Dual norms and image decomposition models

Jean-François Aujol [1,2] *& Antonin Chambolle [3] †

[1] Laboratoire J.A.Dieudonné, UMR CNRS 6621
[2] ARIANA, projet commun CNRS/INRIA/UNSA
[3] CMAP (UMR 7641), Ecole Polytechnique
91128 Palaiseau Cedex, France

Abstract

Following recent work by Y. Meyer, decomposition models into a geometrical component and a textured component have recently been proposed in image processing. In such approaches, negative Sobolev norms have seemed to be useful to modelize oscillating patterns. In this paper, we compare the properties of various norms that are dual of Sobolev or Besov norms. We then propose a decomposition model which splits an image into three components: a first one containing the structure of the image, a second one the texture of the image, and a third one the noise. Our decomposition model relies on the use of three different semi-norms: the total variation for the geometrical componant, a negative Sobolev norm for the texture, and a negative Besov norm for the noise. We illustrate our study with numerical examples.

Key-words: Total variation minimization, BV, texture, noise, negative Sobolev spaces, negative Besov spaces, image decomposition.

1 Motivation

1.1 Decomposition models

Decomposing an image into meaningful components is one of the major goals of image processing. A first range of models are denoising models: in such models, the image is assumed to have been corrupted by a noise, and what we want is to remove the noise. There are mainly two kinds of approaches. A first one is based on a process on the wavelet coefficients of the image [10, 15, 9, 14, 13, 17]. A second one is based on the resolution of the PDE associated to the minimization of an energy involving some norm of the gradient [19, 8, 17].

More recently, following [17], decomposition models into a geometrical component and a textured component have been proposed [21, 18, 4, 20]. The first three models rely on the use of a negative Sobolev norm to characterize textures. [20] uses appropriate dictionaries to characterize textures, and at the end even gets a decomposition of the image into three components: a first one containing the geometrical information of the image, a second one the texture, and a last one the noise. In this paper, we would like to study more deeply the characteristic of negative Sobolev norms, and then propose a decomposition model which would split an image into three components: a first one containing the structure of the image, a second one the texture of the image, and a third one the noise.

1.2 Some functional spaces

Although we will work only in the discrete setting in the rest of the paper, we first introduce various functional spaces that have been considered in image processing for describing piecewise smooth or oscillatory components in images. Throughout this subsection, we denote by Ω an open connected set of $\mathbb{R}^2$ with Lipschitz boundary.

The use of different functional spaces is a common tool in image processing, especially in minimization processes (see [3, 17] and references). For instance, the space BV of functions with bounded variation is often used. It enjoys the nice feature of containing functions with discontinuities along lines, that can represent the edges in an image. We recall here its definition [2].

Definition 1.1. $BV(\Omega)$ is the subspace of functions $u \in L^1(\Omega)$ such that the following quantity, called the

Total Variation of u, is finite:

$$J(u) = \sup\left\{\int_\Omega u(x)\operatorname{div}(\xi(x))dx \text{ such that }\right.$$
$$\left.\xi \in C_c^1(\Omega;\mathbb{R}^2), \|\xi\|_{L^\infty(\Omega)} \le 1\right\} \qquad (1.1)$$

The total variation plays a key-role in the Rudin-Osher-Fatemi (ROF) model [19]. Yves Meyer has recently investigated this model in [17]. Unsatisfied with the results, he has introduced a space G for oscillating patterns. This space happens to be very close to the dual space of BV. In such a space, oscillating patterns have a small norm: this a usefull property to catch them with an energy minimization process. We recall here the definition of G:

Definition 1.2. G is the Banach space composed of the distributions f which can be written $f = \partial_1 g_1 + \partial_2 g_2 = \operatorname{div}(g)$, with g_1 and g_2 in $L^\infty(\Omega)$. On G, the following norm is defined:

$$\|f\|_G = \inf\left\{\|g\|_{L^\infty(\Omega,\mathbb{R}^2)} \,/\, f = \operatorname{div}(g)\right\} \qquad (1.2)$$

with $\|g\|_{L^\infty(\Omega,\mathbb{R}^2)} = \operatorname{ess\,sup}_{x\in\Omega}\sqrt{|g_1|^2 + |g_2|^2}(x)$

In fact, with these notations, G is exactly $W^{-1,\infty}(\Omega)$, the dual space of $W_0^{1,1}(\Omega)$ [1]. And it is shown in the discrete case in [4] that $\|.\|_G$ is the polar semi-norm of $J(.)$ (see Proposition 2.1).

In [17], the author has proposed a new decomposition algorithm: it splits an original image f into a sum $u + v$ of a bounded variation component and a component containing the oscillating part of the image by minimizing: $J(u) + \|v\|_G$. But, due to the G norm, this decomposition is quite difficult to compute. Algorithms have been proposed to solve Meyer's problem or close ones [21, 4, 5]. These works have confirmed that G is a very good space to modelize oscillating patterns, and especially textures.

In [18], the G norm is replaced by another negative Sobolev norm. For the reader convenience, we recall here the definition of a Sobolev space(see [1]):

Definition 1.3. If $p \in (1,+\infty)$, then $W^{1,p}(\Omega)$ is the space of functions in $L^p(\Omega)$, and such that their distributional derivative also belongs to $L^p(\Omega)$. $W_0^{1,p}(\Omega)$ is the closure of $C_c^1(\Omega)$ in $W^{1,p}(\Omega)$. Let $q = \frac{p}{p-1}$ (so that $\frac{1}{p} + \frac{1}{q} = 1$). We denote by $W^{-1,q}(\Omega)$ the dual space of $W_0^{1,p}(\Omega)$.

Oscillating patterns have a small norm in a negative Sobolev space. And such a norm may be easier to minimize than the G norm.

In [17], the author also suggests the use of another dual space to model oscillating patterns: $E = \dot{B}_{-1,\infty}^\infty$. This is the dual space of $\dot{B}_{1,1}^1$.

Definition 1.4. $\dot{B}_{1,1}^1$ is the usual homogeneous Besov space (see [17]). Let $\psi_{j,k}$ an orthonormal base composed of smooth and compactly supported wavelets. $\dot{B}_{1,1}^1$ is a subspace of $L^2(\mathbb{R}^2)$, and a function f belongs to $\dot{B}_{1,1}^1$ if and only if: $\sum_{j\in\mathbb{Z}}\sum_{k\in\mathbb{Z}^2}|c_{j,k}|2^{j/2} < +\infty$, where $c_{j,k}$ are the wavelet coefficients of f.

Definition 1.5. The dual space of $\dot{B}_{1,1}^1$ is the Banach space $E = \dot{B}_{-1,\infty}^\infty$. It is characterized by the fact that the wavelet coefficients of a generalized function in $E = \dot{B}_{-1,\infty}^\infty$ belong to $l^\infty(\mathbb{Z}\times\mathbb{Z}^2)$.

In practice, we will use the Haar wavelet (resp. the Daubechie wavelet with eight or ten vanishing moments) denoted by *Haar* (resp. *Daub8* or *Daub10*) [16].

Remark We have the following result (stated in [17]):

$$\dot{B}_{1,1}^1 \subset \dot{BV} \subset L^2 \subset G \subset E = \dot{B}_{-1,\infty}^\infty \qquad (1.3)$$

where $\dot{BV}$ is the homogeneous version of BV: $\dot{BV} = BV/\{u \in BV \,/\, \nabla u = 0\}$. In fact, we will see that BV is well adapted for the geometrical part of an image, G for the texture part and $E = \dot{B}_{-1,\infty}^\infty$ for the noise.

In the rest of the paper, we will then introduce the discrete versions of the norms of these spaces

1.3 Overview of the paper

The aim of this work is to construct a model which decompose an image into three components, a first one containing the geometrical information, a second one containing the texture, and a last one containing the noise. To separate these components, we want to try to find appropriate norms that are well adapted to measure the relative importance of each components. As suggested in [17], we are going to use dual norms of the standard Sobolev norms to characterize the oscillating patterns.

In the first part of this paper, we introduce the discrete framework in which we will place ourself, and which is an appropriate one in image processing. In a second part, we compute dual norms used in [17, 21, 18, 4]. This will tells us which dual norms are appropriate to capture oscillating patterns. Then, in a third part, we introduce a mathematical framework to minimize dual norms. This will precise the study of the second part. In the fourth part, we propose a new $u + v$ model, which happens to be a powerful denoising model. And in the last part, we finally construct our $u + v + w$ model and show some numerical experiments.

2 Discretization

2.1 Notations

From now on, we consider the discrete case We take here the same notations as in [7]. 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 scalar product $(u,v)_X = \sum_{1 \leq i,j \leq N} u_{i,j} v_{i,j}$ and the norm $\|u\|_X = \sqrt{(u,u)_X}$. We will also use the following norms ($p \geq 1$):

$$\|u\|_p = \left(\sum_{1 \leq i,j \leq N} |u_{i,j}|^p \right)^{1/p} \qquad (2.4)$$

(remark that $\|u\|_X = \|u\|_2$).

2.1.1 Total variation

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: $(\nabla u)_{i,j} = ((\nabla u)_{i,j}^1, (\nabla u)_{i,j}^2)$. with

$$(\nabla u)_{i,j}^1 = \begin{cases} u_{i+1,j} - u_{i,j} & \text{if } i < N \\ 0 & \text{if } i = N \end{cases} \qquad (2.5)$$

and

$$(\nabla u)_{i,j}^2 = \begin{cases} u_{i,j+1} - u_{i,j} & \text{if } j < N \\ 0 & \text{if } j = N \end{cases} \qquad (2.6)$$

The discrete total variation of u is then defined by:

$$J(u) = \sum_{1 \leq i,j \leq N} |(\nabla u)_{i,j}| \qquad (2.7)$$

2.1.2 Sobolev semi-norms

We also use the following norms ($p \geq 1$):

$$\|\nabla u\|_p = \left(\sum_{1 \leq i,j \leq N} |\nabla u_{i,j}|^p \right)^{1/p} \qquad (2.8)$$

(remark that $\|\nabla u\|_Y = \|\nabla u\|_2$ and $J(u) = \|\nabla u\|_1$). We can now introduce the following semi-norm:

$$\|u\|_{1,p} = \|\nabla u\|_p \qquad (2.9)$$

This is a norm on the space $X_0 = \left\{ u \in X, \sum_{i,j} u_{i,j} = 0 \right\}$. It can be seen as a discrete version of a Sobolev norm.

