma^2 + |v|^2}\right) dx \\ &= \int \frac{-1}{2}\frac{|u|^2 - |v|^2}{\sqrt{\gamma^2 + |u|^2}} dx \\ &\quad + \int\left(\sqrt{\gamma^2 + |u|^2} - \sqrt{\gamma^2 + |v|^2}\right) dx \end{aligned}$$

Using the notation $a = \sqrt{\gamma^2 + |u|^2}$ and $b = \sqrt{\gamma^2 + |v|^2}$, we get:

$$\begin{aligned} G(v,u) - F(v) &= \int\left(\frac{-1}{2}\frac{a-b}{a} + a - b\right) dx \\ &= \int \frac{(a-b)^2}{2a} dx \geq 0 \qquad (61) \end{aligned}$$

□

The following lemma holds:

Lemma 1 *We have for all m*:

$$F(u^{m+1}) \leq F(u^m) \qquad (62)$$

and

$$\lim_{m \to +\infty} \|u^{m+1} - u^m\| = 0 \qquad (63)$$

Proof From Proposition 6, we have $F(u^{m+1}) \leq G(u^{m+1}, u^m)$. But from (58), we get $G(u^{m+1}, u^m) \leq G(u^m, u^m) = F(u^m)$. We thus deduce inequality (62).

We now concentrate on proving (63). From Proposition 6, we have:

$$\begin{aligned} F(u^{m+1}) &\leq G(u^{m+1}, u^m) \\ &= F(u^m) + \langle u^{m+1} - u^m, \nabla F(u^m)\rangle \\ &\quad + \frac{1}{2}\langle u^{m+1} - u^m, \mathcal{A}(u^m)(u^{m+1} - u^m)\rangle \\ &= F(u^m) - \frac{1}{2}\langle u^{m+1} - u^m, \mathcal{A}(u^m)(u^{m+1} - u^m)\rangle \end{aligned}$$

where we have used (59). We thus deduce that (since $\lambda_{\min}(\mathcal{A}(u)) \geq 1$):

$$\begin{aligned} \frac{1}{2}\|u^{m+1} - u^m\|^2 &\leq \frac{1}{2}\langle u^{m+1} - u^m, \mathcal{A}(u^m)(u^{m+1} - u^m)\rangle \\ &\leq F(u^m) - F(u^{m+1}) \end{aligned}$$

We finally get that:

$$\|u^{m+1} - u^m\| \leq \sqrt{2(F(u^m) - F(u^{m+1}))} \qquad (64)$$

We have just seen before that $F(u^m)$ is a positive, monotone decreasing sequence. Hence $F(u^m)$ is a convergent sequence, and in particular $F(u^m) - F(u^{m+1}) \to 0$, which concludes the proof. □

We are now in position to prove the convergence of the fixed point iteration as stated in Proposition 4:

Proof From (60), one sees that u^m is uniformly bounded. Therefore, up to a subsequence, u^m converges to some v. Moreover, from Lemma 1, we see that u^{m+1} also converges to v. Passing to the limit in (60), we see that $v = u$ where u is the unique minimizer of (56). We conclude that the whole sequence u^m goes to u. □

We end this section by stating a result about the convergence rate of the fixed point algorithm (42). We denote by $\tilde{u}$ the solution of Problem (56). We use the following notations:

$$\gamma^m = \frac{G(\tilde{u}, u^m) - F(\tilde{u})}{\frac{1}{2}\langle \tilde{u} - u^m, \mathcal{A}(u^m)(\tilde{u} - u^m)\rangle} \qquad (65)$$

and

$$\eta = 1 - \lambda_{\min}(\mathcal{A}(\tilde{u})^{-1}\nabla^2 F(\tilde{u})) \qquad (66)$$

Proposition 7

1. $F(u^{m+1}) - F(\tilde{u}) \leq \gamma^m(F(u^m) - F(\tilde{u}))$.

2. *$\eta < 1$ and $0 \leq \gamma^m \leq \eta$, for m sufficiently large. In particular, $F(u^m)$ has a linear convergence rate of at most η.*
3. *u^m is r-linearly convergent with a convergent rate of at most $\sqrt{\eta}$.*

Proof We refer the interested reader to the proof of Theorem 6.1 in [19]. □

5 Image Deconvolution

In this section, we consider the problem of image deconvolution. We explain how Bermùdez-Moreno algorithm (9) can be applied to this problem, by using the iterative approach of [9, 27]. In all the previous sections, we have considered the denoising problem:

$$\frac{1}{2\mu}\|u-f\|^2 + \phi_\beta(u) \tag{67}$$

with the convention that $\phi_0(u) = \int_\Omega |Du|$. As probably noticed by the reader, Bermùdez-Moreno scheme can be applied to functional of the type:

$$\frac{1}{2\mu}\|Au-f\|^2 + \phi_\beta(u) \tag{68}$$

provided that A is an easily invertible operator. However, in the case of image deblurring, the operator A is ill-posed, and we can therefore not apply Bermùdez-Moreno scheme directly. A possible alternative is to use an iterative approach as proposed in [27] or [9]. This type of approach has grown very popular and is now widely used to handle sparsity constraints [27]. Here we use the presentation of [9]. The trick of the method lies in the following result:

Proposition 8 *Let B a linear positive symmetric invertible operator with $\|B\| < 1$. Let $C = B(I-B)^{-1}$. Then, for all u we have*:

$$\langle Bu, u\rangle = \inf_w \|u-w\|^2 + \langle Cw, w\rangle \tag{69}$$

Moreover, the minimum is reached for

$$w = (I+C)^{-1}(u) = (I-B)(u) \tag{70}$$

Here, we choose $\nu > 0$ such that $\nu A^* A < 1$, and we set $B = \nu A^* A$. Let us set:

$$H(u,w) = \frac{1}{2\mu\nu}\left(\|u-w\|^2 + \langle Cw, w\rangle\right) + \frac{1}{2\mu}\left(\|f\|^2 - 2\langle Au, f\rangle\right) \tag{71}$$

Using Proposition 8, it is easy to see that

$$\frac{1}{2\mu}\|f-Au\|^2 = \inf_w H(u,w) \tag{72}$$

Let us now define

$$F(u,w) = H(u,w) + \phi_\beta(u) \tag{73}$$

Let us consider the following algorithm:

$$\begin{cases} w^n = (I - A^*A)(u^n) \\ u^{n+1} = \operatorname{argmin}_u\left(\frac{1}{2\mu\nu}\|w^n + \nu A^* f - u\|^2 + \phi_\beta(u)\right) \end{cases} \tag{74}$$

Setting $v^n = w^n + \nu A^* f$, it can be written:

$$\begin{cases} v^n = u^n + \nu A^*(f - Au^n) \\ u^{n+1} = \operatorname{argmin}_u\left(\frac{1}{2\mu\nu}\|v^n - u\|^2 + \phi_\beta(u)\right) \end{cases} \tag{75}$$

The following convergence result is shown in [9]:

Proposition 9 *Let X the Euclidean space $\mathbb{R}^{N\times N}$. The sequence (u^n, v^n) defined by scheme (75) is such that $(u^n, v^n) \to (u, v)$ in $X \times X$ with $(u, v - \nu A^* f)$ minimizer of (73).*

In practice, to solve the second line of (75) we use scheme (24) if $\beta = 0$ and scheme (33) if $\beta > 0$. Notice that (75) can also be interpreted as a Forward-Backward splitting algorithm [25] to solve problem (68).

Numerical Results (75) is very easy to implement. Notice that in such an iterative approach, one of the key point is to be able to solve each iteration efficiently, which is the case with scheme (24) or (33). We show some numerical results on Fig. 3. As expected, total variation regularization deconvolution gives sharper edges, whereas smoothed total variation based deconvolution preserves better the textures.

