Variational methods in image processing

Ali Haddad & Yves Meyer

August 14, 2004

1 Introduction

In 1985, David Mumford and Jayant Shah proposed to model black and white still images by functions of bounded variation. More precisely, in their model, an image f is a sum $u + v$ between a sketch u and a second term v which takes care of the textured components and of some additive noise. The objects which are contained in f belong to the sketch u. These objects are assumed to be delimited by contours with finite lengths and u is a geometric-type image. It is then natural to assume that u is a function of bounded variation. This assumption will however be questioned in this paper.

In 1992, Stanley Osher, Leonid Rudin, and Emad Fatemi simplified the Mumford & Shah model. Their model will be named the ORF model. Both models are equipped with an algorithm which yields the 'right decomposition' $f = \bar{u} + \bar{v}$ of a given image. This 'right decomposition' minimizes a certain energy. Both models are giving some access to the meaning of an image, since they are aimed at detecting the objects contained in it.

In these notes we will mostly focus on the ORF model. Here is our first result: Let us assume that we a priori know the objects which are contained in an image f and that the ORF algorithm is applied to f. It will be proved that some pieces belonging to the (physical) objects are always incorporated into the $\bar{v}$ component by the algorithm (see Theorem 3 below). Therefore Stanley Osher, Luminita Vese, and the second author have been looking for other models. One of these models is analyzed in these notes and, in some cases, performs much better than the ORF model, as proved in Theorem 11.

These notes are organized as follows. Sections 2, 3 and 4 contain introductory remarks on image processing, neurophysiology, and modeling. Sections 5 to 7 are devoted to the standard ORF model. A first alternative to the ORF model is studied in Section 8: The BV norm is replaced by a Besov norm in this new model which does not change the norm of indicator functions of rectifiable domains. This alternative model leads to an algorithm which is similar to Donoho's wavelet shrinkage. In Section 9 to 11 the ORF model is applied to a class of textured images and the drawbacks of the ORF model are analyzed. Section 12 is devoted to the new Osher-Vese model.

2 The primary visual cortex

A black and white image E is defined as a function $f(x) = f(x_1, x_2)$ which is named the grey-level of $x \in E$. We set $f(x) = 0$ if the point x is black in the given image and $f(x) = 1$ if x is bright white. Then $f(x) \in [0, 1]$ and the value of $f(x)$ is the grey-level at the point x. The pair $(x, f(x))$ is named a pixel. The bounded function $f(x)$ is not continuous in general. Jump discontinuities of $f(x)$ are playing a key role in image processing. Such discontinuities are generated by the edges of the objects which are present in the given image. At this level of the discussion the grey-level f is a bounded measurable function.

This immediately raises a fundamental issue, since it is clear that most bounded measurable functions do not correspond to natural images. Similarly a random sequence of letters is not a poem or a novel. A natural image has a meaning and meaningful images should be adequately modeled.

A first approach to modeling natural images consists in studying the laws which govern their production. The geometrical organization of the surrounding world and the laws of optics are concerned. But one should not forget to take in account the processing achieved by the human vision system. This processing can be used as a clue for modeling natural images. David Hubel writes in [31]:

In collaboration with Tosten Wiesel and Margaret Livingstone, *I have attempted to build upon the work of Ramón y Cajal in an attempt to obtain a detailed understanding of the physiology of one small part of the cerebral cortex-the striate cortex, or the primary visual cortex... Our main effort has been to determine how the visual information coming from the eye is handled and transformed by the brain...Cells in the primary visual cortex, to which the optic nerve projects (with one intermediate nucleus interposed) are far more exacting in their stimulus requirements. The commonest type of cells fires most vigorously not to a circular spot, but to a short line segment-to a dark line, a bright line, or to an edge boundary between dark and light. Furthermore each cell is influenced in its firing by a restricted range of line orientations: a line more than about 15 to 30 degrees form the optimum generally evokes no response. Different cells prefer different orientations, and no one orientation, vertical, horizontal or oblique is represented more than any other. These observations, made in 1958, had not been predicted and came as a complete surprise...A cell in the left hemisphere might respond to a bright red line oriented at 45 degrees to the horizontal, in a small region of the right visual field, but fail to respond to vertical lines, horizontal lines, or white and black lines...*

At an early stage of our research Torsten Wiesel and I began to ask whether the visual cortex was fully connected at birth, or whether the organization we had seen in adult animals required postnatal experience for its development. It is fair to say that the prevailing opinion of psychologists in the fifties and sixties was that the wiring of the brain and certainly of the cerebral cortex, was largely a product of post-natal experience.

We began by showing that a newborn animal that had one eye covered for a period as short as a few weeks became blind in that eye. We could show that the blindness was not the result of some abnormality in the eye, but of miswiring in the cortex, since the responses from the retina were normal, whereas stimulating the eye that had been covered evoked no reponse from the *cortex. These effects of eye occlusion could be produced only during a critical period of several months subsequent to birth. The period of plasticity in this system was therefore strictly limited. Clinical experience, for example with congenital cataracts, suggests that in humans the critical period extends for several years from birth. Surprisingly, blindfolding both eyes in young monkeys produced far less severe effects, even though the animals became blind in both eyes. Presumably the blindness is due in part at least to effects on higher visual areas. This difference in the effects of monocular and binocular occlusion must have been produced by competition, rather than by simple disuse.*

Perhaps the most surprising result of these studies was the demonstration that at birth, this region of cortex responds very much like adult cortex. Orientation selectivity, for example, is present to a striking degree, and the orientation columns have begun to develop. We conclude that the deprivation effects that we had seen were the result of deterioration of connections present in large part at birth, rather than simply a failure to adquire connections as a result of lack of a learning experience. Specific cortical connections, at least for this region of cortex, are thus partly innate and partly the result of experience, and not, as the previous generation of psychologists had thought, exclusively the result of experience.

This work in young cats and monkeys, especially the concept of the critical period, led to a means of prevention of one of the commonest forms of blindness, the monocular blindness (amblyopia) that result results from strabismus ("cross eye" or "wall eye"). By operating surgically at an early age to correct the condition it is possible to preserve vision in the two eyes, and even prevent the loss of binocular depth perception that almost always results from strabismus. As often happens in science, this beneficial result of our research was quite unforseen. We did not set out to prevent a specific form of blindness; the discoveries came as a by-product of studying the normal brain. Ironically, had we set at the very beginning to understand amblyopia, we would have begun *by working on the eye, not he brain, and would not have gotten anywhere. I suspect no funding agency would have supported a proposal to study amblyopia by making microelectrode recordings from the cortex. The deprivation studies began as a side project and did not require any additional resources.*

Understanding the human brain is one of the greatest challenges that scientists face. It is surely the most complex machine in the known universe. It has commonly been thought that understanding the brain is a hopeless quest, since the main tool we have to work with is our own brain. I have never seen any merit in this logic, and the work I have described here, together with research in hundred of laboratories, has surely already led to remarkable progress in understanding the brain, even though it is just a beginning.

Are these discoveries bridging the gap between neurophysiology and image processing ? Given my scientific training, I would ask another question: *Is the processing performed by the primary visual cortex adequately modeled by an atomic decomposition ?* Under the generic name of *atomic decompositions*, we find many algorithms ranging from *Fourier analysis* to *wavelet analysis*, including *Principal Component Analysis*, and *Independent Component Analysis*. The last two algorithms are aimed at finding simple and universal *building blocks* inside a given class of signals or images. This will be clarified in the next sections. An example of *building blocks* is given by phonemes in speech processing. Phonemes are the shortest pieces of sound which still have a linguistic meaning. David Hubel tells us that some cells of the primary visual cortex are detecting the simplest geometric entities which still have a meaning in an image. Can this processing be viewed as an atomic decomposition ?

A second observation concerns the role of edge detection in David Hubel's work. Some cells in the primary visual cortex are specialized in detecting line segments with a given orientation. The paramount importance of geometry in this processing advocates for a class of models where one is looking for the simplest geometric structures which fit the data. These remarks pave the road to the models which were mentioned in our introduction. The first one is the Mumford-Shah model. A similar but simpler model was proposed by Stanley Osher, Leonid Rudin, and Emad Fatemi and will be treated here. An alternative to the ORF model will be studied in Section 8 and the corresponding algorithm is similar to a wavelet shrinkage which, in a sense, reconciles some atomic decompositions with some variational approaches. As it was already told, our discussion will naturally lead to the Osher-Vese model.

3 Atomic decompositions

In the Webster's dictionary, the word *analysis* is defined as follows:

> *A breaking up of a whole into its parts so as to find out their nature.*

Therefore analyzing a function, a signal, an image means breaking it into simpler entities named *building blocks*, *time-frequency atoms*, or *wavelets*. This is what we call an *atomic decomposition.* This paradigm sprang out in signal or image processing and in mathematics in the seventies. Given a class $\mathcal{C}$ of signals, images or functions, we are looking for the simplest entities which generate all elements $f \in \mathcal{C}$. These entities are named *atoms* and are the building blocks from which we can construct any $f \in \mathcal{C}$ by suitable linear combinations. This definition implies two facts:

(a) *for discovering the atoms, the signals $f \in \mathcal{C}$ need to be analyzed*

(b) *a synthesis is needed for checking that the atoms generate the signals $f \in \mathcal{C}$.*

Atomic decompositions were seminal in many other scientific fields. In chemistry, the first building blocks were pure elements. Next we find molecules and atoms. In physics these atoms are further decomposed into simpler entities. Examples abound in other sciences. In signal or image processing, the definition of the building blocks, the type of processing one should use to decompose a signal into a series of building blocks, and the rules to be used in the synthesis are context dependent as it will be shown.