We consider the polar semi-norm associated to (2.9) (with $1/p + 1/q = 1$):

$$\|v\|_{-1,p} = \sup_{\|u\|_{1,q}=1} (v,u)_X = \sup_{\|\nabla u\|_q=1} (v,u)_X \qquad (2.10)$$

This is the discrete norm we use for negative Sobolev spaces.

2.1.3 G-norm

We also introduce a discrete version of the divergence operator. We define it by analogy with the continuous setting by 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:

$$
\begin{aligned}
(\operatorname{div}(p))_{i,j} = & \begin{cases} p_{i,j}^1 - p_{i-1,j}^1 & \text{if } 1 < i < N \\ p_{i,j}^1 & \text{if i=1} \\ -p_{i-1,j}^1 & \text{if i=N} \end{cases} \qquad (2.11) \\
+ & \begin{cases} p_{i,j}^2 - p_{i,j-1}^2 & \text{if } 1 < j < N \\ p_{i,j}^2 & \text{if j=1} \\ -p_{i,j-1}^2 & \text{if j=N} \end{cases}
\end{aligned}
$$

From now on, we will use these discrete operators. We are now in position to introduce the discrete version of Meyer's space G.

Definition 2.6.

$$G = \{ v \in X \ / \ \exists g \in Y \text{ such that } v = \operatorname{div}(g) \} \qquad (2.12)$$

and if $v \in G$:

$$
\begin{aligned}
\|v\|_G = & \inf \{ \|g\|_\infty \ / \ v = \operatorname{div}(g), \qquad (2.13) \\
& g = (g^1, g^2) \in Y, |g_{i,j}| = \sqrt{(g_{i,j}^1)^2 + (g_{i,j}^2)^2} \}
\end{aligned}
$$

where $\|g\|_\infty = \max_{i,j} |g_{i,j}|$.

Moreover, we will denote:

$$B_{G(\mu)} = \mu B_G = \{ v \in G \ / \ \|v\|_G \leq \mu \} \qquad (2.14)$$

We recall that the Legendre-Fenchel transform of F is given by $F^*(v) = \sup_u (u,v)_X - F(u)$ (see [11]). The two following results are proved in [4].

Proposition 2.1. $J(u) = \sup_{v \in B_G} (u,v)_X$, and $\|v\|_G = \sup_{J(u) \leq 1} (u,v)_X$.

We see that $J(.)$ (resp. $\|.\|_G$) is the polar of $\|.\|_G$ (resp. $J(.)$).

Proposition 2.2. The space G identifies with the following subspace: $X_0 = \{ v \in X \ / \ \sum_{i,j} v_{i,j} = 0 \}$.

2.2 Total variation minimization as a projection

Introduction: Since J defined by (1.1) is homogeneous of degree one (i.e. $J(\lambda u) = \lambda J(u)$ $\forall u$ and $\lambda > 0$), it is then standard (see [11]) that J^* is the indicator function of some closed convex set, which turns out to be the set B_G defined by (2.14):

$$J^*(v) = \chi_{B_G}(v) = \begin{cases} 0 & \text{if } v \in B_G \\ +\infty & \text{otherwise} \end{cases} \qquad (2.15)$$

This can be checked out easily (see [7] for details). In [7], the author proposes a nonlinear projection algorithm to minimize the ROF model. The problem is:

$$\inf_{u \in X} \left(J(u) + \frac{1}{2\lambda} \|f - u\|_X^2 \right) \qquad (2.16)$$

We have the following result, which comes from standard convex duality theory [11]:

Proposition 2.3. *([7]): The solution of (2.16) is given by: $u = f - P_{\lambda B_G}(f)$. where P is the orthogonal projector on λB_G (defined by (2.14)).*

We use the following algorithm to compute $P_{\lambda B_G}(f)$. It indeed amounts to finding:

$$\min \left\{ \|\lambda \mathrm{div}\,(p) - f\|_X^2 : p \ / \ |p_{i,j}| \le 1 \ \forall i, j = 1, \ldots, N \right\} \qquad (2.17)$$

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

$$p_{i,j}^{n+1} = \frac{p_{i,j}^n + \tau(\nabla(\mathrm{div}\,(p^n) - f/\lambda))_{i,j}}{1 + \tau |(\nabla(\mathrm{div}\,(p^n) - f/\lambda))_{i,j}|} \qquad (2.18)$$

In [7] is given a sufficient condition ensuring the convergence of the algorithm: it is shown that as long as $\tau \le 1/8$, then $\lambda \mathrm{div}\,(p^n)$ converges to $P_{\lambda B_G}(f)$ as $n \to +\infty$.

3 Computing dual norms

Contrarily to the computation of the E norm which is easily obtained once the wavelet decomposition of the image has been performed, the other dual norms are less easy to compute.

3.1 Computing G norm

We use the projection algorithm (2.18).

Let f be an image with zero mean. We remark that if $\lambda \ge 0$, then: $f = P_{\lambda B_G}(f)$ if and only if $\|f\|_G \le \lambda$. Thus one just need to compute the norm of $f - P_{\lambda B_G}(f)$ to know if λ is larger (if the norm is zero) or smaller (if the norm is positive) than $\|f\|_G$. Thus we can find $\|f\|_G$ by dichotomy. Nevertheless, in our numerical tests, we have found that this method does not give a precise approximation of the G-norm of an image.

3.2 Application to a decomposition algorithm

3.2.1 Introduction

In [17], Meyer has introduced a new model to split a given image f into a sum $u + v$ of a bounded variation component and a component containing the oscillating part of the image. This model has been first successfully implemented by Vese and Osher [21]. A different approach has been proposed in [4, 5]. We will focus on this one, where the decomposition is computed by minimizing a convex functional which depends on the two variables u and v, alternatively in each variable. Each minimization is based on the projection algorithm (2.18).

The problem to solve is the following (f being the degraded image):

$$\inf_{(u,v) \in X \times \mu B_G} \left(J(u) + \frac{1}{2\lambda} \|f - u - v\|_X^2 \right) \qquad (3.19)$$

where $\mu B_G = \{v \in G/\|v\|_G \le \mu\}$, and $\|v\|_G$ is defined by (1.2), and $J(u)$ by (1.1).

The algorithm to solve this problem used in [4, 5] needs thus the two parameters λ and μ. The parameter λ controls the L^2-norm of the residual $f - u - v$. The smaller λ is, the smaller the L^2 norm of the residual $f - u - v$ is. The larger μ is, the more v contains information, and therefore the more u is averaged. In fact, the choice of λ is easy. One just needs to set it very small. For instance, in all the examples presented hereafter, we have chosen $\lambda = 0.1$, and found out a maximum norm for $f - u - v$ of about 0.5 (for values ranging from 0 to 255). But the μ parameter is much harder to tune. It controls the G norm of the oscillating component v. Some examples of the value of such a norm are displayed in Subsection 3.5.

3.2.2 Choice of the μ parameter

Assume the image u has been degraded by a gaussian noise of variance σ^2. We would like to remove that noise: hence we want the v component to be a white noise of variance σ^2.

In the ROF method [19], it is suggested to minimize the total variation with the constraint that the residual has exactly a variance of σ^2. If we want to do the same for the algorithm of [4], we have to know the G norm of a Gaussian norm of known standard deviation. Thanks to the homogeneity of the norm, $\|n\|_G/\sigma$ should be roughly constant (at least when the number of points is large enough, given the size of the image), however, the strong nonlinearity of the G norm does not allow us to check rigorously this fact (in other words, that this "constant" is quite insensitive to the particular realization of the Gaussian noise we consider).
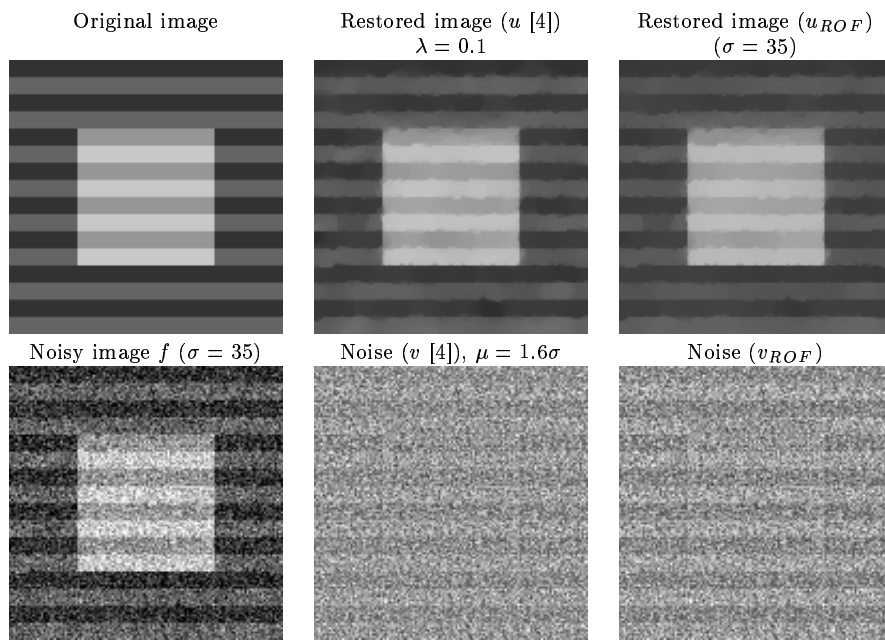
Figure 1: Denoising

We have computed the G norm of white gaussian noise images (of variance σ^2) with the algorithm explained in subsection 3.1. For instance, we have checked that the G norm of a white gaussian noise with variance $\sigma = 1$ is very close to 1.6 for an image of size 128*128. Knowing this, we can then compute the G norm of any white gaussian noise n of size 128*128,

provided we kow its variance σ^2:

$$\|n\|_G \sim 1.6\,\sigma \qquad (3.20)$$

This gives the value of the μ parameter for the algorithm of [4]. Figure 1 shows how this choice of the parameter μ performs on a particular case.