6 Nesterov Algorithms

In the previous section, we have introduced first order numerical schemes (24) or (33) to solve image restoration problems. To see how efficient they are, we have decided to compare them with state of the art first order numerical schemes. It has been shown in [51] that Nesterov schemes are very efficient to solve image restoration problems: they beat all the other existing first order algorithms. These schemes were recently introduced in [38, 39], and they have proved to be a significant improvement in convex optimization. Nevertheless, except in the work by P. Weiss et al. [51], Nesterov schemes have not been applied yet in the image processing community. Notice that contrary to the previous sections, Sect. 6 is not

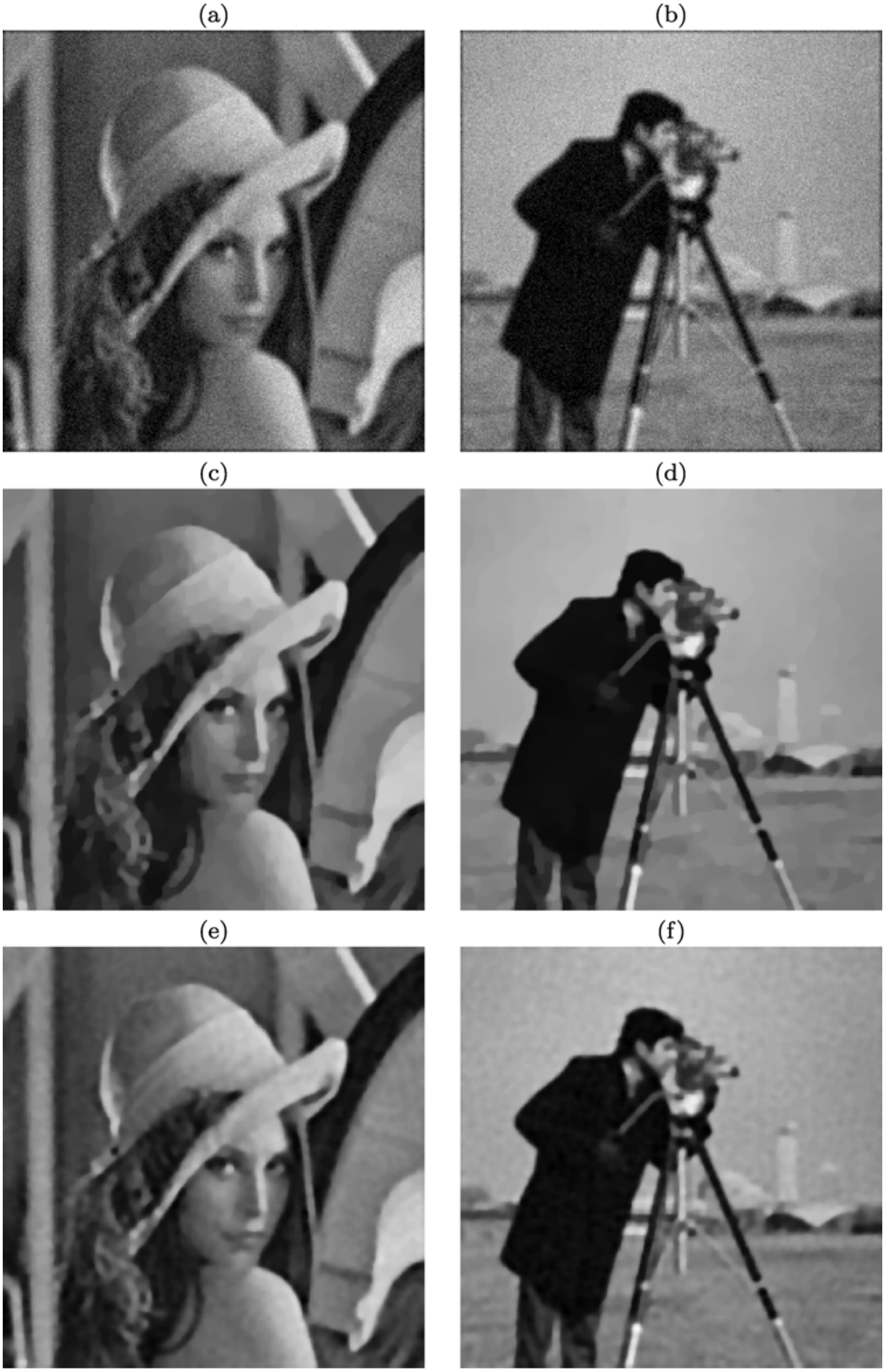


Fig. 3 (**a**) and (**b**): degraded images (convolved with a Gaussian kernel with standard deviation $\eta = 5$, and then degraded by a zero mean Gaussian noise with standard deviation $\sigma = 10$), the original images are on top row of Fig. 1; (**c**) and (**d**): total variation restoration with schemes (75) and (24), with $\lambda = 5$; (**e**) and (**f**): smoothed total variation based restoration with schemes (75) and (43), with $\lambda = 5$ and $\beta = 10$

related to Bermùdez and Moreno framework, and thus it can be read independently. Let us emphasize again that we present Nesterov algorithms because they are optimal first order schemes, and because we want to compare Bermùdez and Moreno approach with state of the art first order schemes.

We first recall Nesterov schemes in Sect. 6.1. Motivated by [51] and our first numerical results, we have decided to implement some improvements of Nesterov schemes recently introduced by Y. Nesterov in [40]. A first variant is presented in Sect. 6.2 and a second one in Sect. 6.3.

6.1 Nesterov Schemes

In [38, 39], Y. Nesterov proposes efficient schemes to minimize functionals such as (11) or (30). We follow here the presentation of [51]. We consider the following minimization problem:

$$\inf_{u\in Q} E(u) \tag{76}$$

where E is a convex Lipschitz differentiable function, and Q a convex closed set. We denote by $\tilde{u}$ a solution of (76). For this type of problem, it can be shown [38] (Theorem 2.1.7, p. 61) that no algorithm (only using the values and gradients of E) has a better rate of convergence than $O(\frac{1}{k^2})$ uniformly on all problems of the form (76) (k is the number of iterations of the algorithm).

In the framework developed by Nesterov, the convergence rate is in term of objective function, and not in term of distance to the minimizer. For instance, a convergence rate of $O(\frac{1}{k^2})$ for problem (76) means that $|E(u^k)-E(\tilde{u})|\leq\frac{C}{k^2}$, where $\tilde{u}$ is the solution of problem (76), and u^k the approximation of $\tilde{u}$ at iteration k. Of course, without further assumption, it gives no information on the convergence rate of u^k to $\tilde{u}$. To get such a piece of information, a coercivity hypothesis is needed for the functional E. Nevertheless, as shown in [51] and in the present paper, the convergence rate obtained with Nesterov's results are in accordance with the behavior of the algorithms (in the sense that if a scheme is supposed to converge as $O(\frac{1}{k^2})$ and a second one as $O(\frac{1}{k})$, it is indeed numerically observed that the first scheme converges faster to the solution).

Let us notice that the constant hidden in the convergence rate in Nesterov's theory is always proportional to $L\times\|u^0-\tilde{u}\|^2$, where L is the Lipschitz constant of ∇E, and u^0 the intial guess for the minimizer $\tilde{u}$.

In [39] is given an $O(\frac{1}{k^2})$ algorithm for solving problem (76) which we detail here-after (it is thus optimal in the sense of Nesterov).

Let $\|.\|$ be a norm and d a convex function such that there exists $\sigma>0$ and x^0 in Q satisfying for all x the inequality: $d(x)\geq\frac{\sigma}{2}\|x-x^0\|^2$.

1. Set $k=0$, $v^0=0$, $x^0\in Q$, L Lipschitz constant of ∇E.
2. Set $k=k+1$, and compute $\eta^k=\nabla E(x^k)$.
3. Set $y^k=\operatorname{argmin}_{y\in Q}(\langle\eta^k,y-x^k\rangle+\frac{1}{2}L\|y-x^k\|^2)$.
4. Set $v^k=v^{k-1}+\frac{k+1}{2}\eta^k$.
5. Set $z^k=\operatorname{argmin}_{y\in Q}(\frac{L}{\sigma}d(x)+\langle v^k,z\rangle)$.
6. Set $x^{k+1}=\frac{2}{k+3}z^k+\frac{k+1}{k+3}y^k$.

Proposition 10 [39] *The previous algorithm ensures that*:

$$0\leq E(y^k)-E(\tilde{u})\leq\frac{4Ld(\tilde{u})}{\sigma(k+1)(k+2)} \tag{77}$$

The idea behind Nesterov's scheme is similar to the one of the conjugate gradient algorithm [23]: the direction of descent is at step $k+1$ is computed by taken into account the information of the complete sequence of gradient $(\nabla E(x^0),\dots,\nabla E(x^k))$, and not only $\nabla E(x^k)$.

Primal Nesterov Algorithm For $\beta>0$, we remind the reader that we set $\phi_\beta(u)=\int\sqrt{\beta^2+|\nabla u|^2}\,dx$. Nesterov algorithm can be used to solve the following problem:

$$\inf_{u\in K_\alpha}\phi_\beta(f+u) \tag{78}$$

where $K_\alpha=\{x\in L^2/\|x\|_2\leq\alpha\}$.