Let us mention a few examples. The oldest option is the windowed Fourier analysis or short-time Fourier analysis. An equivalent formulation was given by Dennis Gabor in 1945 and is today known as *Gabor wavelets.* One dimensional Gabor wavelets are the functions $w_{(t_0,\omega)} = \exp(-(t-t_0)^2)\cos(\omega t+\phi)$. These functions are labelled by a point (t_0, ω) in the time frequency plane. As noticed by Dennis Gabor and also by Léon Brillouin, if a Gabor wavelet $w_{(t_0,\omega)}$ is unfolded into the time-frequency plane, it covers a disc around the point (t_0, ω) of the time-frequency plane. The area of the spot is 2π and this systematic lack of localization is explained by Heisenberg uncertainty principle.

More generally, let us consider a 1-D signal $f(t)$ where t denotes the time variable. Unfolding $f(t)$ in the time frequency plane amounts to trying to draw the graph Γ of the instantaneous frequency $\omega(t)$, viewed as a function of the time variable t. In other words, we wish to write the musical score while listening to the music. Heinsenberg principle says that Γ cannot be a graph, but instead is a thick curve wich covers a large area.

Returning to Gabor wavelets, it becomes natural to evenly space the points (t_0, ω) in the time-frequency plane. Gabor considered the grid of the time frequency plane defined by $t_k = 0, \pm 1, \pm 2, \cdots$, $\omega_l = 0, \pm 2\pi, \pm 4\pi, \cdots$ and the corresponding paving by adjacent rectangles $R_{(k,l)} = [k, k+1] \times [2\pi l, 2\pi(l+1)],\ \ k, l \in \mathbf{Z}^2$. This paving is consistent with Heisenberg uncertainty principle and was also advocated by L. Brillouin. Ingrid Daubechies unveiled the drawbacks of these Gabor wavelets. Indeed the closed linear span of the Gabor wavelets is $L^2(\mathbf{R}^2)$ but the expansion of a function in $L^2(\mathbf{R}^2)$ as a series of Gabor wavelets is unstable, since it needs extremely large coefficients. However if $t_k = k$ is replaced by $t_k = \alpha k$ where $0 < \alpha < 1$, then the corresponding Gabor wavelets form an overcomplete system or a frame [21].

Other solutions were proposed Ronald Coifman and the second author. These new algorithms are named *local trigonometric bases* and are aimed at analyzing quasi-stationary audio signals [32], [35]. Expanding a signal into a local trigonometric basis begins with a segmentation of the given signal. This segmentation is using some specific windows which are incorporated in the algorithm. An optimal choice of the windows will provide the sparsest representation. Once the signal is segmented, the next step is using an alternstive Fourier series expansion. A 2-D version of this algorithm exists but the segmentation is much more constrained than in 1-D. That explains why 2-D local trigonometric bases cannot be used as edge-detectors, as opposed to Candès-Donoho's curvelets [12].

Time-scale algorithms became popular in the early eighties. They emerged at about the same time in distinct communities. In image processing and computer vision, one should mention the fundamental achievements by David Marr [36] who was studying the processing performed by the human vision system. Marr discovered time-scale wavelets in connection with edge-detection. Marr's wavelets are isotropic and the mother-wavelet is defined as $\psi(x) = \Delta g$ where $g(x) = \exp(-|x|^2)$ and Δ is the laplacian. The other Marr wavelets are $\psi_{a,b} = a^{-1}\psi(\frac{x-b}{a})$ where $a > 0$ is the scale and $b \in \mathbf{R}^2$ is the position of the wavelet. It is striking to observe that this pioneering work by David Marr, the discovery of pyramidal algorithms by Burt and Adelson, and the fundamental work by Alexandre Grossmann and Jean Morlet were achieved at about the same time. The reader is referred to [9] or [35].

Let us return to David Hubel. *Is it possible to model the cells in the primary visual cortex by some wavelets ψ_λ which should combine the frequency localization of the Gabor wavelets together with the localization of the Grossmann-Morlet wavelets in the space domain ?* A single orthonormal basis consisting of wavelets with these three degrees of freedom (scale, frequency and position) does not exist. This leads to five alternative solutions which range from the best-bases algorithms to ICA. let us be more specific.

Ronald Coifman proposed to have access to a library containing infinitely many orthonormal bases. One freely picks a basis in the library and this freedom provides us with a finer tuning of the scales, the frequencies and the positions of the time-frequency atoms which are used. A selection algorithm yields the basis which fits a given signal or a given image optimally.

Coifman's best basis algorithm is an adaptive algorithm and yields a context dependent atomic decomposition. The 'matching pursuit algorithm' introduced by Stéphane Mallat is similar in spirit. One starts with an extremely large overcomplete collection $\mathcal{C}$ of time-scale-frequency wavelets ψ. These wavelets are normalized by $\|\psi\|_2 = 1$. Given a signal or an image f, one picks a wavelet $\psi_1 \in \mathcal{C}$ yielding the largest correlation $c_1 =< f, \psi_1 >$ with f.

Then $c_1\psi_1$ is substracted from f which yields a first residual f_1. One proceeds on this residual and obtains a sequence f_n, $n \geq 1$, of residuals. Finally the *matching pursuit* algorithm provides us with the series expansion $f = \sum_1^\infty f_n$.

The multi-layered representations of images were introduced and studied by Amir Averbuch, Ronald Coifman, and François G. Meyer. They reconcile *atomic decompositions* with the $u + v$ models of Mumford & Shah. In [38] we read

> *The underlying assumption behind standard compression methods is that the basis (e.g. the DCT basis, or a wavelet basis) used for compression is well adapted to most images. Our approach follows a completely different direction: – instead of forcing all images to adapt to one single basis, we use a collection of libraries of bases to represent a single image.*
>
> *The main contribution of this work is a new paradigm for image compression. We describe a new multi-layered representation technique for images. An image is encoded as the superposition of one main approximation, and a sequence of residuals. The strength of the multi-layered method comes from the fact that we use different bases to encode the main approximation and the residuals. For instance, we can use: – a wavelet basis to encode a coarse main approximation of the image, – wavelet packet bases to encode textured patterns, – brushlet bases to encode localized oriented textured features, etc.*
>
> *During each iteration of the algorithm, we code the residual part in a lossy way: we only retain the most significant structures of the residual part, which results in a sparse representation of the residual. Because each layer has a sparse representation in the associated basis, the superposition of layers achieves a very compact coding. By selecting different wavelet and wavelet packet bases, we allow different features to be discovered in the image. Furthermore, with each new basis we can encode and compensate for the artifacts created by the quantization of the coefficients of the previous bases. This multi-layered representation has a lot of*

beautiful potential applications in many areas of image processing, and image coding.

In a striking paper, entitled *Emergence of simple-cell receptive field properties by learning a sparse code for natural images,* in *Nature*, pp.607-609, 1996, (381), B. A. Olshausen and D. J. Field propose the following paradigm: the highly specialized tasks of the cells studied by David Hubel could be the result of an evolutionary process which selected this solution. The selection criterion is concision and robustness. That is why Olshausen and Fields aimed at finding the most concise representation of the class of all natural images. Instead of processing all natural images, they used some pictures of landscapes found in a data basis. For discovering the most concise representation, they used an algorithm named *ICA*, or Independent Component Analysis. *ICA* will be defined in these notes. It happened that the building blocks found by Olshausen and Fields are close to some of the wavelets belonging to Coifman's libraries. A similar approach can be found in "Unsupervised learning" by H. B. Barlow (Neural computation 1, pp. 295-311, 1989). H. B. Barlow also suggests that mammals use concise and efficient representations to process natural images.

Before closing this section, let us mention a sriking discovery by Emmanuel Candès and David Donoho [11], [12] where atomic decompositions and geometry are truly reconciled. In full contrast with Coifman's best basis search, Candès and Donoho where looking for <u>one</u> basis which would optimally compress all cartoon images. Here a cartoon image consists of finitely many regions Ω_j, $1 \leq j \leq N$, delimited by smooth boundaries Γ_j. The cartoon image is required to be smooth inside each Ω_j with jump discontinuities across the boundary Γ_j of Ω_j. If $e_j,\ j \in \mathbf{N}$, is an orthonormal basis, we denote by $\rho_j(f)$ the norm $\|f - f_j\|_2$ where f_j denotes the best approximation to f by a linear combination $f_j = \alpha_1 e_{n_1} + \ldots + \alpha_j e_{n_j}$. Using a standard wavelet expansion, we obtain $\rho_j(f) = j^{-1/2}$. Using a Fourier series expansion, the error would have been $j^{-1/4}$ and with Donoho-Candès curvelets, the error (still computed in L^2) is reduced to $j^{-1}(\log j)^{3/2}$. Up to the logarithmic term, this convergence rate is optimal. I cannot resist quoting Candès and Donoho. In [12] they write

In fact neuroscientists have identified edge-processing neurons in the earliest and most fundamental stages of the processing pipeline *upon which mammalian visual processing is built...This article is motivated by fundamental questions concerning the mathematical representation of objects containg edges: what is the sparsest representation of functions $f(x_1, x_2)$ that contain smooth regions but also edges ?*

4 Independent Component Analysis

When it exists, *Independent Component Analysis* improves on *Principal Component Analysis.* In both cases, we are given a collection $\mathcal{C} = \{X(\omega),\ \omega \in \Omega\}$ of vectors $X(\omega)$ belonging to some Hilbert space $\mathcal{H}$ and a probability law $dP(\omega)$ on Ω. This probability law indicates how likely a specific image is to be found in $\mathcal{C}$. We shall assume that the center of gravity $g = \int_\Omega X(\omega)dP(\omega)$ of $\mathcal{C}$ is 0. If $e_n, n \in \mathbf{N}$, is an orthonormal basis of $\mathcal{H}$, we denote by $r_n(\mathcal{C}) = E[\|f - \sum_{0\le m\le n} c_m e_m\|^2]$ the expected squared error in the expansion of $f \in \mathcal{C}$. Then the Karhunen-Loève basis is built as follows: the vectors $e_0, e_1, \cdots, e_n, \cdots$ are inductively defined by requiring that the next error $r_{n+1}(\mathcal{C})$ is minimized by the choice of e_{n+1}. If we are given a cloud $\mathcal{C}$ consisting of M points $X_j \in \mathbf{R}^N,\ 1 \le j \le M$, and if the center of gravity g of $\mathcal{C}$ is 0, we let D be the line through 0 which provides us with the best fit to $\mathcal{C}$ and we consider the hyperplane Π which is perpendicular to D. We then consider the orthogonal projection $\mathcal{C}_1$ of $\mathcal{C}$ on Π and we proceed on this new cloud of points. If $\mathcal{C}$ is the unit cube $[-1/2, 1/2]^N \subset \mathbf{R}^N$ equipped with the Lebesgue measure, then any orthonormal basis of $\mathbf{R}^N$ is a KL basis. The second order statistics of the cube and the ball are identical. In full contrast with this undetermination, *Independent Component Analysis* yields the right basis together with the correct information that a cube is not a ball.