3.3 Computing $\|.\|_{-1,p}$ norms $(p \in (1, +\infty))$

Let $p \in (1, +\infty)$. Let f be a function with zero mean (i.e. $\sum_{1 \le i,j \le N} f_{i,j} = 0$). We want to compute $\|f\|_{-1,p}$ (see (2.10)). Let us set $q = \frac{p}{p-1}$ (so that $\frac{1}{p} + \frac{1}{q} = 1$). Let us recall that $\|u\|_{1,q} = \|\nabla u\|_q$ (see (2.9)) and:

$$\|f\|_{-1,p} = \sup_{\|u\|_{1,q}=1} (f,u)_X = \sup_{\|\nabla u\|_q=1} (f,u)_X \quad (3.21)$$

We look at the following minimization problem:

$$\min_{u \in X} (F(u) - (f,u)_X) \qquad (3.22)$$

where $F(u) = \frac{1}{q} \sum_{1 \le i,j \le N} |\nabla u_{i,j}|^q = \frac{1}{q}\|\nabla u\|_q^q$. By analogy with the continuous setting, we define a discrete Laplacian operator by setting $\Delta u = \mathrm{div}\,(\nabla u)$.

Proposition 3.4. *If $\tilde{u}$ is the minimizer of (3.22), then we have for $1 \le i,j \le N$:*

$$\left(\mathrm{div}\,(|\nabla \tilde{u}_{i,j}|^{q-2} \nabla \tilde{u}_{i,j})\right)_{i,j} = f_{i,j} \qquad (3.23)$$

Moreover, we also have:

$$\|f\|_{-1,p}^p = \|\nabla \tilde{u}\|_q^q = \sum_{1 \le i,j \le N} |\nabla \tilde{u}_{i,j}|^q = qF(\tilde{u}) \quad (3.24)$$

In particular, when $p = 2$, we get: $\Delta \tilde{u} = f$, and $\|f\|_{-1,2} = \|\nabla \tilde{u}\|_Y = \sqrt{\sum_{1 \le i,j \le N} |\nabla \tilde{u}_{i,j}|^2}$.

Proof. (3.23) is the Euler-Lagrange equation of (3.22), and its computation is standard once one has remarked that $(F'(u), h)_X = |\nabla u|^{q-2}(\nabla u, \nabla h)_X$.

We concentrate now on (3.24). It is well known that the Legendre-Fenchel transform of $\frac{1}{q}\|f\|_{1,q}^q$ is $\frac{1}{p}\|f\|_{-1,p}^p$ [11]. We therefore deduce that

$$-\frac{1}{p}\|f\|_{-1,p}^p = \min_u (F(u) - (f,u)_X) = F(\tilde{u}) - (f,\tilde{u})_X \qquad (3.25)$$

The solution $\tilde{u}$ of (3.22) satisfies $F(\tilde{u}) = \frac{1}{q}\sum_{1 \le i,j \le N} |\nabla \tilde{u}_{i,j}|^q = \frac{1}{q}\|\nabla \tilde{u}\|_q^q$, and by Euler's identity we have $(F'(\tilde{u}), \tilde{u})_X = qF(\tilde{u})$. We thus deduce that: $(f,\tilde{u})_X = qF(\tilde{u})$. From (3.25), we get: $-\frac{1}{p}\|f\|_{-1,p}^p = (1-q)F(\tilde{u}) = -\frac{q}{p}F(\tilde{u})$.

■

5

Table 1: A striking example (see Figure 2)

Images	TV	L^2	$\|.\|_{-1,2}$	G	E (Daub10)
textured image	1 000 000	9 500	33 000	360	749
geometric image	64 600	9 500	300 000	2000	355
Gaussian noise ($\sigma = 85$)	2 100 000	9 500	9 100	120	287

textured image   geometric image   Gaussian noise ($\sigma = 85$)
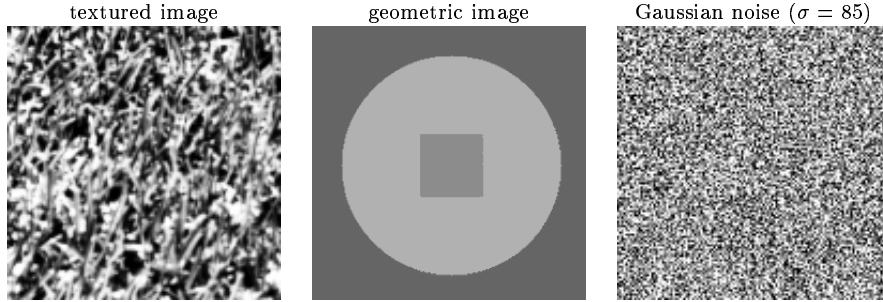
Figure 2: A striking example

3.4 Algorithm to compute $\|f\|_{-1,p}$

From Propositions 3.4, we see that a way to compute $\|f\|_{-1,p}$ is by minimizing (3.22) and by using the formula (3.24). When $p = 2$, then it is immediate to compute $\|f\|_{-1,2} = \sqrt{(-f, \Delta^{-1}f)_X}$ by using the discrete Fourier transform (see the end of subsection 3.6).

3.5 Numerical results

In Figure 2, the three images have the same L^2 norm. Table 1 presents the values of differents norms. It clearly illustrates the superiority of the G norm over the L^2 norm to capture oscillating patterns in minimization processes (the G norm is much smaller for the texture image and the noise image than the geometric image), as claimed in [17]. It also illustrates why the use of the E norm is well adapted to separate the noise (the noisy image has the smallest E norm).

As can be seen on Table 2, there appears a very strong correlation between the L^2 norm and the $\|.\|_{-1,2}$ norm of the image of a white gaussian noise. This will be confirmed by the results of the next subsection.

3.6 $\|.\|_{-1,2}$ norm of a white gaussian noise

We consider an image $n = (n_{i,j})$ of size $N \times N$. To simplify, we use periodic conditions in this subsection (but only in this subsection). We therefore not use (2.5), (2.6) and (2.11), but a discrete gradient and a discrete divergence with periodic conditions. We assume that n is an image of a white gaussian noise; i.e. for all (i, j), $n_{i,j}$ follows a gaussian law whose probability density function reads: $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-x^2}{2\sigma^2}\right)$ where σ^2 is the variance of the noise. We recall that if Z has a probability density function p_Z, than we

Table 2: Images of size 128*128 (see Figures 3, 4, 5)

Images	TV	L^2	G	$\|.\|_{-1,2}$	E (Daub10)
param1	1 000 000	9 500	360	33 000	730
param2	800 000	9 500	300	26 000	690
param3	980 000	9 450	320	30 000	890
param4	1 100 000	9 450	290	22 000	630
circles	18 000	3 000	2000	18 000	280
G20	570 000	2 500	31	2 300	73
G30	850 000	3 900	50.5	3 700	110
G50	1 400 000	6 300	84	6 100	180
G85	2 100 000	9 500	120	9 100	290

param1   param2

param3   param4
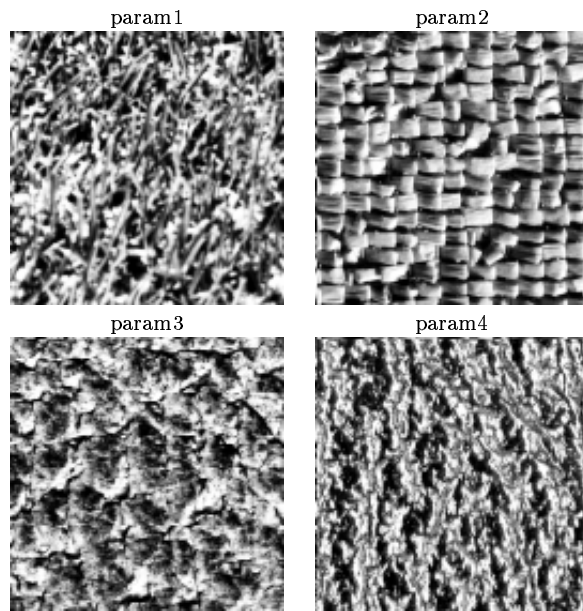
Figure 3: Textured images

circles

Figure 4: BV image

G20 ($\sigma = 20$)   G50 ($\sigma = 50$)   G85 ($\sigma = 85$)
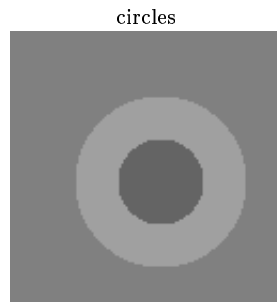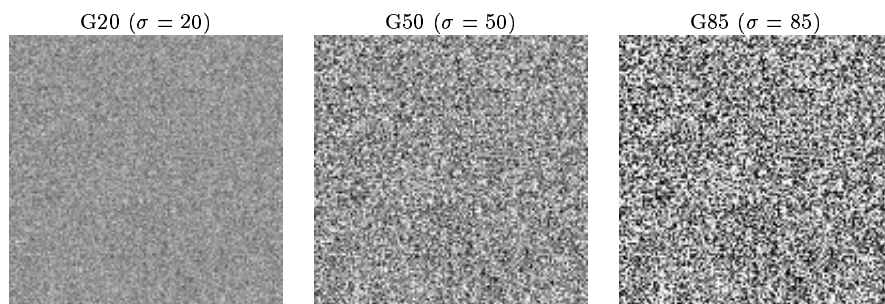
Figure 5: Gaussian noise images

denote its expectancy by $E(Z) = \int z p_Z(z)\, dz$.

Proposition 3.5.

$$E(\|n\|^2_{-1,2}) = C.E(\|n\|^2_X) \qquad (3.26)$$

where C is a constant which depends only on the size of the image.

This proposition gives a simple test to check if an image can be a white gaussian noise or not. This result can also be useful when tuning the parameters of a denoising algorithm using the $\|.\|_{-1,2}$ norm.