This problem is equivalent to problem (30) (see [18] for a complete analysis). The advantage of formulation (78) is that Nesterov's scheme can directly be applied. See [51] (Algorithm 2, p. 12) for a detailed implementation of this algorithm. We will refer to it as the *primal Nesterov algorithm*. We just give here the sketch of the algorithm (P_{K_α} is the orthogonal projection onto K_α):

1. Set $k=0$, $v^0=0$, $x^0=0$, $L=\|\mathrm{div}\,\|^2/\beta=8/\beta$.
2. Set $k=k+1$, and compute $\eta^k=-\mathrm{div}\,(\frac{\nabla(x^k+f)}{\sqrt{\beta^2+|\nabla(x^k+f)|^2}})$.
3. Set $y^k=P_{K_\alpha}(x^k-\eta^k/L)$, with $K_\alpha=\{x\in L^2/\|x\|_2\leq\alpha\}$.
4. Set $v^k=v^{k-1}+\frac{k+1}{2}\eta^k$.
5. Set $z^k=P_{K_\alpha}(-v^k/L)$.
6. Set $x^{k+1}=\frac{2}{k+3}z^k+\frac{k+1}{k+3}y^k$.
7. The output of the algorithm is: $u=y^{\lim}+f$.

Dual Nesterov Algorithm Of course, due to the non-differentiability in zero of the total variation, Nesterov scheme cannot be applied directly to problem (11). The basic idea is to apply Nesterov's scheme to the dual version of (11), that is to: $\inf_{f-u\in G_\mu}\frac{1}{2}\|u\|^2$, where G_μ is given by (22), i.e.:

$$\inf_{q\in K} E(q) \tag{79}$$

where $E(q)=\frac{1}{2}\|f-\mu\mathrm{div}\,q\|^2$ and $K=\{x\in L^2\times L^2/\|x\|\leq 1\}$. If we denote by $\tilde{u}$ the solution of (11), and by $\tilde{q}$ the solution of (79), we have $\tilde{u}=f-\mu\mathrm{div}\,\tilde{q}$.

See [51] (Algorithm 3, p. 20) for a detailed implementation of this algorithm. We will refer to it as the *dual Nesterov algorithm*. We just give here the sketch of the algorithm (P_K is the orthogonal projection onto K):

1. Set $k=0$, $v^0=0$, $x^0=0$, $L=\mu\|\mathrm{div}\,\|^2=8\mu$.
2. Set $k=k+1$, and compute $\eta^k=\nabla(f-\mu\mathrm{div}\,(x^k))$.
3. Set $y^k=P_K(x^k-\eta^k/L)$, with $K=\{x\in L^2\times L^2/\|x\|\leq 1\}$.
4. Set $v^k=v^{k-1}+\frac{k+1}{2}\eta^k$.
5. Set $z^k=P_K(-v^k/L)$.

6. Set $x^{k+1} = \frac{2}{k+3} z^k + \frac{k+1}{k+3} y^k$.
7. The output of the algorithm is: $u = f - \mu \operatorname{div}(y^{\lim})$.

Notice that in the *dual Nesterov algorithm*, the set K is included in $L^2 \times L^2$; whereas in the case of the *primal Nesterov algorithm*, the set K_α is embedded in L^2.

In [51], very good numerical results are reported both for the *primal* and the *dual Nesterov algorithms* (much better than steepest gradient descent for instance). We have therefore decided to use them as reference in the comparisons presented here-after. We remind the reader that for the projected gradient scheme (24) for solving (11), we mention in Sect. 3.4 that the convergence rate is $O(\frac{1}{k})$. This was already an improvement over the $O(\frac{1}{\sqrt{k}})$ bound for non differentiable function [38] (Theorem 3.2.1, p. 138). With the *dual Nesterov algorithm*, we have now a $O(\frac{1}{k^2})$ algorithm for solving problem (11).

6.2 Accelerated Nesterov Algorithm

In [40], Y. Nesterov proposes a way to speed up the minimization algorithms introduced in [38]. The idea is to improve the estimation of the Lipschitz constant of the functional to minimize (in view of equation (77)). In this subsection, we show how it can be used for image restoration. Consider the general minimization problem

$$\inf_u E(u) + \psi(u) \tag{80}$$

We set $\phi(u) = E(u) + \psi(u)$, and:

$$\psi(u) = \chi_Q(u) = \begin{cases} 0 & \text{if } u \in Q \\ +\infty & \text{otherwise} \end{cases} \tag{81}$$

Problem (80) is therefore the same as (76). As previously, E is a convex Lipschitz differentiable function, and Q a convex closed set. We denote by $\tilde{u}$ a solution of (80). We set:

$$T_L(y) = \operatorname{argmin}_{x \in Q} m_L(y, x) \tag{82}$$

with

$$m_L(y, x) = E(y) + \langle \nabla E(y), x - y \rangle + \frac{L}{2} \|x - y\|^2 + \psi(x) \tag{83}$$

Moreover, it is shown in [40] that

$$\phi'(T_L(y)) = L(y - T_L(y)) + \nabla E(T_L(y)) - \nabla E(y) \tag{84}$$

In [40] is given an efficient algorithm for solving problem (80):

- Set $k = 0$, $A^0 = 0$, $v^0 = 0$, $x^0 \in Q$, $L_0 = L$ Lipschitz constant of ∇E, $\psi_0(x) = \frac{1}{2}\|x - x^0\|^2$. Set $\gamma_u > 1$ and $\gamma_d \geq 1$.
- Set $L = L_k$.

```
REPEAT: Set $a = \frac{1+\sqrt{1+4A^k L}}{2L}$.
        Set $y = \frac{A^k x^k + a v^k}{A^k + a}$, and compute $T_L(y)$.
        If: $\langle \phi'(T_L(y)), y - T_L(y) \rangle < \frac{1}{2L} \|\phi'(T_L(y))\|_2^2$,
        then $L = \gamma_u L$.
 UNTIL: $\langle \phi'(T_L(y)), y - T_L(y) \rangle \geq \frac{1}{2L} \|\phi'(T_L(y))\|_2^2$
DEFINE  $y^k = y$, $M_k = L$, $a^{k+1} = a$, $A^{k+1} = A^k + a^{k+1}$,
        $L_{k+1} = M_k / \gamma_d$, $x^{k+1} = T_{M_k}(y^k)$,
        $\psi_{k+1}(x) = \psi_k(x) + a^{k+1}(E(x^{k+1})$
        $+\langle \nabla E(x^{k+1}), x - x^{k+1} \rangle + \psi(x))$,
        $v^{k+1} = \operatorname{argmin}_x \psi_{k+1}(x)$.
Output: the output of the algorithm is $u = x^{\lim}$.
```

The following convergence result is shown in [40]:

Proposition 11 [40] *Let L_E the Lipschitz constant of ∇E. Assume that $0 < L_0 \leq L_E$. Then the previous algorithm ensures that*:

$$0 \leq \phi(x^k) - \phi(\tilde{u}) \leq \frac{4\gamma_u L_E \|\tilde{u} - x^0\|^2}{k^2} \tag{85}$$

where we recall that $\phi(u) = E(u) + \psi(u)$.

To apply this new algorithm, the only points to check are how to solve (82) and how to compute v^k. This is explained by the two following lemmas.

Lemma 2 *The solution of problem* (82) *is given by*:

$$T_L(y) = P_Q\left(y - \frac{1}{L}\nabla E(y)\right) \tag{86}$$

with P_Q orthogonal projection onto Q.

Proof It is easy to see that:

$$m_L(y, x) = C(y) + \frac{L}{2}\left\| x - \left(y - \frac{1}{L}\nabla E(y)\right)\right\|_2^2 + \psi(x) \tag{87}$$

where $C(y)$ is a function depending only on y. The result of the lemma follows from the fact that $\psi = \chi_Q$. □

Lemma 3 $v^k = \operatorname{argmin}_x \psi_k(x)$ *is given by*

$$v^k = P_Q\left(x^0 - \sum_{p=1}^{k} a^p \nabla E(x^p)\right) \tag{88}$$

with P_Q orthogonal projection onto Q.

Proof Remembering that $\psi_0(x) = \frac{1}{2}\|x - x^0\|^2$, it is easy to see that:

$$\psi_k(x) = C(k) + \sum_{p=1}^{k} a^p \psi(x) + \frac{1}{2}\left\| x - x^0 + \sum_{p=1}^{k} a^p \nabla E(x^p)\right\|_2^2$$

(89)

where $C(k)$ is a function depending only on k. The result of the lemma follows from the fact that $\psi = \chi_Q$. □

In practice, as proposed in [40], we use $\gamma_u = \gamma_d = 2$.

Application to Problem (78) The above algorithm can directly be applied to (78), with

$$E(u) = \phi_\beta(u+f) = \int \sqrt{\beta^2 + |\nabla(f+u)|^2}\,dx \quad \text{and}$$

$$\psi(u) = \chi_{K_\alpha}(u) \tag{90}$$

where $K_\alpha = \{x \in L^2 / \|x\|_2 \leq \alpha\}$. Of course, one has: $\nabla E(u) = -\mathrm{div}\,(\frac{\nabla(f+u)}{\sqrt{\beta^2+|\nabla(f+u)|^2}})$.