Let us now consider *Independent Component Analysis.* As in the KL decomposition, we are studying a family $\mathcal{C} = \{X(\omega),\ \omega \in \Omega\}$ of signals and we are given probability law $dP(\omega)$ on Ω. For the sake of simplicity, we first assume $\mathcal{H} = \mathbf{R}^N$. Then $X(\omega) = x(j, \omega),\ 1 \le j \le N,\ \omega \in \Omega$, is sampled on $\{1, 2, \cdots, N\}$. One is interested in knowing if these signals could be written as a linear combination of m independent sources $s_1(\omega), ..., s_m(\omega),\ \omega \in \Omega$.

We then would have

$$x(j,\omega) = \sum_1^m \alpha_{(j,k)}\, s_k(\omega) \quad (1 \le j \le N) \tag{1}$$

where the matrix $A = (\alpha_{(j,k)})_{1\le j\le N, 1\le k\le m}$ is determinisic. In other words $X(\omega) = AS(\omega),\ \omega \in \Omega$, where $X(\omega),\ \omega \in \Omega$, is the given collection and where both the $N \times m$ matrix A and the sources $S(\omega)$ are unknown. When it exists, the *ICA* decomposition of a stochastic process $X(t,\omega)$, $t \in [a,b]$, is defined in a similar way and reads

$$X(t,\omega) = \sum_0^\infty \alpha_k(\omega) f_k(t) \tag{2}$$

where the random variables $\alpha_k(\omega)$ are independent. Similarly a KL decomposition yields a series expansion (2) where the random variables $\alpha_k(\omega)$ are orthogonal with respect to the probability law. In the Gaussian case, a KL decomposition is always an *ICA.* In general they differ.

Most families $\mathcal{C}$ of signals do not have an *ICA.* These is the bad news. But if *ICA* exists, it brings a lot of information. Here are two examples.

Our first example is the ramp process. The ramp process $X(t,\omega)$ is indexed by $\omega \in [0,1]$ and defined by $X(t,\omega) = t$ si $0 \le t < \omega$ and $X(t,\omega) = t-1$ si $\omega \le t \le 1$. If a Karhunen-Loève decomposition is applied to the ramp, the result is the trigonometric system $\sin(k\pi t), k \ge 1$, and this representation is suffering from two main drawbacks. It is slow and does not provide the right information (which is the jump discontinuity at ω). But *ICA* does not exist for the ramp. One cannot write

$$X(t,\omega) = \sum_0^\infty \alpha_k(t)\phi_k(\omega) \tag{3}$$

where $\phi_k(\omega)$ are independent random varaiables. The best one can do is to apply an algorithm which yields the independent components when these components exist. Then the *ICA algorithm* provides us with the wavelet series expansion of the ramp. The convergence speed is exponential. This leads to the conclusion that *ICA* algorithms do their best when they are applied to signals with strong transients or to images.

The second example is the unit cube. Then the *ICA* of the cube is unique (up to a permutation) and yields the most concise and meaninful description of a cube. The sources are here the usual coordinates, the matrix A is the identity, and the cube is defined by $-1/2 \leq x_j \leq 1/2$, $1 \leq j \leq N$. Let us stress that a KL expansion is an *ICA* in the Gaussian case, since then independence is implied by orthogonality. The ramp process and the unit cube are not Gaussian processes.

5 The Osher-Rudin-Fatemi model

The paradigm of *atomic decompositions* was seminal in the preceding sections. Atomic decompositions are bridging the gap between computational methods and neurophysiology. We now take another road and study some variational approaches to image analysis. The Osher-Rudin-Fatemi model (ORF) will be singled out. In this model, an image will be viewed as something that can be drawn by a painter. Neurophysiology is replaced by art. Indeed they both rely on the remarkable properties of the human brain. The ORF model is motivated by ill-posed inverse problems and the BV norm is a regularization which preserves edges [44]. Denoising is a particular example. As it will be proved in these notes, *The Osher-Rudin-Fatemi algorithm is indeed performing a thresholding which is similar to the wavelet shrinkage defined by David Donoho.* More precisely the ORF algorithm has the following property: there exists a norm $\|\cdot\|_*$ and a threshold $\frac{1}{2\lambda}$ such that an image f with a norm $\|f\|_* \leq \frac{1}{2\lambda}$ is put to 0 and an image with a norm larger than the threshold is reduced by a fixed amount (see Theorem 2 and its two corollaries). This is similar to Donoho's wavelet shrinkage. The ORF algorithm can be viewed as being a shrinkage where the emphasis is put on the geometry. Here comes the bad news. The ORF algorithm and Donoho's wavelet shrinkage present the same drawback. In a systematic way pieces which belong to the true objects contained in an image are wiped out by the algorithm and viewed as belonging to the textured component. The Osher-Vese algorithm performs better (Theorem 11).

The ORF model is a simplified variant on the Mumford & Shah model.

In both models, a given image is optimally split into two parts u and v. In the Mumford & Shah model, the given image f is defined over a domain Ω, the u component belongs to the subspace SBV of BV, which consists of functions in BV whose distributional gradient does not contain a singular diffuse measure. In other words, this distributional gradient ∇u is the sum between an L^1 function and a measure carried by a one dimensional singular set K. Then the Mumford & Shah penalty on the $u(x)$ component is a sum $J(u)$ between two terms. The first term is the one-dimensional Hausdorff measure of K. The second one is the square of the L^2 norm of the gradient of $u(x)$ calculated on the complement of this singular set K. The third term of the $J(u)$ functional is the square of the L^2 norm of $v(x)$. Then the functional to be minimized is

$$J(u) = \int_{\Omega \setminus K} |\nabla u|^2 \, dx + \alpha \mathcal{H}^1(K) + \beta \|v\|_{L^2}^2$$

where $\mathcal{H}^1(K)$ denotes the 1-dimensional Hausdorff measure of K. The two positive parameters α and β need to be tuned effectively. As is the case for the parameter λ in the Osher-Rudin-Fatemi model, if α is small, then too many objects will be detected.

This model raises beautiful mathematical problems which have much to do with the theory of minimal surfaces [22]. In the Mumford and Shah model, textures are treated as noise.

The ORF model is similar. An image $f(x)$ is again a sum between two components u and v. The first component u takes care of the objects which are included in the given image f. These objects can be human beings, animals, some furniture or other items. The main assumption concerning these objects says that they can be drawn by a painter since they are delimited by some boundaries which have finite lengths. This observation indicates that the space BV of functions of bounded variation will play a key role in the ORF model. This will become clearer if some properties of BV are reviewed.

One is tempted to say that the indicator function $\chi_E(x)$ of a domain E belongs to the space BV of functions of bounded variation if and only if its boundary ∂E has a finite length and that the BV norm of $\chi_E(x)$ equals the length of the boundary. This statement is clearly valid in the $\mathcal{C}^1$ case but is not true in general. The BV norm will be soon defined. In order to treat the general case, De Giorgi defined the reduced boundary $\partial^* E$ of a measurable set E and proved that the BV norm of χ_E is the 1-dimensional Hausdorff measure of its reduced boundary.

For defining this reduced boundary, let us denote by $B(x, r)$ the ball centered at x with radius r. We then follow De Giorgi:

Definition 1 *The reduced boundary $\partial^* E$ of E is the set of points x belonging the closed support of $\mu = \nabla \chi_E$ such that the following limit exists*

$$\lim_{r \to 0} \quad \frac{\mu\{B(x, r)\}}{|\mu|\{B(x, r)\}} = \nu(x) \tag{4}$$

An indicator function χ_E belongs to BV if and only if $\partial^* E$ has a finite 1-dimensional Hausdorff measure. More generally a function $f(x)$ defined on $\mathbf{R}^2$ belongs to BV if (a) $f(x)$ vanishes at infinity in a weak sense and (b) the distributional gradient ∇f of $f(x)$ is a (vector valued) bounded Borel measure. An apparently weaker definition reads as follows: f belongs to BV if its distributional gradient is a (vector valued) bounded Borel measure. Then is is easily proved that $f = g + c$ where c is a constant and g tends to 0 at infinity in the weak sense. Moreover g belongs to $L^2(\mathbf{R}^2)$ and we have

$$\|g\|_2 \leq \frac{1}{2\sqrt{\pi}} \|f\|_{BV} \tag{5}$$

The proof of (5) relies on the co-area identity and will be delayed after Theorem 1.