Proof. We split the proof into two steps:

Step 1: We begin by computing $E(\|n\|^2_{-1,2})$. We consider the functional:

$$\inf_u F(u) \qquad (3.27)$$

where $F(u) = \frac{1}{2}\|\nabla u\|^2_Y - (n,u)_X$. F is convex and lsc (lower semi continuous). Hence (see [12] for instance) there exists $\tilde{u}$ solving problem (3.27). As in Subsection 3.3, we get:

$$(\nabla \tilde{u}, \nabla h)_Y = (n,h)_X \;\; \forall h \text{ with zero mean.} \qquad (3.28)$$

We denote by δ the image such that $\delta_{0,0} = 1$, and $\delta_{i,j} = 0$ otherwise. We denote by W the solution of the problem: $\inf_u \int_\Omega \left(\frac{1}{2}|\nabla u|^2 - \delta u\right)$, and thus we have $(\nabla W, \nabla h)_Y = (\delta, h)_X$ for all h with zero mean. We introduce a discrete convolution: $f * g(x,y) = \sum_{i,j} f_{i,j} g_{x-i,y-j}$. It is easy to check that $\delta * f = f * \delta = f$ for all f (in fact, W is the Green function associated to problem (3.27)).

Let us now consider $u = n * W$; i.e. $u = \sum_{i,j} u_{i,j}$, with $u_{i,j}(x,y) = n_{i,j}W_{x-i,y-j}$. u is the solution of problem(3.27). If h has zero mean, then $(\nabla u, \nabla h)_Y = (n * \nabla W, \nabla h)_Y = (\nabla W, \bar{n} * \nabla h)_Y = (\delta, \bar{n} * h)_X = (n * \delta, h)_X = (n, h)_X$ (we have used the fact that $(n * a, b)_X = (a, \bar{n} * b)_X$ where $\bar{n}_{i,j} = -n_{i,j}$). From Proposition 3.4, we know that $\|n\|_{-1,2} = \|\nabla u\|_2$. We notice that $\nabla u = n * \nabla W$. Hence:
$\sum_{i,j}|\nabla u_{i,j}|^2 = (n * \nabla W, n * \nabla W)_Y = (n, n * W)_X = \sum_{k,l} n_{k,l} \; n * W(k,l)$, i.e. $\sum_{i,j}|\nabla u_{i,j}|^2 = \sum_{k,l} n_{k,l} \sum_{i,j} n_{i,j} W_{k-i,l-j} = \sum_{i,j,k,l} n_{i,j} n_{k,l} W_{k-i,l-j}$.
And thus:

$$E\left(\sum_{i,j}|\nabla u_{i,j}|^2\right) = E\left(\sum_{i,j,k,l} W_{k-i,l-j} n_{i,j} n_{k,l}\right)$$
$$= \sum_{i,j,k,l} W_{k-i,l-j} E\left(n_{i,j} n_{k,l}\right)$$

As the $n_{i,j}$ are independant, we have $E(n_{i,j} n_{k,l}) = E(n_{i,j}) E(n_{k,l})$ whenever $(i,j) \neq (k,l)$. Moreover, $E(n_{i,j}) = 0$, and $E(n_{i,j}^2) = \sigma^2$.

Hence: $E\left(\sum_{i,j}|\nabla u_{i,j}|^2\right) = \sigma^2 \sum_{i,j} W_{i-i,j-j}$. We thus get:

$$E\left(\sum_{i,j}|\nabla u_{i,j}|^2\right) = N^2 \sigma^2 W_{0,0} \qquad (3.29)$$

We can also observe that: $\sum_{i,j}|\nabla W_{i,j}|^2 = (\nabla W, \nabla W)_Y = (\delta, W)_X = W_{0,0}$. So we can rewrite (3.29): $E(\|\nabla u\|^2_Y) = N^2 \sigma^2 \|\nabla W\|^2_Y$ (and from Proposition 3.4 we know that $\|\nabla u\|_Y = \|n\|_{-1,2}$).

Step 2: We now want to compute $E(\|n\|^2_X)$. We have: $\|n\|^2_X = \sum_{i,j}(n_{i,j})^2$. Hence $E(\|n\|^2_X) = \sum_{i,j} E\left((n_{i,j})^2\right)$. And we get: $E(\|n\|^2_X) = N^2 \sigma^2$. We then get the result of Proposition 3.5 with $C = \|\nabla W\|^2_Y = W_{0,0} = \|\delta\|^2_{-1,2}$.

∎

Remark: Computation of the constant C of Proposition 3.5 Let us denote by m_δ the mean of δ, and by $\dot{\delta} = \delta - m_\delta$. From (3.28), we get that: $\Delta W = -\delta$. So formally, we can write $W = -\Delta^{-1}(\dot{\delta})$. We have shown that $C = \|\dot{\delta}\|^2_{-1,2} = \|\nabla W\|^2_Y$. We thus have $C = (\nabla \Delta^{-1}(\dot{\delta}), \nabla \Delta^{-1}(\dot{\delta}))_Y = (-\dot{\delta}, \Delta^{-1}(\dot{\delta}))_X$. There remains to compute $\Delta^{-1}(\dot{\delta})$. This can easily be done with the Discrete Fourier Transform (DFT). We recall that the DFT of a given discrete image $(f(m,n))$ ($0 \leq m \leq N-1$ and $0 \leq n \leq N-1$) is given by ($0 \leq p \leq N-1$ and $0 \leq q \leq N-1$):

$$\mathcal{F}(f)(p,q) = F(p,q) = \qquad (3.30)$$
$$\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} f(m,n) e^{-j(2\pi/N)pm} e^{-j(2\pi/N)qn}$$

and the inverse transform is:

$$f(m,n) = \frac{1}{N^2}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1} F(p,q) e^{j(2\pi/N)pm} e^{j(2\pi/N)qn} \qquad (3.31)$$

Moreover, we also have $\|\mathcal{F}(f)\|^2_X = N^2\|f\|^2_X$ and $(\mathcal{F}(f), \mathcal{F}(g))_X = N^2(f,g)_X$.

Let us remark that $\mathcal{F}(\delta)(p,q) = 1 \;\forall p$, and that $\mathcal{F}(\dot{\delta})(p,q) = 0$ if $p = q = 0$ and 1 otherwise.

We have $\Delta f(m,n) = f(m+1,n) + f(m-1,n) + f(m,n+1) + f(m,n-1) - 4f(m,n)$. Standard computations lead to:

$$\mathcal{F}(\Delta f)(p,q) = \qquad (3.32)$$
$$2\left(\cos\left(\tfrac{2\pi}{N}p\right) + \cos\left(\tfrac{2\pi}{N}q\right) - 2\right)\mathcal{F}(f)(p,q)$$

We deduce that, if f has zero mean, then for $(p,q) \neq (0,0)$

$$\mathcal{F}(\Delta^{-1}f)(p,q) = \qquad (3.33)$$
$$\frac{1}{2\left(\cos\left(\frac{2\pi}{N}p\right) + \cos\left(\frac{2\pi}{N}q\right) - 2\right)}\mathcal{F}(f)(p,q)$$

8

Table 3: Values of $\sqrt{C} = \|\dot{\delta}\|_{-1,2} = \|\nabla W\|_X$

Image size	16^2	32^2	64^2	128^2	256^2	512^2	1024^2	2048^2	4096^2	8192^2	16384^2
$\sqrt{C}$	0.69	0.77	0.84	0.91	0.97	1.01	1.07	1.12	1.17	1.22	1.26

Table 4: Numerical computations with 1000 images of white gaussian noise n with standard deviation equal to 1 (for the X norm and the $\|.\|_{-1,2}$ norm)

Image size	16^2	32^2	64^2	128^2	256^2	512^2
$E(\|n\|_X)$	16.0	32.0	64.0	128	256	512
$\sqrt{Var(\|n\|_X)}$	0.71	0.69	0.72	0.73	0.73	0.70
$E(\|n\|_{-1,2})$	11.2	24.7	54	116	247	523
$\sqrt{Var(\|n\|_{-1,2})}$	0.97	1.8	3.3	6.0	12	22
$\frac{E(\|n\|_X)}{E(\|n\|_{-1,2})}$	0.70	0.77	0.84	0.91	0.96	1.02

Hence $\mathcal{F}(\Delta^{-1}\dot{\delta})(0,0) = 0$ and if $(p,q) \neq (0,0)$:

$$\mathcal{F}(\Delta^{-1}\dot{\delta})(p,q) = \frac{1}{2\left(\cos\left(\frac{2\pi}{N}p\right) + \cos\left(\frac{2\pi}{N}q\right) - 2\right)} \tag{3.34}$$

so that

$$C = -(\dot{\delta}, \Delta^{-1}\dot{\delta})_X =$$

$$\frac{1}{N^2} \sum_{(p,q)\neq(0,0)} \frac{1}{2\left(2 - \cos\left(\frac{2\pi}{N}p\right) + \cos\left(\frac{2\pi}{N}q\right)\right)} \tag{3.35}$$

We have computed this way the constant C for different image sizes (see Table 3).

Table 4 confirms the results of Table 3. To fill in Table 4, for each size (from 16*16 to 512*512), we have simulated 1000 images of white Gaussian noise with standard deviation σ equal to 1.0 (to normalize our experiment). Then we have computed the X norm and the $\|.\|_{-1,2}$ norm of each image. We have computed the experimental mean and and also the experimental standard deviation. The values of $\frac{E(\|n\|_X)}{E(\|n\|_{-1,2})}$ of Table 4 fit the values of $\sqrt{C}$ of Table 3, as expected from Proposition 3.5. But the experimental standard deviations of these norms do not have the same behaviour: the standard deviation of the X remains almost constant, whereas the standard deviation of the $\|.\|_{-1,2}$ norm seems to be an increasing function of the size of the image. Nevertheless, the standard deviation of the $\|.\|_{-1,2}$ norm seems to increase slower then its mean, and as soon as the size of the image is large enough, the standard deviation remains small with repect to the mean (which can therefore be considered as a quite good estimate of the $\|.\|_{-1,2}$ norm).

Table 5 shows the results we get for the same experiment with the TV and the E norm (we recall that $J(f) = \|f\|_{TV}$).