One just has to set $x^0 = 0$, $L = \|\mathrm{div}\,\|^2/\beta = 8/\beta$. The solution is given by $f + x^{\mathrm{lim}}$. Notice that here, the projection onto $Q = K_\alpha$ is straightforward: $P_{K_\alpha}(x) = \frac{\alpha x}{\max\{\alpha, \|x\|_2\}}$.

We will refer to this algorithm as the *accelerated primal Nesterov algorithm*.

Application to Problem (11) The basic idea is to apply the accelerated Nesterov scheme to the dual version of (11), that is to (79)), i.e.:

$$\inf_q E(q) + \psi(q) \tag{91}$$

with $E(q) = \frac{1}{2}\|f - \mu \mathrm{div}\, q\|_2^2$ and $\psi(q) = \chi_K(q)$ with $K = \{g \in L^2 \times L^2, \sqrt{g_1^2 + g_2^2} \leq 1\}$. We therefore have: $\nabla E(q) = \nabla(f - \mu \mathrm{div}\, q)$.

One just has to set $u^0 = 0$, $L = \mu\|\mathrm{div}\,\|^2 = 8\mu$. The solution is given by $f - \mu \mathrm{div}\, x^{\mathrm{lim}}$. Notice that here, the projection onto $Q = K$ is straightforward: $P_K(x_1, x_2) = \frac{1}{\max\{1, \|x\|\}}(x_1, x_2))$, with $x = (x_1, x_2)$ and $\|x\| = \sqrt{x_1^2 + x_2^2}$.

We will refer to this algorithm as the *accelerated dual Nesterov algorithm*.

6.3 Variant for the Accelerated Nesterov Algorithm

In [40], Y. Nesterov proposes in fact a more general algorithm than the one we have presented in Sect. 6.2. We show here how it can be used to solve image restoration problems. We still consider the general minimization problem

$$\inf_u E(u) + \psi(u) \tag{92}$$

But this time ψ is assumed to be a strongly convex function with parameter $\mu_\psi > 0$: in the case when ψ is C^2, it means that the smallest eigenvalue of $\nabla^2 \psi$ is $\mu_\psi > 0$.

We set $\phi(u) = E(u) + \psi(u)$. As previously, E is a convex Lipschitz differentiable function We denote by $\tilde{u}$ a solution of (92). Moreover, in [40] is given an efficient algorithm for solving problem (92): this is exactly the algorithm presented in Sect. 6.2, the only difference being that in the step **REPEAT**, instead of setting $a = \frac{1+\sqrt{1+4A^k L}}{2L}$, we set:

$$a = \frac{b + \sqrt{b^2 + 4A^k b}}{2} \quad \text{with } b = \frac{1 + \mu_\psi A^k}{L} \tag{93}$$

The following convergence result is shown in [40]:

Proposition 12 [40] *Let L_E the Lipschitz constant of E, and μ_ψ the convexity parameter of ψ. Assume that $0 < L_0 \leq L_E$. Then the previous algorithm ensures that* (85) *still holds. Moreover, we also have*:

$$0 \leq \phi(x^k) - \phi(\tilde{u})$$
$$\leq \gamma_u L_E \|\tilde{u} - x^0\|^2 \left(1 + \sqrt{\frac{\mu_\psi}{8\gamma_u L_E}}\right)^{-2(k-1)} \tag{94}$$

Notice that (84) still holds in this case. To apply this new algorithm, the only points to check are how to solve (82) and how to compute v^k. We particularize the problem, and we consider the restoration problem (30), i.e. in (92) we take:

$$E(u) = \phi_\beta(u+f) = \int \sqrt{\beta^2 + |\nabla(f+u)|^2}\,dx \quad \text{and}$$

$$\psi(u) = \frac{1}{2\mu}\|u\|^2 \tag{95}$$

Notice that we have:

$$L_E = \|\mathrm{div}\,\|^2/\beta = 8/\beta \quad \text{and} \quad \mu_\psi = \frac{1}{\mu} \tag{96}$$

The two following lemmas hold.

Lemma 4 *The solution of problem* (82) *is given by*:

$$T_L(y) = \frac{Ly - \nabla E(y)}{L + \frac{1}{\mu}} \tag{97}$$

Proof It is easy to see that:

$$\nabla_x(m_L(y,x)) = \nabla E(y) + L(x-y) + \frac{x}{\mu} \tag{98}$$

□

Lemma 5 $v^k = \mathrm{argmin}_x \psi_k(x)$ *is given by*:

$$v^k = \frac{1}{1 + \frac{\sum_{p=1}^k a^p}{\mu}} \left(x^0 - \sum_{p=1}^k a^p \nabla E(x^p)\right) \tag{99}$$

Proof Remembering that $\psi_0(x) = \frac{1}{2}\|x - x^0\|^2$, it is easy to see that:

$$\psi_k(x) = \frac{1}{2}\|x - x^0\|^2 + \sum_{p=1}^{k} a^p \psi(x) + \sum_{p=1}^{k} a^p (E(x^p) + \langle \nabla E(x^p), x - x^p \rangle) \tag{100}$$

□

In the next section, we will refer to this algorithm as the *variant of the accelerated primal Nesterov algorithm*. In practice, we take $x^0 = 0$, $\gamma_u = 2$ and $\gamma_d = 2$.

7 Numerical Examples

In this section, we present some numerical examples with the schemes introduced in this paper. See also [6] for more numerical results. Notice that all the experiments presented in this paper were run with Matlab, on a laptop with a processor at 2 GHz and 2 Gb of RAM. In all the presented algorithms, the cost of one iteration of the algorithm is proportional to the size of the image. This cost is around 0.03 second for a 256×256 image with either the fixed point algorithm (43), the projected gradient algorithm (24), or Chambolle projection algorithm (27). The primal or dual Nesterov algorithms (Sect. 6.1) have a cost per iteration which is twice higher. This cost per iteration is between 8 and 10 times higher with the different variants of Nesterov algorithms (Sects. 6.2 and 6.3).

In Sect. 7.1, we consider the case of smoothed total variation regularization, and in Sect. 7.2 we are interested in total variation regularization.

7.1 Smoothed Total Variation Regularization

Restoration Results Obtained with the New Scheme (43) We illustrate here the efficiency of scheme (43) (based on Bermùdez and Moreno framework) to solve problem (30). This new scheme (43) has the advantages of being simple and stable. Moreover, it seems quite fast (less then 2 seconds for a 256×256 image to get a normalized L^2 error smaller than 0.5 with $\beta = 1$, the images having their values in the range [0; 255]). On Fig. 4, we show the restoration results we get on the noisy images of Fig. 1. The curvature parameter β of (43) is fixed to 10. As expected, the textures are better preserved with this model than with total variation regularization (compare with Fig. 2), but the edges are not as sharp.

Influence of the Number of Iterations in the Fixed Point Loop (42) We now want to see the speed of convergence of (43), and how it depends on the number of iterations in loop (42), and on the parameter β. For different values of β, we compute an *ideal* image by running 10 000 iterations of (43) with 500 iterations for the fixed point (42). We can then compute at each iteration the L^2 error between a computed image with (43) and the target *ideal* image. On Fig. 5, we show the behavior of the algorithm with respect to the number of iterations for the fixed point iteration, for different values of β. Clearly, it shows that 1 iteration is a very good choice: this will be our choice until the end of the paper. It is also clear that the convergence of (43) is much faster for large values of β.

Notice that there exist some theoretical convergence results about iterative schemes using an inner fixed-point loop. For instance, it is shown in [12] that one fixed-point iteration of the method of [46] is enough to get convergence (for the algorithm of [46]). This confirms the numerical observation made in this paper that in the inner fixed-point loop, one iteration may be sufficient to get convergence.

Comparisons with Nesterov Schemes On Fig. 6, we compare our new algorithm (43) with the *primal Nesterov algorithm* (Sect. 6.1), the *accelerated primal Nesterov algorithm* (Sect. 6.2), and the *variant of the accelerated primal Nesterov algorithm* (Sect. 6.3). Notice that since algorithm (43) uses a fixed point iteration, we refer to it as *fixed point* algorithm in the caption of Fig. 6. The convergence speed of these three last algorithms depends on the Lipschitz constant of the energy to minimize: the smaller this constant, the faster the method. It thus means here the larger β, the faster the method. Notice that here the images we consider have their range in [0, 255] (while for instance in [51] the images are normalized in [0, 1]: this has some impact on the values β proposed here).