The condition at infinity says that $f \star \varphi$ tends to 0 at infinity whenever φ is a function in the Schwartz class. For instance, any function in $L^p(\mathbf{R}^2)$, $1 \leq p < \infty$, tends to 0 at infinity in this weak sense. An equivalent definition of BV is given by the following observation. If f belongs to BV, then there exists a constant C such that for every compactly supported continuous function g, the convolution product $h = f \star g$ belongs to $\mathcal{C}^1$ and satisfies $\|\nabla h\|_\infty \leq C\|g\|_\infty$. Conversely this property characterizes BV.

When the Osher-Rudin-Fatemi model is being used, a specific definition of the BV norm is crucially needed. Indeed the ORF model amounts to minimizing a functional which contains this BV norm. We will impose that this norm be isotropic. Let us begin by the simple case where ∇f belongs to $L^1(\mathbf{R}^2)$. Then the BV-norm of f will be defined as $\|f\|_{BV} = \int |\nabla f(x)|\, dx$. We then write $f \in \dot{W}^{1,1}$ and this function space will be useful in what follows. If ∇f is a general Borel measure, our simply minded approach does not work but pave the way to the following definition. We write $\mu_j = \partial_j f$ and we define the Borel measure σ by $\sigma = |\mu_1| + |\mu_2|$. By the Radon-Nikodym theorem we have $\mu_j = \theta_j(x)\sigma$, $j = 1, 2$, where $\theta_j(x)$ are Borel functions with values in $[-1, 1]$. Finally the Borel measure $|\nabla(f)|$ is defined by

$$|\nabla(f)| = \sqrt{\theta_1^2 + \theta_2^2}\, \sigma \tag{6}$$

We can conclude:

Definition 2 *The BV norm of f is the total mass of the Borel measure $|\nabla(f)|$.*

With an obvious abuse of language, we write $\|f\|_{BV} = \int \sqrt{|\mu_1|^2 + |\mu_2|^2}\, dx$. With these new notations the co-area identity reads as follows [4], [15].

Theorem 1 *Let $f(x)$ be a real valued measurable function defined on the plane and belonging to BV. Let us denote by Ω_t, $t \in \mathbf{R}$, the measurable set defined by*

$$\Omega_t = \{x \in \mathbf{R}^2 \mid f(x) > t\} \tag{7}$$

Let $\partial^*\Omega_t$ be the reduced boundary of Ω_t and $l(t)$ the 1-dimensional Hausdorff measure $\mathcal{H}^1(\partial^*\Omega_t)$. Then one has

$$\|f\|_{BV} = \int_{-\infty}^{+\infty} l(t)dt \tag{8}$$

In other words the sum of all the lengths of the level sets of f yields the BV norm of f. This identity needs to be completed with the following observation

$$f(x) = \lim_{m\to\infty} [\int_{-m}^{\infty} \chi_{\Omega_t}(x)dt - m] \tag{9}$$

where the limit is taken with respect to the weak* topology of BV.

A first approximation to this theorem was given in the pioneering work by Fleming and Rishel and Theorem 1 was completed by De Giorgi. An advised reference is [10].

Roughly speaking Theorem 1 says that any function f in BV can be written as a convex combination of indicator functions of domains with rectifiable boundaries. If a positive constant γ exists such that the norm $\|\cdot\|$ in a functional Banach space B satisfies $\|\chi_E\| \leq \gamma\, \mathcal{H}^1(\partial^* E)$ for any Borel set E, then BV is contained in B and we have $\|f\| \leq \gamma\, \|f\|_{BV}$ in full generality. This remark is applied when $B = L^2(\mathbf{R}^2)$. Then the isoperimetric inequality yields $\gamma = \frac{1}{2\sqrt{2\pi}}$ and the proof of (5) is complete. In many examples an alternative version of (9) is useful. If $f \in BV$ we denote by f^+ and f^- the functions $\sup(f(x), 0)$ and $\sup(-f(x), 0)$. Then f^+ and f^- both belong to BV. When f is a non-negative function in BV, (9) takes the following simple form $f(x) = \int_0^\infty \chi_{\Omega_t}(x)dt$. A final remark concerns the approximation of a function $f \in BV$ by simple functions. Let us pave the plane with congruent squares $Q(j,k),\ k \in \mathbf{Z}^2$, the size-length of $Q(j,k)$ being $2^{-j},\ j \in \mathbf{N}$. We then have

Proposition 1 *If $f \in BV$, we denote the mean value of f on $Q(j,k)$ by $m(j,k)$ and we consider the simple function f_j defined by $f_j(x) = m(j,k)$ on each $Q(j,k)$. Then we have $\|f_j\|_{BV} \leq 4\|f\|_{BV}$ and f_j weakly converges to f as j tends to infinity.*

The simple proof of Proposition 1 is left to the reader. An other useful remark is given by the following lemma

Lemma 1 *If $\theta(t)$ is a Lipschitz function of the real variable t and if $f \in BV$, then*

$$\|\theta(f)\|_{BV} \leq \|\frac{d}{dt}\theta\|_\infty \|f\|_{BV} \tag{10}$$

In what follows $\theta(t) = 0$ if $t \leq 0, \theta(t) = t$ if $0 \leq t \leq 1$, and $\theta(t) = 1$ if $t \geq 1$.

The ORF algorithm aims at decomposing an image f into a sum $u + v$ between two components. The first component u takes care of what might be drawn by a painter using a pencil. Therefore this component is adequately modeled by a function in BV. The texture and the noise belong to the component v. Then u is a sketch of the given image and this sketch u captures the main geometric features of the given image f. The v component is more complex and is not described by a functional Banach space in the ORF model. In the ORF model, $v \in L^2(\mathbf{R}^2)$ since both f and u are square-integrable. The ORF algorithm depends on a tuning parameter $\lambda > 0$. Objects with size less than $\frac{1}{2\lambda}$ will be treated as some texture and wiped out from u. We then arrive to the definition of the algorithm.

Definition 3 *Let $f \in L^2(\mathbf{R}^2)$. Then the ORF decomposition $f = u + v$ of f minimizes the functional $J(u) = \|u\|_{BV} + \lambda \|v\|_2^2$ among all decompositions of f as a sum between a function $u \in BV$ and $v \in L^2(\mathbf{R}^2)$.*

Later on the ORF model will be generalized and the space BV will be replaced by other functional spaces. This will pave the way between the ORF algorithm and the famous wavelet shrinkage.

6 Properties of the ORF algorithm

Let us denote by $\dot{W}^{1,1}$ the closure in BV of the linear space of smooth functions with compact support. In other words $f \in \dot{W}^{1,1}$ means $\nabla f \in L^1$. Here $\dot{W}^{1,1}$ denotes the homogeneous version of the standard Sobolev space. The dual space of $\dot{W}^{1,1}$ is the Banach space G consisting of all generalized functions g which can be written as

$$g = \operatorname{div} H \tag{11}$$

where $H = (h_1, h_2) \in L^\infty \times L^\infty$. The norm of g in G is denoted by $\|g\|_*$ and is defined as being the infimum of $\|H\|_\infty$ where this infimum is computed over all decompositions (11) of g. Here and in what follows $\|H\|_\infty = \sup_{x \in \mathbf{R}^2} |H(x)|$, $|H| = \sqrt{|h_1|^2 + |h_2|^2}$. Then $L^2 \subset G$ and the space G_0 is defined as the closure of L^2 in G. We have [37]

Lemma 2 *The Banach space BV is the dual space of G_0 and the norm in BV is the dual norm.*

This implies the following property: if a sequence u_j of functions belonging to BV converges to u in the distributional sense and if $\|u_j\|_{BV} \le 1$, then u belongs to BV and $\|u\|_{BV} \le 1$. This weak compactness property implies the existence of the optimal ORF decomposition and uniqueness is standard.

The following lemma will be needed.

Lemma 3 *If both u and v belong to L^2, then we have*

$$|\int u(x)v(x)\,dx| \leq \|u\|_{BV}\|v\|_* \tag{12}$$

For proving it, it suffices to approach u in L^2 by a sequence of functions in $\dot{W}^{1,1}$. The details can be found in [37].

An image f satisfies $0 \leq f(x) \leq 1$ and the u component inherits this property. The proof of this remark is easy. If $f = u + v$ is the ORF decomposition, we have $f = \theta(f) = \theta(u) + w$. We then have $\|\theta(u)\|_{BV} \leq \|u\|_{BV}$ and $|w(x)| = |\theta(f) - \theta(u)| \leq |f(x) - u(x)|$. Therefore $\|w\|_2 \leq \|v\|_2$ and the uniqueness of the ORF decomposition implies $u = \theta(u)$ as announced.

Antonin Chambolle made the following crucial remark. Let F_λ be the closed convex subset of $L^2(\mathbf{R}^2)$ defined by $\|\cdot\|_* \leq \frac{1}{2\lambda}$. Then we have

Theorem 2 *The ORF decomposition $f = \bar{u} + \bar{v}$ of a function $f \in L^2(\mathbf{R}^2)$ is given by*

$$\bar{v} = \operatorname{Arg\,inf}\{\|f - v\|_2;\ \ v \in F_\lambda\} \tag{13}$$

The proof of this theorem will be given in the following section. For the time being, let us comment on Theorem 2.