3.7 An algorithm to solve the Osher-Solé-Vese model

We come back here to the model proposed by Osher-Solé-Vese in [18]. We recall that the considered problem is:

$$\inf_u \left(J(u) + \frac{1}{2\lambda}\|f - u\|_{-1,2}^2 \right) \tag{3.36}$$

The following result holds:

Proposition 3.6. *If $\hat{u}$ is a solution of problem (3.36), then $\hat{v} = \Delta^{-1}\left(\frac{\hat{u}-f}{\lambda}\right)$ is a solution of the dual problem:*

$$\inf_v \left(\frac{\lambda}{2}\|\nabla v\|_Y^2 - (f,v)_X + J^*(v) \right) \tag{3.37}$$

Proof. There are many ways to show this (see [11]). For instance, one may reproduce the proof of Proposition 2.3 given in [7]. Thanks to Proposition 3.4, we

Table 5: Numerical computations with 1000 images of white gaussian noise n with standard deviation equal to 1 (for the TV norm and the E norm)

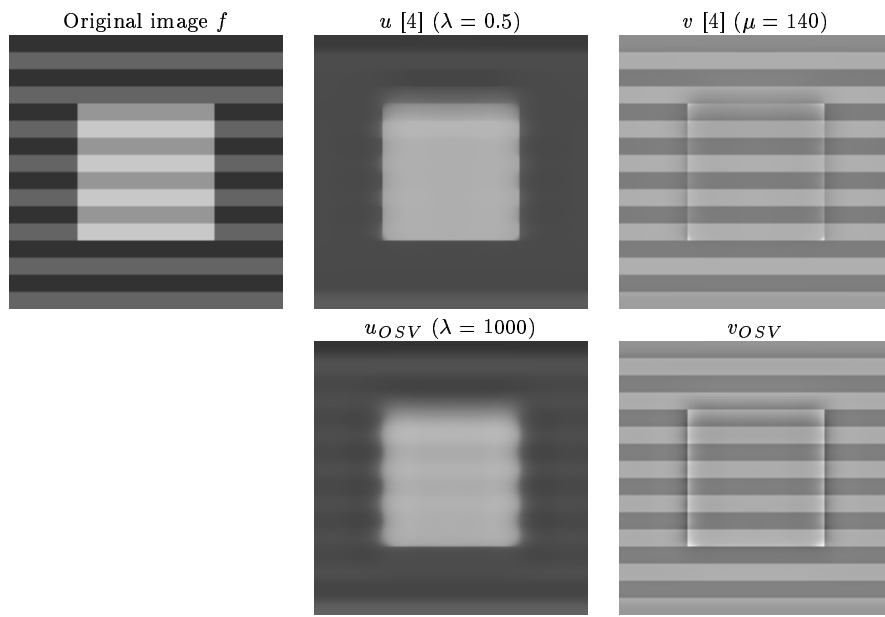| Image size	16^2	32^2	64^2	128^2	256^2	512^2
$E(J(n))$	392	1 670	6 920	28 100	113 000	455 000
$\sqrt{Var(J(n))}$	20	43	90	180	360	690
$E(\|n\|_E)$	2.81	3.59	3.63	3.98	4.30	4.59
$\sqrt{Var(\|n\|_E)}$	0.40	0.34	0.32	0.32	0.28	0.26

9

Figure 6: Decomposition (the parameters are tuned so that both v_{OSV} and the v component got with the algorithm of [4] have the same L^2 norm)

have $\|f-u\|^2_{-1,2} = (-f+u, \Delta^{-1}(f-u))_X$. If u is a minimiser of (3.36), then: $0 \in -\Delta^{-1}\left(\frac{u-f}{\lambda}\right) + \partial J(u)$, i.e.: $\Delta^{-1}\left(\frac{u-f}{\lambda}\right) \in \partial J(u)$. Hence $u \in \partial J^*\left(\Delta^{-1}\left(\frac{u-f}{\lambda}\right)\right)$, i.e.: $0 \in \frac{f-u}{\lambda} - \frac{f}{\lambda} + \frac{1}{\lambda}\partial J^*\left(\Delta^{-1}\left(\frac{u-f}{\lambda}\right)\right)$. We then set $v = \Delta^{-1}\left(\frac{u-f}{\lambda}\right)$, and we get: $0 \in -\Delta v - \frac{f}{\lambda} + \frac{1}{\lambda}\partial J^*(v)$. v is thus a solution of problem (3.37).

∎

Algorithm Thanks to Proposition 3.6, we have $\hat{u} = f + \lambda\Delta\hat{v}$. Hence, to compute $\hat{u}$, we just need to compute $\hat{v}$ solution of problem (3.37). As J^* is given by (2.15), $\hat{v}$ is the solution of the following problem:

$$\min\left\{\frac{\lambda}{2}\|\nabla\text{div}\,(p)\|^2_Y + (\nabla f, p)_Y : p \text{ such that}\right.$$
$$\left.|p_{i,j}| \le 1 \,\forall i, j = 1, \ldots, N\right\} \quad (3.38)$$

By using the same arguments as in [7], one sees that this problem can be numerically solved by a fixed point method: $p^0 = 0$, and

$$p_{i,j}^{n+1} = \frac{p_{i,j}^n - \tau(\nabla(\Delta\text{div}\,(p^n) + f/\lambda))_{i,j}}{1 + \tau|(\nabla(\Delta\text{div}\,(p^n) + f/\lambda))_{i,j}|} \quad (3.39)$$

And as in [7], we show that if τ is small enough, $f + \lambda\Delta\text{div}\,p^n \to \hat{u}$ as $n \to \infty$.

Numerical results Figure 7 is to be compared with Figure 1. We can see that as a denoising algorithm, the OSV model seems to outperforms both the ROF model and the algorithm of [4]. Figure 6 shows a comparison between the OSV model and the algorithm of [4]. As a decomposition algorithm (that is as an algorithm which splits a given image into a geometrical component and an oscillatory component), the algorithm of [4] seems to outperform the OSV model.

4 Minimization of dual norms

4.1 Some convex analysis

Proposition 4.7. *Let F be a proper lower semi-continuous convex function. Then the two following statements are equivalent:*

1. $\tilde{u}$ is a solution of

$$\min_u\left(F(u) + \frac{1}{2\lambda}\|f - u\|^2_X\right) \quad (4.40)$$

2. $\tilde{w} = f - \tilde{u}$ is a solution of

$$\min_w\left(F^*(w/\lambda) + \frac{1}{2\lambda}\|f - w\|^2_X\right) \quad (4.41)$$

Moreover, if F is 1 homogeneous (this is the case when F is a semi norm), then F^* is the indicator function of a closed convex set K.
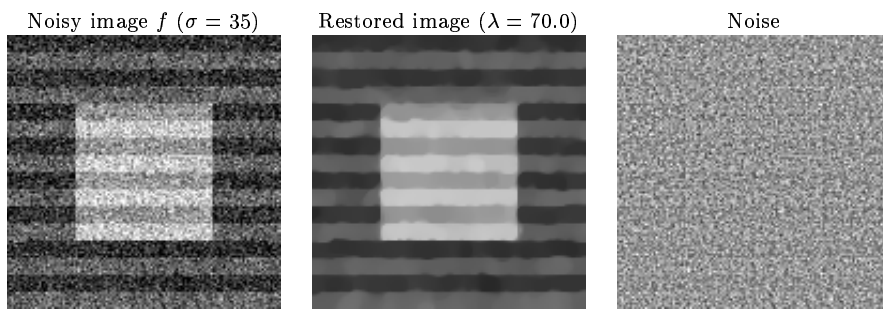
10

Figure 7: Denoising stripes with the Osher-Sole-Vese model

Examples

1. If $F(u) = J(u)$ (defined by (1.1)), then
   $F^*(w) = \chi_{\{v \ / \ \|v\|_G \leq 1\}}(w)$.

2. If $F(u) = \|u\|_G$ (defined by (1.2)), then
   $F^*(w) = \chi_{\{v \ / \ J(v) \leq 1\}}(w)$.

3. If $F(u) = \|u\|_{-1,p}$, then
$$F^*(w) = \chi_{\{v \ / \ \|\nabla v\|_q \leq 1\}}(w) \qquad (4.42)$$

4. If $F(u) = \|u\|_{\dot{B}^1_{1,1}}$, then
   $F^*(w) = \chi_{\{v \ / \ \|v\|_{\dot{B}^{\infty}_{-1,\infty}} \leq 1\}}(w)$.

Proof. *(of Proposition 4.7)*. This is a standard duality result for variational problems [11]. $\tilde{u}$ is a solution of (4.40) if and only if $0 \in (1/\lambda)(\tilde{u} - f) + \partial F(\tilde{u})$, i.e if and only if $(1/\lambda)(f - \tilde{u}) \in \partial F(\tilde{u})$, which is equivalent to $\tilde{u} \in \partial F^*\left(\frac{f-\tilde{u}}{\lambda}\right)$, i.e $0 \in f - \tilde{u} - f + \partial F^*\left(\frac{f-\tilde{u}}{\lambda}\right)$, which precisely means that $\tilde{w} = f - \tilde{u}$ is a solution of (4.41).

∎

4.2 Application to $u + v$ models

Let us now consider a functional of the form:
$$\min_{u,v} \left(F(u) + \frac{1}{2\lambda}\|f - u - v\|_X^2 + L(v) \right) \qquad (4.43)$$

where F and L are two convex functions. A natural algorithm to minimize such a functional consists in minimizing succesively with respect to each variable, that is to solve sucessively the two following problems:

$$\min_u \left(F(u) + \frac{1}{2}\|f - u - v\|_X^2 \right) \qquad (4.44)$$

$$\min_v \left(\frac{1}{2}\|f - u - v\|_X^2 + L(v) \right) \qquad (4.45)$$

Let us assume that $F(u) = J(u)$ defined by (1.1). We will consider several possibilities for the function L:
$$L(v) = \mu\|v\|_{-1,p} \qquad (4.46)$$

Comments: Using Proposition 4.7, one sees that $\tilde{v}$ is a solution of problem (4.45) if and only if $\tilde{w} = f - u - \tilde{v}$ is a solution of the problem:

$$\min_w \left(L^*(w) + \frac{1}{2}\|f - u - w\|_X^2 \right) \qquad (4.47)$$

- If $p = 2$ in (4.46), then using (4.42) we see that the minimization of (4.45) amounts to a Tikhonov regularization with H^1 penalization (and one knows that such a regularization is not a good choice in image processing since it does not preserve edges).