To make comparisons, we compute the L^2-norm of the difference between the original image and the *ideal* image (obtained by running (43) with 10 000 iterations). We then set this L^2-norm as the constraint in the *primal Nesterov algorithm* and the *accelerated primal Nesterov algorithm*. It is to be noticed that such a choice makes a small bias in favor of our scheme (43). However, the obtained results are sufficiently convincing to forget this bias.

It can be seen that, the larger β, the faster the algorithms. For large values of β, all the algorithms are fast. However, when β decreases to zero, then scheme (43) seems to bring a significant increase in speed of convergence towards a good approximation. It seems indeed that (43) can lead to a good approximation of the minimizer with few iterations. However, when one is interested in getting a very accurate solution, then the *variant of the accelerated primal Nesterov algorithm* seems to be the best choice. This is in accordance with the result of Proposition 12. Notice that the cost per iteration of scheme (43) is twice lower than for the *primal Nesterov algorithm*, while the *accelerated primal Nesterov algorithm* and its *variant* are between 4 and 5 times slower per iteration than the *primal Nesterov algorithm*.

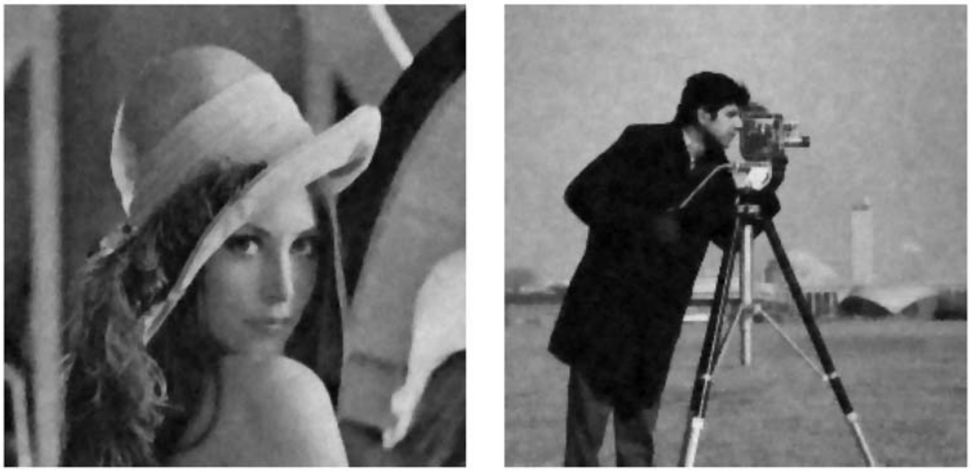
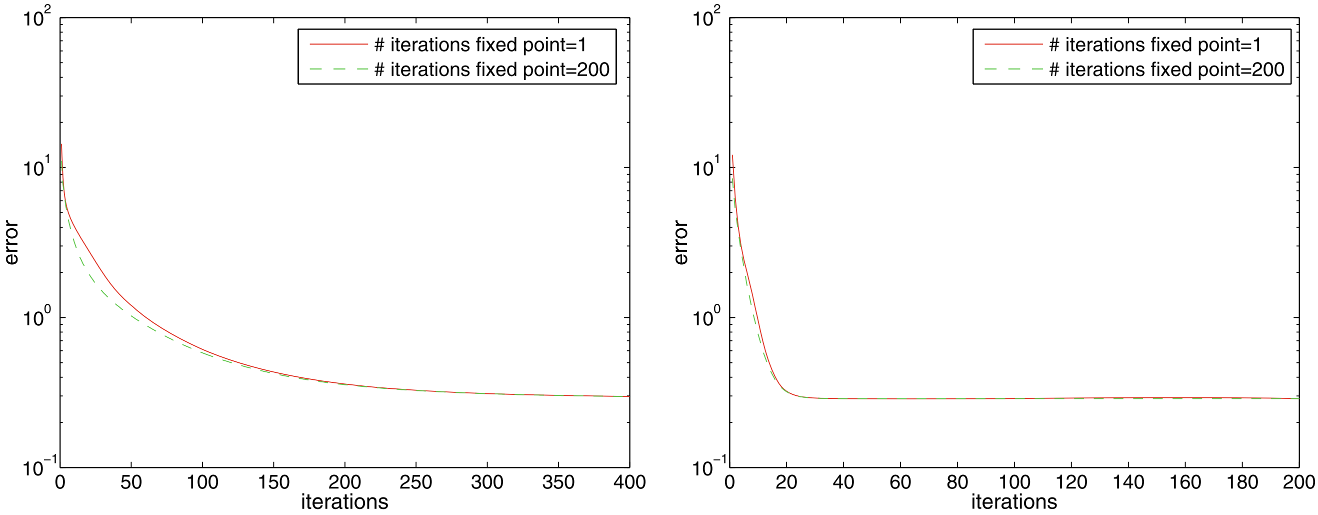


Fig. 4 Smoothed total variation based restoration of the noisy images presented on the bottom row of Fig. 1 with scheme (43) with $\beta = 10$. In both cases, the Lagrange multiplier is $\mu = 30$



(a) $\beta = 0.1$

(b) $\beta = 10$

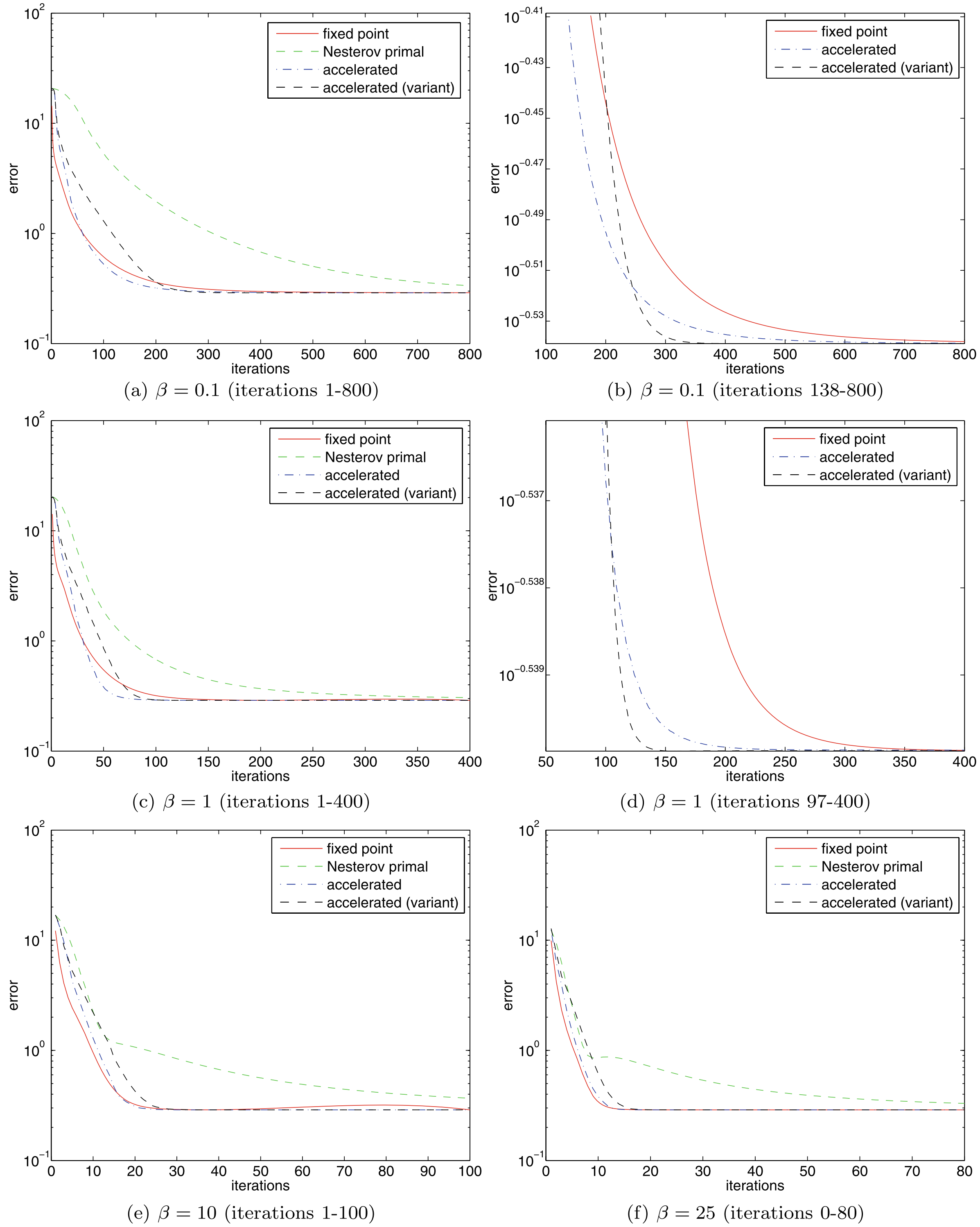
Fig. 5 Comparisons of the number of iterations for the fixed point loop (41) in algorithm (43): 1 or 200. The L^2 error is given with respect to the number of iterations of (43) (vertical logarithmic scale). Graph (**a**) is with $\beta = 0.1$: after 60 iterations of (43), both errors are the same. Graph (**b**) is with $\beta = 10$: after 10 iterations of (43), both errors are the same. We thus advocate to use only 1 iteration for the fixed point iteration (42)

Notice that the quality of the restored image obtained with scheme (43) after a few iterations (10 iterations for $\beta = 25$, 20 iterations for $\beta = 10$, 80 iterations for $\beta = 1$, 200 iterations for $\beta = 0.1$) is visually very good: the normalized L^2 error is then smaller than 0.3. For a restoration purpose, there is no need for the accuracy of the *variant of the accelerated primal Nesterov algorithm*. It is more important to have a fast approximation than a slow and very accurate solution.