Corollary 1 *Let $f \in L^2(\mathbf{R}^2)$. If $\|f\|_* \leq \frac{1}{2\lambda}$, then the ORF decomposition of f is given by $\bar{u} = 0$, $\bar{v} = f$. If $\|f\|_* > \frac{1}{2\lambda}$, it is given by $f = \bar{u} + \bar{v}$ where $\|\bar{v}\|_* = \frac{1}{2\lambda}$ and $\|f - \bar{v}\|_2$ is minimal under that constraint.*

Corollary 1 is now rephrased in a more intuitive way:

Corollary 2 *If $\|f\|_* \leq \frac{1}{2\lambda}$, then f is a texture corrupted by an additive noise and does not contain any object.*

Returning to Corollary 1 and assuming $\|f\|_* > \frac{1}{2\lambda}$, we have $f = \bar{u} + \bar{v}$, $\|v\|_* = \frac{1}{2\lambda}$ and $\|\bar{u}\|_2$ is minimal under these requirements. An other characterization of the optimal pair $(\bar{u}, \bar{v})$ is given by $f = \bar{u} + \bar{v}$, $\|\bar{v}\|_* = \frac{1}{2\lambda}$, and $\int \bar{u}\bar{v} = \|\bar{u}\|_{BV}\|\bar{v}\|_*$. This leads to the following definition:

Definition 4 *A pair (u,v) of two functions in $L^2(\mathbf{R}^2)$ is named an extremal pair if $u \in BV$ and $\int uv\,dx = \|u\|_{BV}\|v\|_*$.*

Theorem 2 can be rephrased into the following assertion

Corollary 3 *If $\|f\|_* > \frac{1}{2\lambda}$, the ORF decomposition $f = \bar{u} + \bar{v}$ of f is characterized by the following two properties: (a) $\|\bar{v}\|_* = \frac{1}{2\lambda}$ and (b) $(\bar{u}, \bar{v})$ is an extremal pair.*

The reader is referred to [37] where Corollary 3 is given a direct proof. Corollaries 1, 2 and 3 are telling us that the ORF algorithm is a shrinkage where the threshold is $\frac{1}{2\lambda}$.

7 The abstract formulation of the ORF algorithm

A proof of Theorem 2 is given now. This proof is valid in a more general context which reads as follows.
Let H be a real Hilbert space. The norm in H is denoted by $|\cdot|$ and the corresponding inner product is $x \cdot y$. Let F be a non-empty closed convex subset of H. Let us define $p : H \mapsto \mathbf{R} \cup \{+\infty\}$ by

$$p(x) = \sup\{x \cdot y;\ y \in F\} \tag{14}$$

This functional p is convex, lower semi-continuous and satisfies $p(\lambda x) = \lambda p(x)$ for $\lambda > 0,\ x \in H$. Let $f \in H$ be given. Among all decompositions $f = x + y$ of f we want to find the one for which the energy $J(x) = p(x) + \lambda |y|^2$ is minimal. We denote by $f = \bar{x} + \bar{y}$ this optimal decomposition. Without loosing generality we can assume $\lambda = 1/2$. It suffices to replace F by $\frac{1}{2\lambda}F$ to obtain the general case. Following Antonin Chambolle we have

Theorem 3 *With the preceding notations the optimal decomposition $f = \bar{x} + \bar{y}$ is given by*

$$\bar{y} = \operatorname{Arg\,inf}\{|f - y|;\ y \in F\} \tag{15}$$

The proof of Theorem 3 is not difficult. It consists in applying the celebrated von Neumann minimax theorem to the functional

$$V(x,y) = x \cdot y + \frac{1}{2}|f - x|^2,\ \ x \in H, y \in F \tag{16}$$

The minimax theorem says the following. Let E be a compact and convex set. Let F be a convex set which does not need to be given a topological structure. We consider a functional $V : E \times F \mapsto \mathbf{R}$. We define $P : E \mapsto \mathbf{R} \cup \{\infty\}$ and $Q : F \mapsto \{-\infty\} \cup \mathbf{R}$ by

$$P(x) = \sup\{V(x, y); \;\; y \in F\} \tag{17}$$

and similarly

$$Q(y) = \inf\{V(x, y); \;\; x \in E\} \tag{18}$$

We obviously have $Q(y) \leq V(x, y) \leq P(x)$ which implies $\sup\{Q(y), \; y \in F\} = \beta \leq \alpha = \inf\{P(x), \;\; x \in E\}$. The minimax theorem yields $\alpha = \beta$ under the following assumptions

Theorem 4 *Let us assume that $x \mapsto V(x, y)$ is convex and lower semi-continuous on E for every $y \in F$. Let us also assume that $y \mapsto V(x, y)$ is concave on F for every $x \in E$. Then there exists an element $\bar{x} \in E$ such that $P(\bar{x}) = \beta$.*

For applying Theorem 4, we fix a large $R > 1$ and define $E \subset H$ as being the closed ball $|x| \leq R$ equipped with the weak topology. The functional V is defined by (16). Theorem 4 says that $P(x) = p(x) + \frac{1}{2}|f - x|^2$ reaches its minimum α at $\bar{x}$. Then $Q(y)$ reaches its maximum β at $\bar{y}$ and $\bar{y}$ is defined by (15). We then have $f = \bar{x} + \bar{y}$ and Theorem 3 is proved. The details of this proof are left to the reader. For proving Theorem 2 it suffices to observe that $p(u) = \|u\|_{BV}$ by Lemma 3.

8 A model based on a Besov norm

We denote by $\mathcal{S}(\mathbf{R}^2)$ the Schwartz class and by $\mathcal{S}'(\mathbf{R}^2)$ the dual space of tempered distributions. A functional Banach space E is defined by the property $\mathcal{S}(\mathbf{R}^2) \subset E \subset \mathcal{S}'(\mathbf{R}^2)$ where the two embeddings are continuous ones. In general $\mathcal{S}(\mathbf{R}^2)$ is not dense in E and we let E_0 denote the closure in E of the space of testing functions. The space E^* is the dual space of E_0 and not of E in general. The norm in E^* is denoted by $\|\cdot\|_*$. If $E = BV$, then E_0 is defined by $f \in L^2$ and $\nabla f \in L^1$. The generalized ORF models we have in mind are based on the energy

$$\mathcal{K}(u) = \|u\|_E + \lambda\|f - u\|_2^2 \tag{19}$$

The optimal decomposition $f = \bar{u} + \bar{v}$ is the one which minimizes $\mathcal{K}(u)$. This optimal decomposition exists and is unique whenever the Banach space E has the following property: for every sequence $u_j \in E$, $j \in \mathbf{N}$, such that $\|u_j\|_E \leq 1$ and $\|u_j - u\|_2 \to 0$, $j \to \infty$, we have $\|u\|_E \leq 1$. With these notations Theorem 3 is still valid and $\bar{v}$ is the argument of $\inf\{\|f - v\|_2; \|v\|_* \leq \frac{1}{2\lambda}\}$.

The Besov space $Z = \dot{B}_1^{1,\infty}$ is close to BV as it will be proved below. Moreover this Besov space is isomorphic to a trivial sequence space and this isomorphism is given by the wavelet expansion. Let us consider now the variant of the ORF algorithm where the BV norm is replaced by the norm in Z. It will then be proved that the u component is obtained by a variant on the standard wavelet shrinkage.

The space BV can be defined by the existence of a constant C such that for every $y \in \mathbf{R}^2$ we have

$$\|f(\cdot - y) - f(\cdot)\|_1 \leq C|y| \tag{20}$$

The lower bound of these constants C is a norm which is equivalent to the usual BV norm of f. Similarly the homogeneous Besov space $Z = \dot{B}_1^{1,\infty}$ is defined by the existence of a constant C such that for every $y \in \mathbf{R}^2$ we have

$$\|f(\cdot - y) + f(\cdot + y) - 2f(\cdot)\|_1 \leq C|y| \tag{21}$$

The embedding $BV \subset Z$ is obvious and should be compared to the well known fact that a Lipschitz function always belongs to the Zygmund class, the converse not being true. It is therefore surprising that for a large collection of functions the two norms are equivalent. Here is the full story. Let $\mathcal{C}_N$ denote the collection of step functions with N levels. In other terms $f = c_1\chi_{E_1} + \cdots + c_N\chi_{E_N}$ where $E_1, \cdots, E_N$ are Borel sets and $c_1, \cdots, c_N$ are N real coefficients. We then have

Theorem 5 *There exist two positive constants C_0 and C_1 such that for every $f \in \mathcal{C}_N$ we have*

$$C_0\|f\|_Z \leq \|f\|_{BV} \leq C_1\, N\, \|f\|_Z \tag{22}$$

The proof of Theorem 5 relies on a deep theorem by Gérard Bourdaud [7]. If $\lambda \in \mathbf{R}$, we define $\theta_\lambda(t) = (t - \lambda)^+$. We then have

Theorem 6 *There exists a constant C_0 such that, for every $f \in Z$ and $\lambda \in \mathbf{R}$, we have*

$$\|\theta_\lambda(f)\|_Z \leq C_0\|f\|_Z \tag{23}$$

Before returning to Theorem 5, let us restate Theorem 6 under the following form:

Corollary 4 *If $b \geq a \geq 0$, we write $I = [a, b]$ and define a function $\tau_I(t)$ of the real variable t by $\tau_I(t) = 0$ if $t \leq a$, $\tau_I(t) = t - a$ if $a \leq t \leq b$, and finally $\tau_I(t) = b - a$ if $t \geq b$. Then we have*

$$\|\tau_I(f)\|_Z \leq 2C_0\|f\|_Z \tag{24}$$

Indeed $\tau_I(t) = \theta_a(t) - \theta_b(t)$. Then the proof of Theorem 5 begins with a simpler case:

Lemma 4 *Let $E \subset \mathbf{R}^2$ be a Borel set and let χ_E the indicator function of E. Then we have*

$$\frac{1}{2}\|\chi_E\|_Z \leq \|\chi_E\|_{BV} \leq \|\chi_E\|_Z \tag{25}$$

The proof of (25) relies on an obvious remark: If a, b, and c all belong to $\{0, 1\}$, then we have

$$|a - b| \leq |a - 2b + c| \tag{26}$$

The proof is trivial and left to the reader. Returning to (25), we apply (26) to $a = \chi_E(x + y), b = \chi_E(x)$, and $c = \chi_E(x - y)$. This yields the right-hand estimate in (25). The left-hand estimate is true for any function $f \in BV$.