- If $p = 1$ in (4.46), then using (4.42) we see that the minimization of (4.45) amounts to a Lipschitz regularization of the gradient: this is not a good choice in image processing either.

- More generally, if $p \in [1, +\infty)$, we see that such a minimization smoothes too much the geometrical part of the image.

Conclusion If one wants to minimize a functional like (4.43) to capture oscillating patterns, then such considerations are in favor of the use of the G norm, as advocated in [17]. Even if such a norm is much more complicated to compute (especially comparing with the $\|.\|_{-1,2}$ norm), it seems worth to use it.

5 A new $u + v$ algorithm

5.1 Principle

In [17], the author proposes to minimize the following functional:

$$\inf_{u+v=f} \left(J(u) + \beta\|v\|_E \right) \qquad (5.48)$$

11

Due to the $E = \dot{B}^{\infty}_{-1,\infty}$ norm, this problem cannot be solved directly. We propose instead to minimize the following functional:

$$\inf_{(u,v)\in X^2}\left(J(u) + B^*(v/\delta) + \frac{1}{2\lambda}\|f - u - v\|^2_X\right)$$

(5.49)

where $B(v) = \|v\|_{\dot{B}^1_{1,1}}$ and therefore $B^*(w/\delta) = \chi_{\{\|w\|_E \leq \delta\}}$. Furthermore, let us denote by

$$\delta B_E = B_{E(\delta)} = \{w \ / \ \|w\|_E \leq \delta\}$$

(5.50)

When λ goes to 0, one can show that solving problem (5.49) gives a solution for (5.48) (using the same arguments as in [4]). To minimize (5.49), we consider the two following problems:

- v being fixed, we search for u as a solution of:

$$\inf_{u\in X}\left(J(u) + \frac{1}{2\lambda}\|f - u - v\|^2_X\right)$$

(5.51)

- u being fixed, we search for v as a solution of:

$$\inf_{v\in\delta B_E}\|f - u - v\|^2_X$$

(5.52)

From Proposition 2.3, we know that the solution of (5.51) is given by:
$\hat{u} = f - v - P_{\lambda B_G}(f - v)$. And the solution of (5.52) is given by: $\hat{v} = P_{\delta B_E}(f - u)$.

To compute this last projection, we consider the dual problem of (5.52):

$$\inf_{w\in X}\left(\frac{1}{2}\|f - u - w\|^2_X + \delta\|w\|_{\dot{B}^1_{1,1}}\right)$$

(5.53)

The solution of this problem $\hat{w}$ is obtained by a wavelet soft-thresholding algorithm with threshold δ (see [9]). We write $\hat{w} = WST(f - u, \delta)$. And from proposition 4.7, we see that $\hat{v} = f - u - \hat{w}$.

5.1.1 Algorithm

1. Initialization:

$$u_0 = v_0 = 0$$

(5.54)

2. Iterations:

$$v_{n+1} = P_{\delta B_E}(f - u_n) = f - u_n - WST(f - u_n, \delta)$$

(5.55)

$$u_{n+1} = f - v_{n+1} - P_{\lambda B_G}(f - v_{n+1})$$

(5.56)

3. Stopping test: we stop if

$$\max(|u_{n+1} - u_n|, |v_{n+1} - v_n|) \leq \epsilon$$

(5.57)

5.2 Mathematical analysis

We follow here [4]. The proofs of the results stated in this subsection are the same as in [4]. The first point we state is the convergence of the proposed algorithm.

Proposition 5.8. *The sequence (u_n, v_n) built in (5.54)–(5.57) converges to $(u_0, v_0) \in X \times X$ the unique solution of problem (5.49).*

To see the link between (5.48) and (5.49), we introduce the following problem:

$$\inf_{u+v=f}(J(u) + B^*(v/\delta))$$

(5.58)

The next result states the link between (5.48) and (5.58)

Proposition 5.9. *Let us fix $\beta > 0$ in problem (5.48). Let $(\hat{u}, \hat{v})$ a solution of problem (5.48). We fix $\delta = \|\hat{v}\|_E$ in (5.58). Then:*

- *$(\hat{u}, \hat{v})$ is also a solution of problem (5.58).*

- *Conversely, any solution $(\tilde{u}, \tilde{v})$ of (5.58) (with $\delta = \|\hat{v}\|_E$) is a solution of (5.48).*

To close the link between (5.48) and (5.49), we have to check what happens when λ goes to 0 in problem (5.49).

Proposition 5.10. *Let us fix $\beta > 0$ in (5.48). Set $\delta = \|\hat{v}\|_E$ in (5.49) and (5.58). Let us denote by $(u_{\lambda_n}, v_{\lambda_n})$ the solution of problem (5.49) (with $\lambda = \lambda_n$). Then, when λ_n goes to 0, any cluster point of $(u_{\lambda_n}, v_{\lambda_n})$ is a solution of problem (5.58).*

5.3 Numerical results and comments

Choice of the parameters The parameter λ controls the L^2 norm of the residual $f - u - v$. The smaller it is, the smaller the norm of the residual becomes. In practice, we just need to set λ small (in our experiments, we have used $\lambda \leq 1$: the L^∞ norm of the residual is then smaller than 1.5 (and the gray values of our images go from 0 to 255)).

The parameter δ represents the E norm of the v component. Having a closer look at the algorithm, we see that δ is the threshold which is used in the wavelet thresholding. In the case of denoising, we can then use the classical threshold $\sigma\sqrt{2\log(N^2)}$ where N^2 is the size of the image [10, 14, 17]. We therefore choose $\delta = \sigma\sqrt{2\log(N^2)}$. In fact, to get better results, we have introduced a weighting parameter η, and we set $\delta = \eta\sigma\sqrt{2\log(N^2)}$. In practice, η is equal to 1 or slightly smaller.
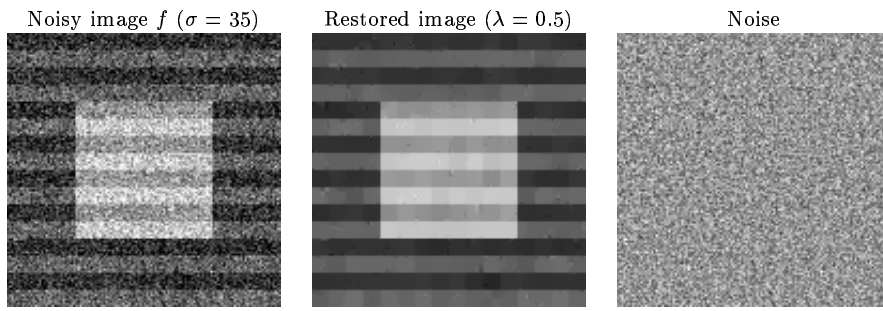
Figure 8: Denoising stripes (algorithm (5.54)–(5.57))

Comments This new algorithm appears to be a very good denoising algorithm for textured images. Indeed, it simultaneously minimizes the total variation of the restored image and a Besov norm (which amounts to a wavelet shrinkage). It benefits from the advantages of both methods. One sees on Figure 9 that the restored image has been more regularized with this new model than with the classical wavelet thresholding. Figure 8 is to be compared with Figure 1 and Figure 7. In this case, our new denoising algorithm clearly outperforms the ROF model [19] or the algorithm of [4]. And it also appears slightly better than the OSV model [18]. The result of Figure 8 is particularly impressive because the pattern and stripes are along the central directions of the main wavelet directions (see also Figures 9, 10 and 12 for comparison).

And on Figure 10, one sees that the small stucture of textures are better preserved than with the algo-rithm of [4]. This is why this model is particularly well suited to denoise textured images.

In [14, 13], the author also proposes to simulta-neously combine the advantage of the total variation minimization and wavelet thresholding. He introduces a general framework which contains both these two ap-proaches. Moreover, he can deal with other dictionar-ies than wavelet bases. But the numerical approach is completely different. He solves a penalization problem which is more complicated.

Therefore, to construct a model which would de-compose an image into three components, a first one containing the geometrical information of the image, a second one containing the texture, and a last one containing the noise, it is then natural to consider the three following norms: the total variation, the G norm, and the E norm.

6 $u + v + w$ model

6.1 Presentation

We propose to minimize the following functional:

$$\inf_{(u,v,w)\in X^3} F(u,v,w) \qquad (6.59)$$

where

$$F(u,v,w) = \qquad (6.60)$$
$$J(u) + J^*\left(\frac{v}{\mu}\right) + B^*\left(\frac{w}{\delta}\right) + \frac{1}{2\lambda}\|f - u - v - w\|_X^2$$

Proposition 6.11. *(6.59) admits a solution $(\hat{u}, \hat{v}, \hat{w})$. Moreover, $\hat{u}$ and $\hat{v} + \hat{w}$ are unique.*

Proof. We split the proof into two steps:
Step 1: Existence
It is standard to check that F is convex and coer-cive on $X \times X \times X$. Hence we deduce the existence of a minimizer $(\hat{u}, \hat{v}, \hat{w})$ (see [6] for further details)
Step 2: Uniqueness of $\hat{u}$ and $\hat{v} + \hat{w}$:
To get this result, we first remark that F is strictly convex on $X \times \mu B_G \times \delta B_E$, as the sum of a convex function and of a strictly convex function, except in the direction $(u, v, -u - v)$. Hence it suffices to check that if $(\hat{u}, \hat{v}, \hat{w})$ is a minimizer of F then for $t \neq 0$, $(\hat{u}+t\hat{u}, \hat{v}+t\hat{v}, \hat{w}-t(\hat{u}+\hat{v}))$ is not a minimizer of F. And since F is convex, we can assume $|t| < 1$. The result is obvious if $\hat{v} + t\hat{v} \in X\backslash\mu B_G$ or $\hat{w} - t(\hat{u} + \hat{v}) \in X\backslash\delta B_E$. Let us show that if $\hat{v}+t\hat{v} \in \mu B_G$ and $\hat{w}-t(\hat{u}+\hat{v}) \in \delta B_E$, then the result is still true. Indeed, we have in this case (since $|t| < 1$):

$$F(\hat{u} + t\hat{u}, \hat{v} + t\hat{v}, \hat{w} - t(\hat{u} + \hat{v})) = F(\hat{u}, \hat{v}, \hat{w}) + tJ(\hat{u})$$
$$(6.61)$$
By contradiction, let us assume that there exists $\hat{t} \neq 0$ such that $\hat{w} - \hat{t}(\hat{u} + \hat{v}) \in \delta B_E$.

$$F(\hat{u} + \hat{t}\hat{u}, \hat{v} + \hat{t}\hat{v}, \hat{w} - \hat{t}(\hat{u} + \hat{v})) \leq F(\hat{u}, \hat{v}, \hat{w}) \quad (6.62)$$

As $(\hat{u}, \hat{v}, \hat{w})$ minimizes F, (6.62) is an equality. From (6.61), we deduce that $J(\hat{u}) = 0$ (since $\hat{t} \neq 0$). There exists therefore $\gamma \in \mathbb{R}$ such that for all (i, j), $\hat{u}_{i,j} = \gamma$.