7.2 Total Variation Regularization

In this section, we consider problem (11). We want to compare five different algorithms. The first one is the projection algorithm of [16]: we refer to it as *Chambolle projection algorithm*. We use $\tau = 0.249$ in (27). The second one is the modification of this algorithm as proposed in [17], and which we proved to be Bermùdez-Moreno algorithm (24) in the case of problem (11): since it is an adaptation of Uzawa method [23] to problem (11), we refer to it as *Uzawa algorithm*. We use $\tau = 0.249$ in (24). The third algorithm we use here is the *dual Nesterov algorithm* presented in Sect. 6.1, as proposed in [51]. Motivated by the results of [39] and [51], we use it as the reference algorithm. The fourth algorithm we use here is the *accelerated dual Nesterov algorithm* of [40] presented in Sect. 6.2. The fifth algorithm we use is our new scheme (43). Since it uses a fixed point algorithm, we refer to it as *fixed point* method.

For a given image and a given regularization parameter μ, a reference *ideal* image is computed by running 10 000



(a) $\beta = 0.1$ (iterations 1-800)

(b) $\beta = 0.1$ (iterations 138-800)

(c) $\beta = 1$ (iterations 1-400)

(d) $\beta = 1$ (iterations 97-400)

(e) $\beta = 10$ (iterations 1-100)

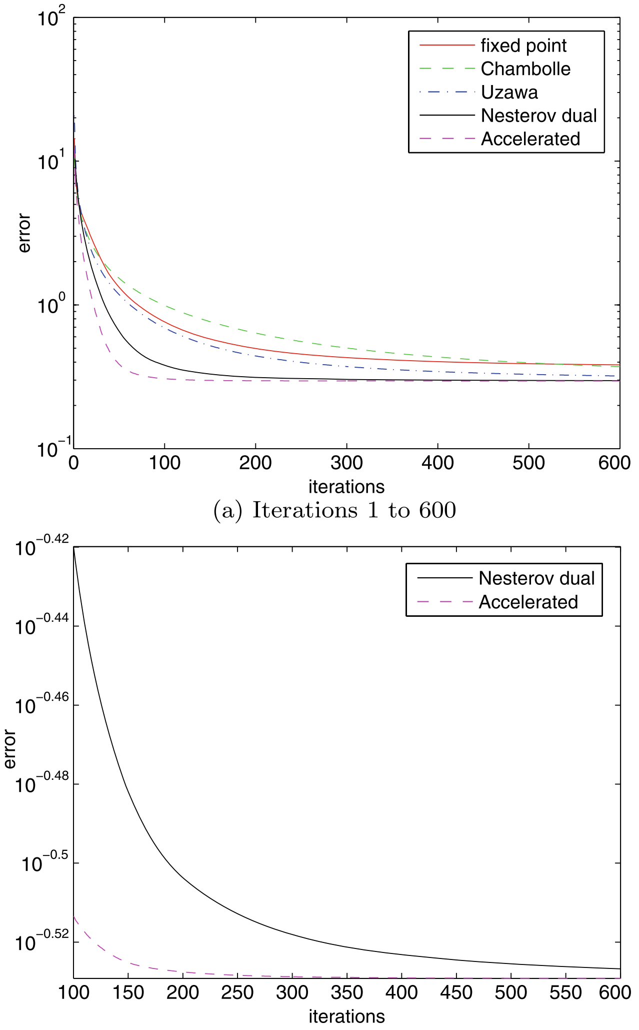
(f) $\beta = 25$ (iterations 0-80)

Fig. 6 Speed of convergence (smoothed total variation regularization): the L^2 error is given with respect to the number of iterations (vertical logarithmic scale). Graph (**a**) and (**b**) is with $\beta = 0.1$; Graph (**c**) and (**d**) with $\beta = 1$; Graph (**e**) with $\beta = 10$; Graph (**f**) with $\beta = 25$. The range of the image is between 0 and 255. On graphs (**a**), (**c**), (**d**), (**e**) and (**f**), from *top* to *bottom* are the speed of convergence of the *fixed point* based algorithm (43), the speed of the *primal Nesterov algorithm*, the speed of the *accelerated primal Nesterov algorithm*, the speed of the *variant of the accelerated primal Nesterov algorithm*. On graphs (**b**) and (**d**), the *primal Nesterov algorithm* is not shown. Notice that the time for 1 iteration of the *primal Nesterov algorithm* is around twice the time for 1 iteration of the *fixed point* based algorithm (43). The *accelerated primal Nesterov algorithm* and its *variant* are between 4 and 5 times slower per iteration than the *primal Nesterov algorithm*. To get a fast approximation, the *fixed point* based algorithm (43) seems to be the best choice (the accuracy is good enough for image restoration). To get a highly accurate solution, the *variant of the accelerated Nesterov scheme* seems to be the most efficient

iterations with the *dual Nesterov algorithm*. Here, the bias will therefore be in favor of the *dual Nesterov algorithm*. However, we think that the results are convincing enough to forget this bias.

A convergence speed result is presented on Fig. 7: we give the L^2-norm of $\tilde{u} - u^n$, where u^n is the computed image at iteration n, and $\tilde{u}$ the *ideal* image to obtain. As can be seen on Fig. 7, the *dual Nesterov* algorithm is faster then *Uzawa algorithm*, which is itself faster than *Chambolle projection algorithm*. The *accelerated dual Nesterov algorithm* seems to be the best choice to get a highly accurate solution (50 iterations to get a normalized L^2 error of 0.3). However, 1 iteration with the *accelerated dual Nesterov algorithm* is around 4 times slower than with the *dual Nesterov algorithm*: the *dual Nesterov algorithm* seems thus a good compromise when one is interested in getting a very good approximation. Nevertheless, 1 iteration with the *dual Nesterov algorithm* is around 2 times slower than with *Uzawa*, *Chambolle*, or scheme (43) (while all three have the same computation time per iteration). For typical image restoration problems (with Gaussian noise), (24) seems 30% faster then (27) (for instance, it takes 70 iterations for (24) to get a normalized L^2 error of 1 while it takes 110 iterations for (27) to get the same accuracy). Algorithm (43) seems to be a good alternative when one is only interested in getting an approximation with a small number of iterations; for instance, scheme (43) is the fastest (in term of computation time) to get a normalized L^2 error of 2 (around 30 iterations).

In [51], the authors explain that the *dual Nesterov algorithm* is much faster then the projected gradient method (24) (*Uzawa* algorithm) for total variation regularization. We confirm that it is indeed much faster when one is interested in computing an accurate solution. Notice also that in [51], the comparison criterion used is the value of the total variation of the computed image. This is indeed the quantity which is controlled in Nesterov's approach for solving (11) (see Proposition 10). Here, the criterion is the L^2 difference of the computed solution for some iteration with the *ideal* solution. Figure 7 is surely in favor of the approach developed in [51]. However, the difference during the first iterations is not that large, and thus the projected gradient algorithm (24) (*Uzawa algorithm*) can still be considered as a good method when one is only interested in getting a fast approximation of the solution.