We now treat the general case in Theorem 5. Without loosing generality, we can assume $c_1 < c_2 < \cdots < c_n$. The corollary of Theorem 6 is applied with $I_N = [c_{N-1}, c_N]$. We then obtain

$$\|(c_N - c_{N-1})\chi_{E_N}\|_Z \leq C_0\|f\|_Z \tag{27}$$

Since $\|\chi_E\|_{BV} \leq \|\chi_E\|_Z$, it implies

$$\|(c_N - c_{N-1})\chi_{E_N}\|_{BV} \leq C_0\|f\|_Z \tag{28}$$

We then consider $\Omega_k = E_k \cup \cdots \cup E_N = \{x; f(x) \geq c_k\}$. The corollary of Theorem 6 is applied to $I_k = [c_{k-1}, c_k]$ (with $c_0 = 0$ by convention) and we obtain

$$\|(c_k - c_{k-1})\chi_{\Omega_k}\|_{BV} \leq C_0 \|f\|_Z \tag{29}$$

But $f = \sum_1^N (c_k - c_{k-1})\chi_{\Omega_k}$ and the triangle inequality ends the proof.

Since the BV norm and the Besov norm are equivalent ones for simple functions, it is tempting to replace the standard ORF energy by

$$\mathcal{K}(u) = \|u\|_Z + \lambda \|v\|_2^2 \tag{30}$$

where $v = f - u$ as before.
A main difference with the standard ORF algorithm is coming from the fact that the Banach space Z is not contained in $L^2(\mathbf{R}^2)$. Instead Z is contained in the Lorentz space $L^{2,\infty}$. A typical example of a function in Z is $|x|^{-1}$ which is not square integrable. Good news are coming. The space Z admits a simple wavelet characterization. Let (ψ_1, ψ_2, ψ_3) be three functions in the Schwartz class such that the collection $\psi_{j,k} = 2^j \psi(2^j x - k)$, $j \in \mathbf{Z}$, $k \in \mathbf{Z}^2$, $\psi \in \{\psi_1, \psi_2, \psi_3\}$, is an orthonormal basis of L^2. We then have

Lemma 5 *A function $f \in L^{2,\infty}$ belongs to $Z = \dot{B}_1^{1,\infty}$ if and only if its wavelet coefficients $c(j,k) = \int f \psi_{j,k}\, dx$ satisfy*

$$\sup_{j \in \mathbf{Z}} \Big(\sum_{k \in \mathbf{Z}^2} |c(j,k)| \Big) = C < \infty \tag{31}$$

Similarly a tempered distribution f belongs to the dual space $Z^* = \dot{B}_\infty^{-1,1}$ if and only if its wavelet coefficients $c(j,k)$ satisfy

$$\sum_j \sup_k |c(j,k)| = C' < \infty \tag{32}$$

We remind the reader that Z^* is not the dual space of Z but rather the dual of the closure in Z of the space of testing functions. The norm of f in $\dot{B}_1^{1,\infty}$ and the infimum of C in (31) are equivalent norms and a similar remark applies to $\dot{B}_\infty^{-1,1}$ and C'. In other words the two Besov spaces $\dot{B}_1^{1,\infty}$ and $\dot{B}_\infty^{-1,1}$ are identified to simple sequence spaces. This is not the case for the space BV of functions of bounded variation. We have [17] and [18]:

Theorem 7 *With same notations, the wavelet coefficients $c(j,k)$ of a function $f \in BV$ belong to weak-l^1. More precisely there exists a constant C which only depends on the wavelet basis such that for every positive λ, we have*

$$\#\{(j,k) \in \mathbf{Z}^3; |c(j,k)| > \lambda\} \leq \frac{C}{\lambda}\|f\|_{BV} \tag{33}$$

However (33) does not characterize BV and (33) does not even imply $f \in \dot{B}_1^{1,\infty}$.

We now turn to the variant of the ORF algorithm. We write $Z = \dot{B}_1^{1,\infty}$ and the new energy is defined by

$$\mathcal{K}(u) = \|u\|_Z + \lambda\|f-u\|_2^2 \tag{34}$$

The optimal decomposition $f = \bar{u} + \bar{v}$ is the one which minimizes $\mathcal{K}(u)$. It exists and is unique. From now on we are cheating and stating that the norm in Z is not defined by (21) but instead by

$$\|u\|_Z = \sup_{j\in\mathbf{Z}} \sum_{k\in\mathbf{Z}^2} |c_u(j,k)| \tag{35}$$

which is an equivalent norm. Here $c_u(j,k)$ are the wavelet coefficients of u. With these notations $\mathcal{K}(u) = \|u\|_Z + \lambda\sigma(u)$ where $\sigma(u) = \sum_j\sum_k |c_f(j,k) - c_u(j,k)|^2$. This variant on the ORF algorithm leads to the following algorithm.

We want to find the sequence $y(j,k)$ which mimimizes

$$\sup_{j\in\mathbf{Z}}\Big(\sum_{k\in\mathbf{Z}^2}|y(j,k)|\Big) + \sum_j\sum_k |c_f(j,k) - y(j,k)|^2 \tag{36}$$

Then the wavelet coefficients of u are $c_u(j,k) = y(j,k),\ (j,k) \in \mathbf{Z}^3$. This optimization problem can be solved by the following procedure. Let a positive number η be given and let us assume that for each j, the sequence $y(j,k),\ k \in \mathbf{Z}^2$, is constricted by a limited budget defined by $\sum_{k\in\mathbf{Z}^2}|y(j,k)| \leq \eta$. We are asked to minimize $\sigma(u)$ within this budget limitation. This problem is named $\mathcal{P}(\eta)$. The minimum of $\sigma(u)$ is denoted by $\omega(\eta)$. Problem $\mathcal{P}(\eta)$ can be decoupled into a sequence of problems. Each one of them is standard and reads as follows: *We are given a sequence x_k, $k \in \mathbf{Z}$, and a positive number η. Then we want to minimize $\sum_k |x_k - y_k|^2$ under the constraint $\sum_k |y_k| \leq \eta$.* This is a variant on the standard wavelet shrinkage. Indeed y_k and x_k should have the same sign and satisfy $0 \leq |y_k| \leq |x_k|$. These two requirements say that the wavelet coefficients are shrunk towards 0. The second step consists in minimizing $\eta + \lambda\omega(\eta)$ over $\eta \in [0, \infty)$ and is left to the reader.

We said we are cheating. Indeed this approach can be questioned. When defined by (31), the norm in $\dot{B}_1^{1,\infty}$ depends on the wavelet basis. It is not translation and rotation invariant.

9 Analysis of textured images

As it was already said, the ORF algorithm is aimed at decomposing an image f into a sum $u + v$ where u represents the objects contained in f while v models the textured components. We shall challenge this working hypothesis with a collection of synthetic images which are explicitely given as a sum between a cartoon image g and a texture h. More precisely we consider a sum $f = g+h$ where g is the simplest sketch we can figure while h is an oscillating function modeling a textured component. For instance $h(x) = m(Nx)\chi(x)$ where $m(x)$ is a continuous function which is 2π-periodic in one variable and has a vanishing mean while χ is the indicator function of a rectifiable domain. We write $f_N(x) = g(x)+\chi(x)m(Nx)$ and N will be arbitrarily large. *Can the ORF algorithm be trusted? Does the ORF algorithm yield a decomposition $f_N = \bar{u}_N + \bar{v}_N$ where the sketch $\bar{u}_N$ is close to the original sketch g?* More precisely we expect $\bar{u}_N = g + \epsilon_N, \bar{v}_N = h - \epsilon_N$ where the L^2 norm of the error term ϵ_N tends to 0.

This fairy tale is untrue, since the ORF algorithm applied to $f = g$ does not get g back but instead a new function $\tilde{g}$. A more precise statement will be given in Lemma 10. The best to be expected in the general setting is $\bar{u}_N = \tilde{g} + \epsilon_N, \bar{v}_N = h + g - \tilde{g} - \epsilon_N$. This is true, as Theorem 8 will tell us.