1. If $\gamma = 0$, then $\hat{u} = 0$. Thus $(\hat{u} + \hat{t}\hat{u}, \hat{v} + \hat{t}\hat{v}, \hat{w} - \hat{t}(\hat{u} + \hat{v})) = (\hat{u}, \hat{v} + \hat{t}\hat{v}, \hat{w} - \hat{t}\hat{v})$.

2. If $\gamma \neq 0$, then $\hat{w} - \hat{t}(\hat{u} + \hat{v})$ cannot belong to
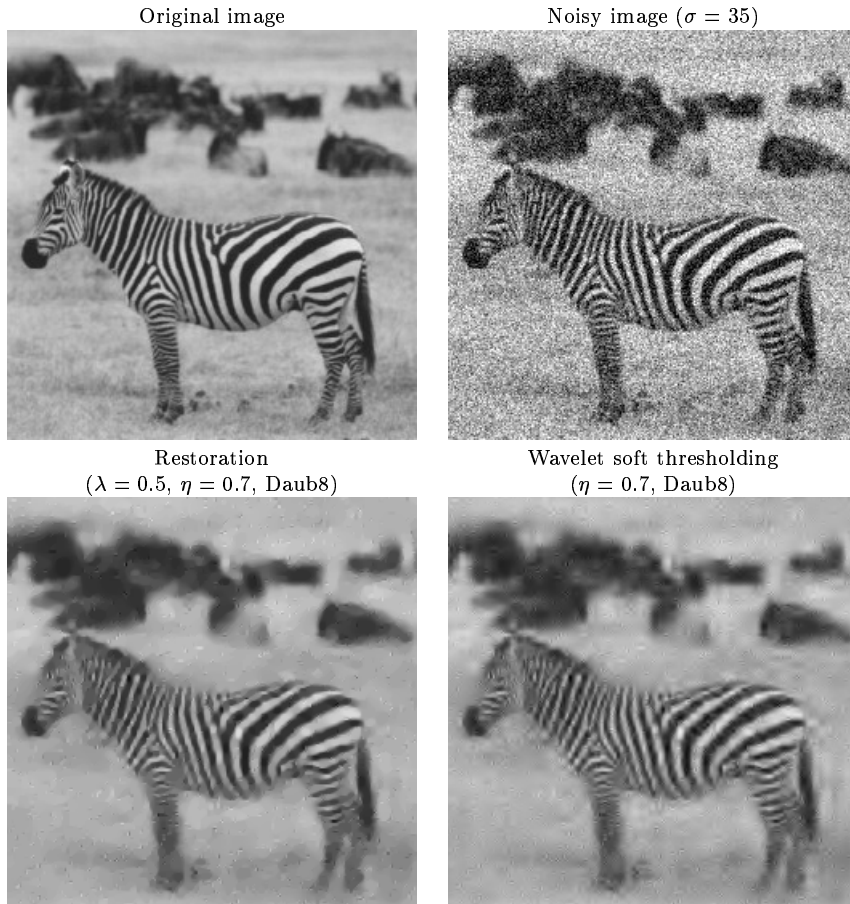
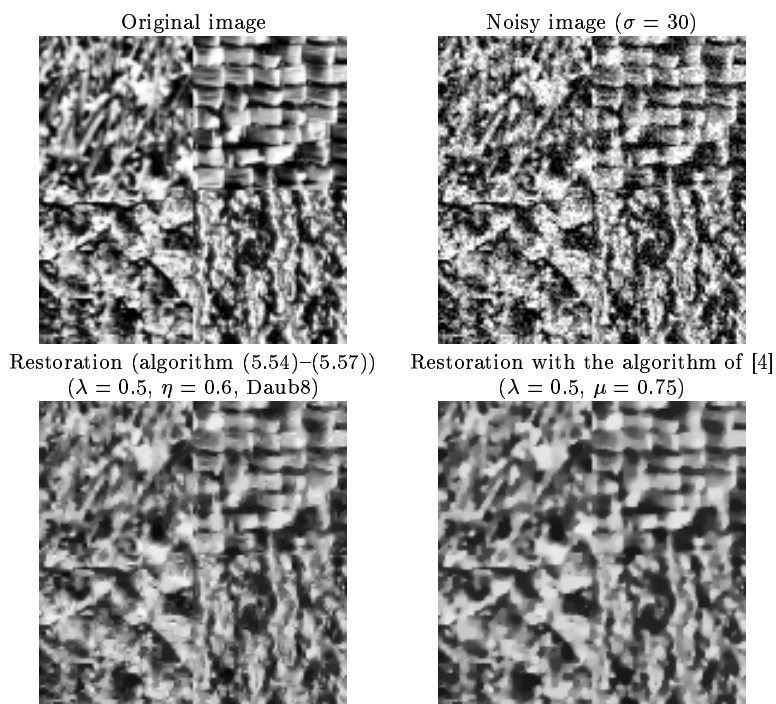Figure 9: Denoising a zebra (left: algorithm (5.54)–(5.57) , and right: wavelet soft thresholding)

Figure 10: Texture image

δB_E since its mean is not 0. This contradicts our assumption.

To solve (6.59), we consider the three following problems:

- v and w being fixed, we search for u as a solution of:

$$\inf_{u \in X} \left(J(u) + \frac{1}{2\lambda} \|f - u - v - w\|_X^2 \right) \qquad (6.63)$$

- u and w being fixed, we search for v as a solution of:

$$\inf_{v \in \mu B_G} \|f - u - v - w\|_X^2 \qquad (6.64)$$

- u and v being fixed, we search for w as a solution of:

$$\inf_{w \in \delta B_E} \|f - u - v - w\|_X^2 \qquad (6.65)$$

From Proposition 2.3, we know that the solution of (6.63) is given by:
$\hat{u} = f - v - w - P_{\lambda B_G}(f - v - w)$. The solution of (6.64) is simply given by: $\hat{v} = P_{\mu B_G}(f - u - w)$. And remembering Section 5, the solution of (6.65) is given by: $\hat{w} = P_{\delta B_E}(f - u - v) = f - u - v - WST(f - u - v, \delta)$, where $WST(f - u - v, \delta)$ stands for the wavelet soft-thresholding of $f - u - v$ with threshold δ.

6.2 Algorithm

1. Initialization:

$$u_0 = v_0 = w_0 = 0 \qquad (6.66)$$

2. Iterations:

$$\begin{aligned}
w_{n+1} &= P_{\delta B_E}(f - u_n - v_n) \qquad (6.67) \\
&= f - u_n - v_n - WST(f - u_n - v_n, \delta)
\end{aligned}$$

$$v_{n+1} = P_{\mu B_G}(f - u_n - w_{n+1}) \qquad (6.68)$$

$$u_{n+1} = f - v_{n+1} - w_{n+1} - P_{\lambda B_G}(f - v_{n+1} - w_{n+1}) \qquad (6.69)$$

3. Stopping test: we stop if

$$\max(|u_{n+1} - u_n|, |v_{n+1} - v_n|, |w_{n+1} - w_n|) \le \epsilon \qquad (6.70)$$

Proposition 6.12. *The sequence (u_n, v_n, w_n) is such that u_n converges to $\hat{u} \in X$, and if $(\hat{v}, \hat{w})$ is a cluster point of (v_n, w_n) then $(\hat{u}, \hat{v}, \hat{w})$ is a minimizer of (6.59).*

Proof. We first remark that, as we solve successive minimization problems, we have:

$$\begin{aligned}
F(u_n, v_n, w_n) &\ge F(u_n, v_n, w_{n+1}) \\
&\ge F(u_n, v_{n+1}, w_{n+1}) \\
&\ge F(u_{n+1}, v_{n+1}, w_{n+1}) \quad (6.71)
\end{aligned}$$

In particular, the sequence $F(u_n, v_n, w_n)$ is nonincreasing. As it is bounded from below by 0, it thus converges in $\mathbb{R}$. We denote by m its limit. We want to show that

$$m = \inf_{(u,v,w) \in X \times X \times X} F(u, v, w) \qquad (6.72)$$

Without any restriction, we can assume that, $\forall n$, $(u_n, v_n, w_n) \in X \times \mu B_G \times \delta B_E$. As F is coercive and as the sequence $F(u_n, v_n, w_n)$ converges, we deduce that the sequence (u_n, v_n, w_n) is bounded in $X \times \mu B_G \times \delta B_E$. We can thus extract a subsequence $(u_{n_k}, v_{n_k}, w_{n_k})$ which converges to $(\hat{u}, \hat{v}, \hat{w})$ as $n_k \to +\infty$, with $(\hat{u}, \hat{v}, \hat{w}) \in X \times \mu B_G \times \delta B_E$. Moreover, we have, for all $n_k \in \mathbb{N}$ and all w in X:

$$F(u_{n_k}, v_{n_k}, w_{n_k+1}) \le F(u_{n_k}, v_{n_k}, w) \qquad (6.73)$$

We also have, for all $n_k \in \mathbb{N}$ and all v in X:

$$F(u_{n_k}, v_{n_k+1}, w_{n_k+1}) \le F(u_{n_k}, v, w_{n_k+1}) \qquad (6.74)$$

and for all $n_k \in \mathbb{N}$ and all u in X:

$$F(u_{n_k}, v_{n_k}, w_{n_k}) \le F(u, v_{n_k}, w_{n_k}) \qquad (6.75)$$

Let us denote by $\tilde{w}$ a cluster point of (w_{n_k+1}). Considering (6.71), we get (since F is continuous on $X \times \mu B_G \times \delta B_E$):

$$m = F(\hat{u}, \hat{v}, \hat{w}) = F(\hat{u}, \hat{v}, \tilde{w}) \qquad (6.76)$$

By passing to the limit in (6.67), we get: $\tilde{w} = P_{\delta B_E}(f - \hat{u} - \hat{v})$. But from (6.76), we know that: $\|f - \hat{u} - \hat{v} - \hat{w}\| = \|f - \hat{u} - \hat{v} - \tilde{w}\|$. By uniqueness of the projection, we conclude that $\tilde{w} = \hat{w}$. Hence $w_{n_k+1} \to \hat{w}$.