(a) Iterations 1 to 600

(b) Iterations 100 to 600

Fig. 7 Speed of convergence (total variation regularization): the L^2 norm of the error is given at each iteration (vertical logarithmic scale). (**a**) gives the speed of convergence for iterations 1 to 600, and (**b**) for iterations 100 to 600. On graph (**a**), from *top* to *bottom* are the speed of convergence of the *fixed point* based algorithm (43) with $\beta = 0.1$, the speed of convergence of *Chambolle* projection algorithm (27) with $\tau = 0.249$, the speed of *Uzawa* scheme (24) with $\tau = 0.249$, the speed of the *dual Nesterov algorithm*, and the speed of the *accelerated dual Nesterov algorithm*. On graph (**b**) are only shown the *dual Nesterov algorithm*, and the *accelerated dual Nesterov algorithm*. 1 iteration with the *accelerated dual Nesterov algorithm* is around 4 times slower than with the *dual Nesterov algorithm*. But 1 iteration with the *dual Nesterov algorithm* is itself around 2 times slower than with *Uzawa*, *Chambolle*, or scheme (43) (while all three have the same computation time per iteration). To get a highly accurate solution, the *accelerated dual Nesterov* algorithm seems to be the best choice. However, the *dual Nesterov algorithm* seems to be the best compromise when one is only interested in getting a good approximation (which is the case for image restoration)

Dual Nesterov Algorithm for Solving (30) In view of Fig. 7, one should be tempted to use the *dual Nesterov algorithm* for solving (30). It is easy to compute the dual problem. If we denote by $\tilde{u}$ the solution of (30), then we have $\tilde{u} = f - \mu \operatorname{div} \tilde{p}$ with $\tilde{p}$ solution of:

$$\inf_{p \in K} \frac{1}{2\mu} \|\mu \operatorname{div} p - f\|^2 - \beta \int \sqrt{1 - |p|^2}\, dx \tag{101}$$

where $K = \{p \in L^2 \times L^2 \ / \ \|p\|_\infty \leq 1\}$. However, the gradient of the functional in (101) is not Lipschitz, and we therefore cannot use directly the *dual Nesterov algorithm*.

Acknowledgements The author would like to thank Vicent Caselles, Antonin Chambolle and Pierre Weiss for fruitful discussions about this work. The author also would like to thank the anonymous reviewers for their useful comments.

This work has been done with the support of the French "Agence Nationale de la Recherche" (ANR), under grant NATIMAGES (ANR-08-EMER-009), "Adaptivity for natural images and texture representations".

Appendix: Proof of Convergence of Bermùdez-Moreno Algorithm

In this section, we follow [10] and [31] (Chap. II.3). Our goal is to give the reader some intuition on why the result of Theorem 1 holds. We remind the reader that we use the notations: $H_\lambda = \frac{I-L_\lambda}{\lambda}$, with $L_\lambda = (I+\lambda H)^{-1}$ and $H = \partial\phi$ with ϕ proper convex lower semi continuous function. We will use the next lemma:

Lemma 6

$$\frac{1}{\lambda^2}\left\|L_\lambda(v^1) - L_\lambda(v^2)\right\|^2 + \left\|H_\lambda(v^1) - H_\lambda(v^2)\right\|^2 \leq \frac{1}{\lambda^2}\|v^1 - v^2\|^2 \tag{102}$$

Proof This is an immediate consequence of definitions (7). □

Problem (3) is related to:

$$\forall z, \quad \langle Au, z-u\rangle + \psi(z) - \psi(u) \geq \langle g, z-u\rangle \tag{103}$$

The relation is given by the next lemma (whose proof is straightforward (see [31, Proposition 2.2, p. 37]):

Lemma 7 *u is solution of* (103) *if and only if u is solution of* (3).

We remind the reader that $B\partial\phi(B^*u) = \partial\psi(u)$ (see [31, Proposition 5.7, p. 27]). Problem (103) is related to the subdifferential inclusion:

$$g - Au \in B\partial\phi(B^*u) \tag{104}$$

The relation is given by the next proposition:

Proposition 13 *u is solution of* (104) *if and only if u is solution of* (103).

Proof The fact that u solution of (104) implies that u solution of (103) is a direct consequence of the definition of the subdifferential of a convex function [31]. The reciprocal result is more complicated, and we refer the reader to Chap. II.3 of [31] for a detailed proof. □

We will make use of the next lemma (Lemma 2.1 in [10]):

Lemma 8 *H maximal monotone operator. Then the two following conditions are equivalent*:

(i) $y \in H(v)$
(ii) $y = H_\lambda(v+\lambda y)$

An immediate consequence of the previous lemma is the following result:

Proposition 14 *u is a solution of* (104) *if and only if* (u, y) *is a solution of*:

$$\begin{cases} Au = g - By \\ y = H_\lambda(B^*u + \lambda y) \end{cases} \tag{105}$$

We are now in position to prove Theorem 1.

Proof From (102), we get:

$$\begin{aligned}
&\frac{1}{\lambda^2}\|L_\lambda(B^*u+\lambda y) - L_\lambda(B^*u^m + \lambda y^m)\|_E^2 + \|y - y^{m+1}\|_E^2 \\
&\quad\leq \frac{1}{\lambda^2}\|B^*(u-u^m) + \lambda(y-y^m)\|_E^2 \\
&\quad= \|y-y^m\|_E^2 + \frac{2}{\lambda}\langle B^*(u-u^m), y-y^m\rangle_E \\
&\qquad+ \frac{1}{\lambda^2}\|B^*(u-u^m)\|_E^2
\end{aligned} \tag{106}$$

But if we subtract the first line of (9) to the first line of (105), we have: $A(u-u^m) = B(y^m - y)$. Taking the inner product with $(u-u^m)$, we deduce:

$$\begin{aligned}
\langle A(u-u^m), u-u^m\rangle &= \langle B(y^m - y), u-u^m\rangle \\
&= \langle y^m - y, B^*(u-u^m)\rangle
\end{aligned} \tag{107}$$

Hence:

$$\begin{aligned}
\langle y-y^m, B^*(u-u^m)\rangle &= \langle -A(u-u^m), u-u^m\rangle \\
&\leq -\alpha\|u-u^m\|_E^2 \\
&\leq \frac{-\alpha}{\|B^*\|^2}\|B^*(u-u^m)\|_E^2
\end{aligned} \tag{108}$$

We now deduce from (106) that:

$$\begin{aligned}
&\frac{1}{\lambda^2}\|L_\lambda(B^*u+\lambda y) - L_\lambda(B^*u^m + \lambda y^m)\|_E^2 + \|y - y^{m+1}\|_E^2 \\
&\quad\leq \frac{1}{\lambda}\left(\frac{1}{\lambda} - \frac{2\alpha}{\|B^*\|^2}\right)\|B^*(u-u^m)\|^2 + \|y-y^m\|_E^2
\end{aligned} \tag{109}$$

We eventually get that, since $0 < \frac{1}{\lambda} < \frac{2\alpha}{\|B^*\|^2}$, as long as $u^m \neq u$: $\|y - y^{m+1}\|_E < \|y-y^m\|_E$. We deduce that $\|y-y^m\|_E^2$ is a convergent sequence in $\mathbb{R}$. Thus passing to the limit in (109), we get: $\lim_{m\to+\infty}\|B^*(u-u^m)\|_E=0$. Using (108), we eventually get that $u^m\to u$.

There remains to prove that y^m also converges. We first remark that now, passing to the limit in (109), we get: $L_\lambda(B^*u^m+\lambda y^m)\to L_\lambda(B^*u+\lambda y)$. But since $L_\lambda=I-\lambda H_\lambda$, we get with the second line of (105) that: $L_\lambda(B^*u+\lambda y)=B^*u$. From the second line of (9), we get:

$$\begin{aligned}y^{m+1}&=H_\lambda(B^*u^m+\lambda y^m)\\&=y^m+\frac{1}{\lambda}(B^*u^m-L_\lambda(B^*u^m+\lambda y^m))\end{aligned}\tag{110}$$

Passing to the limit, we eventually get that: $\lim_{m\to+\infty}\{y^{m+1}-y^m\}=0$. Now we can conclude that $y^m\rightharpoonup y$ in E weak, since the application

$$v\in E\to H_\lambda(B^*u(v)+\lambda v)\tag{111}$$

with $u(v)$ solution of: $Au=g-Bv$, is non expansive (see [43, Corollary 4, p. 199]). □