Theorem 8 will apply to the ORF algorithm, but also to more general contexts. We begin with a Hilbert space H and $\|\cdot\|$ will denote the corresponding norm. We are given a dense subspace $V \subset H$ together with a norm which is finite on V and denoted by $\|\cdot\|_*$. If $x \notin V$, then $\|x\|_* = \infty$. Let us assume that the norm $\|\cdot\|_*$ is lower semi-continuous on H:

$$\|x_j\|_* \le 1 \text{ and } \lim_{j\to\infty} \|x - x_j\| = 0 \ \Rightarrow \|x\|_* \le 1 \tag{37}$$

Let us denote by $F \subset V$ the closed convex set defined by $\|\cdot\|_* \le 1$ and let $P_F : H \mapsto F$ be the orthogonal projection; $P_F(x) = z$ is the point in F which minimizes $\|z - x\|$. We know from Theorem 2 that the decomposition given by the ORF algorithm is $x = y + z,\ z = P_F(x)$. Writing $R_F = I - P_F$, we have $R_F(x) = y$. The main theorem of this section is the following:

Theorem 8 *For every $x \in H$, $x' \in H$ we have*

$$\|R_F(x') - R_F(x)\| \le 13(\|x\| + \|x'\|)\sqrt{\|x' - x\|_*} \tag{38}$$

Before proving Theorem 8, let us return to the ORF algorithm. Then F will be defined by $\|\cdot\|_* \le \frac{1}{2\lambda}$ where $\|\cdot\|_*$ is the dual norm as in Section 6. Theorem 8 reads

Corollary 5 *We consider the ORF model with a given value of the parameter λ. Let f_1 and f_2 be two functions in $L^2(\mathbf{R}^2)$. Let $f_j = u_j + v_j$ be the ORF decomposition of $f_j,\ j = 1, 2$. Then we have*

$$\|u_2 - u_1\|_2 \le 13(\|f_1\|_2 + \|f_2\|_2)\sqrt{\lambda\|f_2 - f_1\|_*} \tag{39}$$

We now assume that f_2 is a sum $f_2 = f_1 + h$ between a cartoon image f_1 and a textured component h satisfying $\|h\|_* \le \epsilon$. We then have

$$\|u_2 - u_1\|_2 \le 13(\|f_2\|_2 + \|f_1\|_2)\sqrt{\epsilon\lambda} \tag{40}$$

In other words the ORF algorithm does not perform what could have been dreamed, since the cartoon component f_1 is not preserved. Instead it is modified into u_1. This being said, the ORF algorithm acts in a consistent way: when it is applied to f_1 it yields $f_1 = u_1 + v_1$ and when it is applied to $f_2 = f_1 + h$ it yields $f_2 = u_2 + v_2$ where $u_2 = u_1 + O(\sqrt{\epsilon}),\ v_2 = v_1 + h - O(\sqrt{\epsilon})$.

Before proving these estimates, let us make a few remarks. The relevance of (38) comes from the fact that in many applications the norm which controls $x' - x$ is much weaker than the norm which is used in the left-hand side of (38). It is indeed the case for the ORF algorithm. Theorem 8 is not interesting when $\|x'-x\|_* \geq 1$. Indeed the mapping P_F is a contraction and R_F is Lipschitz. We have $\|R_F(x') - R_F(x)\| \leq 2\|x-x'\|$ which improves on (38). We will prove the optimality of the square root in (38). The counter-example which will be given below concerns the standard ORF algorithm. The conclusion in Theorem 8 does not apply to P_F. Otherwise we would have $\|x'-x\| \leq C\sqrt{\|x'-x\|_*}$ whenever $\|x\| \leq 1, \|x'\| \leq 1$, which is not true. A trivial counter-example is given by the usual ORF algorithm applied to $f_1 = 0$ and $f_2(x) = \cos(\omega \cdot x)\phi(x)$. Here ϕ is a smooth bump function and ω is arbitrarily large. Then $\|f_2\|_* \leq \frac{C}{|\omega|}$ while $\|f_2 - f_1\|_2 \geq c > 0$. Finally the weight given by $\|x\| + \|x'\|$ in the right-hand side of (38) cannot be erased. The simplest counter-example does not concern the standard ORF algorithm but the wavelet shrinkage. It is given by $H = l^2(\mathbf{N})$ and $\|x\|_* = \|x\|_\infty = \sup|x_n|$. Here $P_F(x) = \tilde{x}$ is defined by $\tilde{x}_n = x_n$ if $|x_n| \leq 1$ while $\tilde{x}_n = \mathrm{sign} x_n$ if not. In other words $R_F(x) = \hat{x}$ is defined by $\hat{x}_n = 0$ if $|x_n| \leq 1$ and $\hat{x}_n = x_n - \mathrm{sign} x_n$ if not. This operator R_F is a shrinkage which pulls the coefficients back to 0. We now check (38) on the two sequences (x_n) and (x'_n) defined by $x'_n = 1+\epsilon$ if $1 \leq n \leq N$ and $x'_n = 0$ if $n \geq N+1$. When $x_n = 1$ if $1 \leq n \leq N$ and $x_n = 0$ if $n \geq N+1$, we have $\|x'-x\|_\infty = \epsilon$ while $\|R_F(x') - R_F(x)\| = \epsilon\sqrt{N}$. The constant 13 is obviously not optimal.

The proof of Theorem 8 begins with a standard lemma

Lemma 6 *Let $F \subset H$ be a closed convex set. Let $x_0 \in H, z_0 = P_F(x_0),\ y_0 = x_0 - P_F(x_0)$, and $d = \|y_0\|$. If $x_0 = y_1 + z_1$ where $z_1 \in F$ and $\|y_1\| \leq d + \epsilon\|x_0\|$, then we have*

$$\|y_1 - y_0\| \leq 2\sqrt{\epsilon + \frac{\epsilon^2}{2}}\,\|x_0\| \tag{41}$$

The proof is standard and left to the reader.

From now on F is defined by $\|\cdot\|_* \leq 1$ and the other notations of Lemma 6 are kept. We then have

Lemma 7 *If $0 < \epsilon \leq 1,\ 0 < \eta \leq 1,\ x_0 = y_1 + z_1,\ \|y_1\| \leq d + \epsilon\|x_0\|,\ \|z_1\|_* \leq 1 + \eta$, then we have*

$$\|y_1 - y_0\| \leq 9\sqrt{\epsilon + \eta}\|x_0\| \tag{42}$$

For proving (42), we define $\tilde{z}_1 = (1+\eta)^{-1} z_1$ which ensures $\|\tilde{z}_1\|_* \leq 1$ and $\tilde{z}_1 \in F$. We then define $\tilde{y}_1 = x_0 - \tilde{z}_1$. By trivial algebra we have $\tilde{y}_1 = (1+\eta)^{-1} y_1 + \eta(1+\eta)^{-1} x_0$ and the triangle inequality implies

$$\|\tilde{y}_1\| \leq d + \beta\|x_0\|, \quad \beta = \frac{\epsilon+\eta}{1+\eta} - \frac{\eta d}{(1+\eta)\|x_0\|} \leq \frac{\epsilon+\eta}{1+\eta} \tag{43}$$

On the other hand, the definition of $\tilde{y}_1$ and the triangle inequality give

$$\|\tilde{y}_1 - y_1\| \leq \frac{\eta}{1+\eta}(2+\eta)\|x_0\| \tag{44}$$

Finally Lemma 6 is applied to the suboptimal decomposition $x_0 = \tilde{y}_1 + \tilde{z}_1$. This is the non-trivial ingredient of the proof. It yields

$$\|\tilde{y}_1 - y_0\| \leq 2\sqrt{\beta + \frac{\beta^2}{2}}\,\|x_0\| \tag{45}$$

Combining (44) with (45) yields (42).

We now prove Theorem 8 under the following form

Lemma 8 *If $x \in H, x' \in H$, $x = y + z$, $z = P_F(x)$, $x' = y' + z'$, $z' = P_F(x')$, $d = \|y\|$, $d' = \|y'\|$, and $\|x' - x\|_* \leq 1$, then*

$$\|y' - y\| \leq 13\sqrt{\|x' - x\|_*}(\|x\| + \|x'\|) \tag{46}$$

In Lemma 8, F is defined as in Theorem 8. We have $0 \in F$ which implies $\|y\| \leq \|x\|$. Indeed the trivial decomposition $x = x + 0$ is challenging the optimal decomposition $x = y + z$. We also have $\|y'\| \leq \|x'\|$. The proof of Lemma 8 starts with the following fact

$$|d' - d| \leq \frac{\epsilon}{1+\epsilon}(\|x\| + \|x'\|) \tag{47}$$

where $\epsilon = \|x - x'\|_*$.

For proving (47), one writes $x = x' + x - x' = y' + z' + x - x' = y'' + z''$ with $z'' = (1+\epsilon)^{-1}(z' + x - x')$ and $y'' = y' + \frac{\epsilon}{1+\epsilon}(z' + x - x')$. Since $\|z\|_* \leq 1$, the triangle inequality yields $\|z''\|_* \le 1$. It implies $\|y\| \le \|y''\|$ since $x = y+z$ is optimal. The definition of y'' and the triangle inequality yield

$$\|y''\| \le \|y'\| + \frac{\epsilon}{1+\epsilon}\|z' + x - x'\| \tag{48}$$

Since $y' = x' - z'$ and $\|y'\| \le \|x'\|$, we have $\|z' + x - x'\| \le \|x\| + \|x'\|$. Combining this estimate with $\|y\| \le \|y''\|$ and (48), we obtain

$$d \le d' + \frac{\epsilon}{1+\epsilon}(\|x\| + \|x'\|) \tag{49}$$

But x and x' are playing symmetric roles. We also have $d' \le d + \frac{\epsilon}{1+\epsilon}(\|x\| + \|x'\|)$ and (47) follows from these two inequalities.

We now return to $x = y + z$ and write $x = y' + w$ where $w = z' + x - x'$. Then $\|w\|_* \le 1 + \epsilon$ and $\|y'\| \le d + \frac{\epsilon}{1+\epsilon}(\|x\| + \|x'\|)$ by (47). Lemma 7 yields

$$\|y' - y\| \le 13(\|x\| + \|x'\|)\sqrt{\epsilon} \tag{50}$$

as announced.

10 Optimality in Theorem 8

Theorem 8 is optimal in the following sense. The square root in the RHS of (39) cannot be replaced by an exponent larger than 1/2. For proving this remark, let us denote by D the unit disc (centered at 0 with radius 1) and denote by D_ϵ the disc centered at 0, with radius $1+\epsilon$. Consider $f_1 = \chi_D, f_2 = \chi_{D_\epsilon}$. Then we have $\|f_2 - f_1\|_* \le \epsilon$. This follows from a more general observation given by the following lemma.