Let us denote by $\tilde{v}$ a cluster point of (v_{n_k+1}). In the same way as above, we can show that $\tilde{v} = \hat{v}$. Hence $v_{n_k+1} \to \hat{v}$.

By passing to the limit in (6.73) (F is continuous on $X \times \mu B_G \times \delta B_E$), we therefore have for all w:

$$F(\hat{u}, \hat{v}, \hat{w}) \le F(\hat{u}, \hat{v}, w) \qquad (6.77)$$

By passing to the limit in (6.74) (F is continuous on $X \times \mu B_G \times \delta B_E$), we therefore have for all v:

$$F(\hat{u}, \hat{v}, \hat{w}) \le F(\hat{u}, v, \hat{w}) \qquad (6.78)$$

And by passing to the limit in (6.75), for all u:

$$F(\hat{u}, \hat{v}, \hat{w}) \leq F(u, \hat{v}, \hat{w}) \qquad (6.79)$$

(6.77), (6.78) and (6.79) can respectively be rewritten:

$$F(\hat{u}, \hat{v}, \hat{w}) = \inf_{w \in X} F(\hat{u}, \hat{v}, w) \qquad (6.80)$$

$$F(\hat{u}, \hat{v}, \hat{w}) = \inf_{v \in X} F(\hat{u}, v, \hat{w}) \qquad (6.81)$$

$$F(\hat{u}, \hat{v}, \hat{w}) = \inf_{u \in X} F(u, \hat{v}, \hat{w}) \qquad (6.82)$$

But, from the definition of $F(u, v, w)$, (6.80) is equivalent to (see [11]):

$$0 \in -f + \hat{u} + \hat{v} + \hat{w} + \lambda \partial B^* \left(\frac{\hat{w}}{\delta} \right) \qquad (6.83)$$

and (6.81) to

$$0 \in -f + \hat{u} + \hat{v} + \hat{w} + \lambda \partial J^* \left(\frac{\hat{v}}{\mu} \right) \qquad (6.84)$$

and (6.82) to

$$0 \in -f + \hat{u} + \hat{v} + \hat{w} + \lambda \partial J(\hat{u}) \qquad (6.85)$$

The subdifferential of F at $(\hat{u}, \hat{v}, \hat{w})$ is given by:

$$\partial F(\hat{u}, \hat{v}) = \frac{1}{\lambda} \left(\begin{array}{c} -f + \hat{u} + \hat{v} + \hat{w} + \lambda \partial J(\hat{u}) \\ -f + \hat{u} + \hat{v} + \hat{w} + \lambda \partial J^* \left(\frac{\hat{v}}{\mu} \right) \\ -f + \hat{u} + \hat{v} + \hat{w} + \lambda \partial B^* \left(\frac{\hat{w}}{\delta} \right) \end{array} \right) \qquad (6.86)$$

And thus, according to (6.85), (6.84) and (6.83), we have:

$$\left(\begin{array}{c} 0 \\ 0 \\ 0 \end{array} \right) \in \partial F(\hat{u}, \hat{v}, \hat{w}) \qquad (6.87)$$

which is equivalent to: $F(\hat{u}, \hat{v}, \hat{w}) = \inf_{(u,v,w) \in X^3} F(u, v, w) = m$. Hence the whole sequence $F(u_n, v_n, w_n)$ converges towards m the minimum of F on $X \times \mu B_G \times \delta B_E$. From Proposition 6.11, we deduce that the sequence (u_n) converges to $\hat{u}$ and that $(v_n + w_n)$ converges to $\hat{v} + \hat{w}$ when n tends to $+\infty$.

∎

Remark We can do the same kind of Mathematical analysis for our $u + v + w$ model as in Subsection 5.2 (see also [6] for further details).

6.3   Numerical results

Choice of the parameters The parameter λ controls the L^2 norm of the residual $f - u - v$, μ represents the G norm of the v component, and δ represents the E norm of the w component. λ and μ are tuned in the same ways as in Subsection 5.3. We display some numerical results on Figures 11 tand 12.

Comments The model in [20] can be written:

$$\min_{u,v} \left(\|\mathcal{C}u\|_1 + \|\mathcal{D}v\|_1 + \alpha \|f - u - v\|_2^2 + \beta J(u) \right) \qquad (6.88)$$

where $\mathcal{C}$ is a transform adapted to the geometrical part of the image (for instance the curvelet transform), and $\mathcal{D}$ a transform adapted to the textured part of the image (for instance the discrete cosinus transform). The noise $w = f - u - v$ is penalized through its L^2 norm. In fact, the authors of [20] get the noise component as a by-product of their model. But they have not really introduced a norm specific to the noise, contrary to what we have done. Moreover, if the curvelet transform is being used for the geometrical part of the image in (6.88), the computation time is then much longer than with our algorithm.

References

[1] R. Adams. *Sobolev Spaces*. Pure and applied Mathematics. Academic Press, Inc, 1975.

[2] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variations and free discontinuity problems*. Oxford mathematical monographs. Oxford University Press, 2000.

[3] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing*, volume 147 of *Applied Mathematical Sciences*. Springer-Verlag, 2002.

[4] J.F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Decomposing an image: Application to textured images and SAR images, 2003. INRIA Research Report 4704, to appear in JMIV, http://www.inria.fr/rrrt/rr-4704.html.

[5] J.F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Decomposing an image: Application to SAR images. In *Scale-Space '03*, volume 1682 of *Lecture Notes in Computer Science*, 2003.
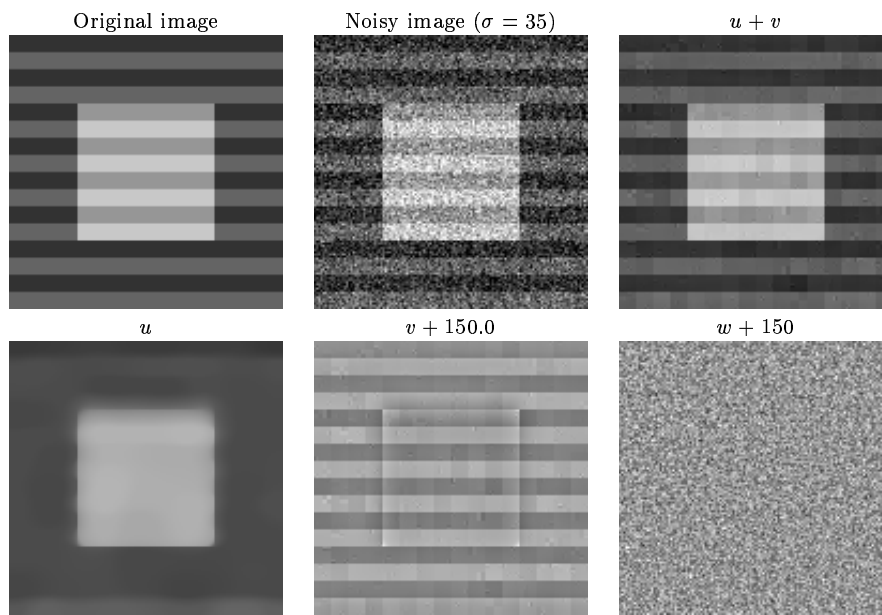
Figure 11: A simple example ($\lambda = 0.5$, $\mu = 120$, $\eta = 1.0$, Haar)

[6] J.F. Aujol and A. Chambolle. Dual norms and image decomposition models, 2004. INRIA Research Report 5130, http://www.inria.fr/rrrt/rr-5130.html.

[7] A. Chambolle. An algorithm for total variation minimization and applications. *JMIV*, 20:89–97, 2004.

[8] A. Chambolle and P.L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(3):167–188, 1997.

[9] A. Chambolle, R.A. De Vore, N. Lee, and B.J. Lucier. Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Transcations on Image Processing*, 7(3):319–335, March 1998.

[10] D.L. Donoho and M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224, December 1995.

[11] I. Ekeland and R. Temam. *Analyse convexe et problèmes variationnels*. Etudes Mathématiques. Dunod, 1974.

[12] L.C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, 1991.

[13] F. Malgouyres. Mathematical analysis of a model which combines total variation and wavelet for image restoration. *Journal of information processes*, 2(1):1–10, 2002.

[14] F. Malgouyres. Minimizing the total variation under a general convex constraint for image restoration. *IEEE transactions on image processing*, 11(12):1450–1456, December 2002.

[15] S.G. Mallat. A theory for multiresolutionsignal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, July 1989.

[16] S.G. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1998.

[17] Yves Meyer. Oscillating patterns in image processing and in some nonlinear evolution equations, March 2001. The Fifteenth Dean Jacquelines B. Lewis Memorial Lectures.

[18] S.J. Osher, A. Sole, and L.A. Vese. Image decomposition and restoration using total variation minimization and the H^{-1} norm. *Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal*, 1(3):349–370, 2003.

[19] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[20] J.L. Starck, M. ELad, and D.L. Donoho. Image decomposition: separation of texture from piecewise smooth content, 2003. To appear in IEEE Transactions on Image Processing.

[21] L.A. Vese and S.J. Osher. Modeling textures with total variation minimization and oscillating patterns in image processing. *Journal of Scientific Computing*, 19:553–572, 2003.

Original image

Noisy image ($\sigma = 20$)

u

$u + v$

$v + 150.0$

$w + 150$

Figure 12: Barbara image ($\lambda = 1.0$, $\mu = 30$, $\eta = 0.6$, Daub8)