References

1. Acar, R., Vogel, C.: Analysis of total variation penalty methods for ill-posed problems. Inverse Probl. **10**, 1217–1229 (1994)
2. Almansa, A., Ballester, C., Caselles, V., Haro, G.: A TV based restoration model with local constraints. J. Sci. Comput. **34**(3), 209–236 (2008)
3. Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variations and Free Discontinuity Problems, Oxford mathematical monographs. Oxford University Press, London (2000)
4. Andreu-Vaillo, F., Caselles, V., Mazon, J.M.: Parabolic Quasilinear Equations Minimizing Linear Growth Functionals. Progress in Mathematics, vol. 223. Birkhäuser, Basel (2002)
5. Aubert, G., Kornprobst, P.: Mathematical Problems in Image Processing. Applied Mathematical Sciences, vol. 147. Springer, New York (2002)
6. Aujol, J.-F.: Some algorithms for total variation based image restoration. CMLA Report 08-21 (2008). http://hal.archives-ouvertes.fr/hal-00260494/fr/
7. Aujol, J.-F., Gilboa, G.: Constrained and SNR-based solutions for TV-Hilbert space image denoising. J. Math. Imaging Vis. **26**(1–2), 217–237 (2006)
8. Aujol, J.-F., Aubert, G., Blanc-Féraud, L., Chambolle, A.: Image decomposition into a bounded variation component and an oscillating component. J. Math. Imag. Vis., **22**(1), 71–88 (2005)
9. Bect, J., Blanc-Féraud, L., Aubert, G., Chambolle, A.: A l1-unified variational framework for image restoration. In: ECCV 04. Lecture Notes in Computer Sciences, vol. 3024, pp. 1–13. Springer, Berlin (2004)
10. Bermùdez, A., Moreno, C.: Duality methods for solving variational inequalities. Comput. Math. Appl. **7**, 43–58 (1981)
11. Bioucas-Dias, J., Figueiredo, M.: Thresholding algorithms for image restoration. IEEE Trans. Image Process. **16**(12), 2980–2991 (2007)
12. Bonesky, T., Bredies, K., Lorenz, D., Mass, P.: A generalized conditional gradient method for nonlinear operator equations with sparsity constraints. Inverse Probl. **23**, 2041–2048 (2009)
13. Bresson, X., Chan, T.: Fast minimization of the vectorial total variation norm and applications to color image processing. UCLA CAM report, 07-25 (2007)
14. Brezis, H.: Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert. North Holland, Amsterdam (1973)
15. Brezis, H.: Analyse Fonctionnelle. Théorie et Applications. Mathématiques Appliquées pour la Maitrise. Masson, Paris (1983)
16. Chambolle, A.: An algorithm for total variation minimization and applications. J. Math. Imaging Vis. **20**, 89–97 (2004)
17. Chambolle, A.: Total variation minimization and a class of binary MRF models. In: EMMCVPR 05. Lecture Notes in Computer Sciences, vol. 3757, pp. 136–152. Springer, Berlin (2005)
18. Chambolle, A., Lions, P.L.: Image recovery via total variation minimization and related problems. Numer. Math. **76**(3), 167–188 (1997)
19. Chan, T., Mulet, P.: On the convergence of the lagged diffusity fixed point method in total variation image restoration. SIAM J. Numer. Anal. **36**(2), 354–367 (1999)
20. Chan, T., Shen, J.: Image Processing and Analysis—Variational, PDE, Wavelet, and Stochastic Methods. SIAM, Philadelphia (2005)
21. Chan, T., Golub, G., Mulet, P.: A nonlinear primal-dual method for total variation-based image restoration. SIAM J. Sci. Comput. **20**(6), 1964–1977 (1999)
22. Charbonnier, P., Blanc-Feraud, L., Aubert, G., Barlaud, M.: Deterministic edge-preserving regularization in computed imaging. IEEE Trans. Image Process. **6**(2) (2007)
23. Ciarlet, P.G.: Introduction à l'Analyse Numérique Matricielle et à l'Optimisation. Masson, Paris (1982)
24. Combettes, P.L., Pesquet, J.: Image restoration subject to a total variation constraint. IEEE Trans. Image Process. **13**(9), 1213–1222 (2004)
25. Combettes, P.L., Wajs, V.: Signal recovery by proximal forward-backward splitting. SIAM J. Multiscale Model. Simul. **4**(4), 1168–1200 (2005)
26. Darbon, J., Sigelle, M.: Image restoration with discrete constrained total variation part I: Fast and exact optimization. J. Math. Imaging Vis. **26**(3), 277–291 (2006)
27. Daubechies, I., Defrise, M., De Mol, C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Commun. Pure Appl. Math. **57**, 1413–1457 (2004)
28. Dobson, D., Vogel, C.: Convergence of an iterative method for total variation denoising. SIAM J. Numer. Anal. **34**, 1779–1791 (1997)
29. Durand, S., Fadili, J., Nikolova, M.: Multiplicative noise removal using L^1 fidelity on frame coefficients. CMLA Report, 08-40 (2008)
30. Duval, V., Aujol, J.-F., Vese, L.: Projected gradient based color image decomposition. In: SSVM 09 (2009)
31. Ekeland, I., Temam, R.: Analyse Convexe et Problèmes Variationnels, 2nd edn. Grundlehren der mathematischen Wissenschaften, vol. 224. Dunod, Paris (1983)
32. Facciolo, G., Almansa, A., Aujol, J.-F., Caselles V.: Irregular to regular sampling, denoising and deconvolution. SIAM J. Multiscale Model. Simul. (2009, in press)
33. Fadili, J., Starck, J.-L.: Monotone operator splitting for fast sparse solutions of inverse problems (2009, submitted)
34. Fu, H., Ng, M., Nikolova, M., Barlow, J.: Efficient minimization methods of mixed l1-l1 and l2-l1 norms for image restoration. SIAM J. Sci. Comput. **27**(6), 1881–1902 (2006)
35. Goldfarb, D., Yin, W.: Second-order cone programming methods for total variation based image restoration. SIAM J. Sci. Comput. **27**(2), 622–645 (2005)
36. Goldstein, T., Osher, S.: The split Bregman algorithm for L1 regularized problems. UCLA CAM Report, April 2008
37. Meyer Y.: Oscillating patterns in image processing and in some nonlinear evolution equations, March 2001. The Fifteenth Dean Jacquelines B. Lewis Memorial Lectures

38. Nesterov, Y.: Introductory Lectures on Convex Optimization: a Basic Course. Kluwer Academic, Dordrecht (2004)
39. Nesterov, Y.: Smooth minimization of non-smooth functions. Math. Program. (A) **103**(1), 127–152 (2005)
40. Nesterov, Y.: Gradient methods for minimizing composite objective function. Core discussion paper (2007)
41. Ng, M.K., Qi, L., Yang, Y.F., Huang, Y.: On semismooth Newton methods for total variation minimization. J. Math. Imaging Vis. **27**, 265–276 (2007)
42. Nikolova, M., Chan, R.: The equivalence of half-quadratic minimization and the gradient linearization iteration. IEEE Trans. Image Process. **16**(6), 1623–1627 (2007)
43. Pazy, A.: On the asymptotic behavior of iterates of nonexpansive mappings in Hilbert space. Isr. J. Math. **26**, 197–204 (1977)
44. Pazy, A.: Semigroups of Linear Operators and Applications to Partial Differential Equations. Applied Mathematical Sciences, vol. 44. Springer, New York (1983)
45. Polyak, B.: Introduction to Optimization. Translation Series in Mathematics and Engineering. Optimization Software, New York (2004)
46. Ramlau, R., Teschke, G.: A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints. Numer. Math. **104**, 177–203 (2006)
47. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**, 259–268 (1992)
48. Setzer, S.: Split Bregman Algorithm, Douglas-Rachford Splitting and frame shrinkage. Preprint, University of Mannheim (2008)
49. Weisfeld, E.: Sur le point pour lequel la somme des distances de points donnés est minimum. Tôhoku Math. J. **43**, 355–386 (1937)
50. Weiss, P.: Algorithmes rapides d'optimisation convexe. Applications à la restauration d'images et à la détection de changement. Ph.D. thesis, Université de Nice Sophia-Antipolis, December 2008
51. Weiss, P., Aubert, G., Blanc-Feraud, L.: Efficient schemes for total variation minimization under constraints in image processing. SIAM J. Sci. Comput. (2009, in press)
52. Yuan, J., Schnörr, C., Steidl, G.: Convex Hodge decomposition and regularization of image flows. J. Math. Imaging Vis. **33**(2), 169–177 (2009)
53. Zhu, M., Chan, T.F.: An efficient primal-dual hybrid gradient algorithm for total variation image restoration, May 2008. UCLA CAM Report 08-34
54. Zhu, M., Wright, S.J., Chan, T.F.: Duality-based algorithms for total variation image restoration, May 2008. UCLA CAM Report 08-33

Jean-François Aujol studied Mathematics at the Ecole Normale Supérieure in Cachan. He received his Ph.D. in Applied Mathematics from Nice-Sophia-An-tipolis University in June 2004. In 2004–2005, he was successively an assistant researcher in UCLA (Applied Mathematics Department), and then a research ingeneer at ENST (TSI department) in Paris. Since then, Jean-François Aujol is a scientist researcher with CNRS, at Centre de Mathématiques et de leurs Applications (CMLA, ENS Cachan, CNRS, UniverSud), France. He is interested in mathematical problems related to image processing.