Lemma 9 *If g is a radial function and if $h(r) = \frac{1}{r}\int_0^r sg(s)\,ds$, then*

$$\|g\|_* = \|h\|_\infty \tag{51}$$

The proof of Lemma 9 is straightforward. It suffices to write $g(r) = \frac{\partial}{\partial x}(\frac{x}{r}h) + \frac{\partial}{\partial y}(\frac{y}{r}h)$ which follows from the definition of h. Then $g = \mathrm{div} H$, $H = (\frac{x}{r}h, \frac{y}{r}h)$ and $\|H\|_\infty = \|h\|_\infty$. This implies $\|g\|_* \le \|h\|_\infty$. The lower bound is much easier. The norm of a radial function $f(r)$ in $BV(\mathbf{R}^2)$ is simply $2\pi\int_0^\infty |f'(r)|r\,dr$.

A lower bound of the dual norm $\|g\|_*$ is given by $2\pi \sup |\int_0^\infty g(r)f(r)\,dr|$ where this supremum is computed on the collection of all testing functions satisfying $2\pi \int_0^\infty |f'(r)|r\,dr \leq 1$. This ends the proof of Lemma 9.

Lemma 9 implies the following result

Lemma 10 *Let χ_R denote the indicator function of the disc $|x| \leq R$. We then have $\|\chi_R\|_* = R/2$. The ORF decomposition of χ_R is given by $\chi_R = u_R + v_R$ where $u_R = 0$ if $\lambda R \leq 1$ and $u_R = (1 - \frac{1}{\lambda R})\chi_R, v_R = \frac{1}{\lambda R}\chi_R$ if $\lambda R > 1$.*

The first assertion of Lemma 10 is given by Lemma 9. The second one follows from the characterization of the ORF decomposition given by Theorem 2. We have $\|v_R\|_* = \frac{1}{2\lambda}$ and $\int u_R v_R\,dx = \|u_R\|_{BV}\|v_R\|_*$ which ends the proof.

We return to the optimality claim. Assuming $\lambda = 2$, the ORF decomposition of f_1 and f_2 are $f_1 = u_1 + v_1, f_2 = u_2 + v_2$ where $u_1 = \frac{1}{2}f_1, u_2 = (1 - \frac{1}{2(1+\epsilon)})f_2$. It implies $u_2 - u_1 = \frac{\epsilon}{2(1+\epsilon)}f_1 + \frac{1+2\epsilon}{2(1+\epsilon)}\chi_\epsilon$ where χ_ϵ is the indicator function of the annulus $1 \leq |x| \leq 1 + \epsilon$. Finally $\|u_2 - u_1\|_2 \geq \sqrt{\frac{\pi\epsilon}{2}}$ which implies the optimality.

11 Examples of textures

We still consider the original Osher-Rudin-Fatemi algorithm. Here and in what follows, $\|\cdot\|_*$ denotes the norm in the Banach space G which is the dual of $\dot{W}^{1,1}$. The ORF decomposition yields $u = 0$ (no object) whenever $\|f\|_* \leq \frac{1}{2\lambda}$. We now would like to better understand the meaning of this condition. More precisely we aim at showing the following

Proposition 2 *Oscillating patterns have a small $\|\cdot\|_*$ norm.*

This is still a vague statement since we did not define what an oscillating pattern is and the smallness condition in Proposition 2 does not mean anything. Proposition 3 will take care of these issues.

For the time being, let us introduce the following definition:

Definition 5 *Let μ be a non-negative Borel measure. We say that μ is a Guy David measure if there exists a constant C such that for every disc $D(x_0, R)$ centered at x_0 with radius R we have*

$$\mu(D(x_0, R)) \leq CR \tag{52}$$

For instance let Γ be a rectifiable curve in the plane and let σ denote the arc-length on Γ. Then σ is a Guy David measure if and only if Γ is Ahlfors regular. Guy David proved that this condition is equivalent to the boundedness of the Cauchy integral operator acting on $L^2(\Gamma; d\sigma)$. If μ is a signed Borel measure, we say that μ is a Guy David measure if and only if the measure $|\mu|$ satisfies (52) and the optimal C in (52) will then be denoted by $\|\mu\|_D$.

Understanding the dual norm $\|\cdot\|_*$ is much easier if one takes in account the following theorem

Theorem 9 *A non-negative Borel measure μ belongs to G if and only if μ is a Guy David measure.*

Let us stress that Theorem 9 does not hold in dimensions larger than 2. Before proving Theorem 9 let us stress its meaning. We need to prove the following estimate

$$|\int f\, d\mu| \leq C\|f\|_{BV}\|\mu\|_D \tag{53}$$

whenever f belongs to the Sobolev space $\dot{W}^{1,1}$. Let us stress that the LHS of (53) does not have a meaning if $f \in BV$. For instance the integral $\int f(x_1, 0)\, dx_1$ is meaningless if $f \in BV$. Indeed a function in BV does not have a trace on a line and is only defined almost everywhere. We now return to the proof. One way is obvious. It suffices to test (53) on a function f_R defined by $f_R(x) = 1$ on the disc $D(x_0, R)$ with $f_R(x) = 0$ outside the disc $D(x_0, 2R)$ together with $f_R(x) = \frac{2R-|x|}{R}$ in between. Then $\|f_R\|_{BV} = CR$ and $\int f_R\, d\mu \geq \mu(D(x_0, R))$, since μ is non-negative. This ends the easy part of the proof.

The other direction is more involved. It suffices to treat the case where $f(x) \geq 0$. For proving (53) we replace μ by $\mu \star \phi_\epsilon$ where $\phi_\epsilon(x) = \frac{1}{\epsilon^2}\phi(\frac{x}{\epsilon})$. The function ϕ is smooth with compact support and satisfies $\int \phi = 1$. This permits to extend (53) to any function $f \in BV$ and what is even more important it permits to use weak limits in BV. More precisely we can approach a function $f \in BV$ with compactly supported functions in BV. Next these functions are replaced by simple functions, as in Proposition 1. Let us now denote by f a compactly supported simple function. Then $f(x) = \sum_0^N c_j \chi_j(x)$ where $0 \leq c_0 < c_1 < \ldots < c_N$. Here χ_j is the indicator function of a set E_j which is a finite union of squares. We write $\Omega_j = E_j \cup \ldots \cup E_N$. Finally $\|f\|_{BV} = \sum_0^{N-1}(c_{j+1} - c_j) l_j$ where l_j is the length of the boundary $\partial\Omega_j$. Finally $\int f\, d\mu = \sum_0^{N-1} c_j \mu(E_j)$. But $E_j = \Omega_j \setminus \Omega_{j+1}$ which implies $\mu(E_j) = \mu(\Omega_j) - \mu(\Omega_{j+1})$. Using Abel's transformation we obtain $\int f\, d\mu = \sum_0^N (c_j - c_{j-1})\mu(\Omega_j)$. Here $c_{-1} = 0$. It then suffices to use the following *geometrical estimate* $\mu(\Omega) \leq \|\mu\|_D \mathcal{H}^1(\partial\Omega)$ which implies $\mu(\Omega_j) \leq C l_j \|\mu\|_D$ and $\int f\, d\mu \leq C\|f\|_{BV}\|\mu\|_D$. This is Theorem 9. For proving our estimate we write Ω as the finite union of its connected components Ω_m and we denote by L_m the length of $\partial\Omega_m$. Then Ω_m is contained in a disc D_m of radius L_m. It suffices to use (52) for D_m to prove the *geometrical estimate.*

Theorem 9 is not valid in other dimensions. Indeed a domain $\Omega \subset \mathbf{R}^3$ whith $\mathcal{H}^2(\partial\Omega) \leq 1$ can be unbounded.

Guy David measures are needed in understanding the pointwise multipliers of BV.

Definition 6 *We say that a measurable function $m(x)$ is a pointwise multiplier of BV if, for every $f \in BV$, the function $g(x) = m(x)f(x)$ also belongs to BV.*

The closed graph theorem yields the continuity of the mapping $M : f(x) \mapsto m(x)f(x)$. We then have for every $f \in BV$,

$$\|m(x)f(x)\|_{BV} \leq C\|f\|_{BV} \tag{54}$$

It is readily seen that (54) implies $m(x) \in L^\infty$. If $m(x) \in L^\infty$, it suffices to prove (54) when $f \in \dot{W}^{1,1}$. Indeed if $f \in BV$, there exists a sequence $f_j \in \dot{W}^{1,1}$ such that $\|f_j\|_{BV} \leq \|f\|_{BV}$ and $\lim \|f - f_j\|_2 = 0$ which ends the proof of our remark. If $f \in \dot{W}^{1,1}$ we have $\nabla(mf) = m\nabla f + f\nabla m$. Since $m \in L^\infty$, we have $m\nabla f \in L^1$ and we are led to estimating $\|f\nabla m\|_1$. Since $f \in \dot{W}^{1,1}$ implies $|f| \in \dot{W}^{1,1}$, everything ends with the computation of $\int |\nabla m| f\,dx$ when f is a non-negative function in $\dot{W}^{1,1}$. Then Theorem 9 implies the following

Theorem 10 *A measurable function $m(x)$ is a pointwise multiplier of BV if and only if $m \in L^\infty(\mathbf{R}^2)$ and ∇m is a Guy David measure.*

For instance let us consider the indicator function χ_Ω of a domain Ω with a rectifiable boundary. Then χ_Ω is a poinwise multiplier of BV if and only if $\partial\Omega$ is a Guy David curve.

We then apply Theorem 10 to texture analysis. The following corollary preludes the analysis of some textures which are present in some SPOT images.

Corollary 6 *Let us define the multiplier norm of m by $\|m\|_M = \|m\|_\infty + \|\nabla m\|_D$. Then there exists a constant C_0 such that*

$$\|m(x)f(x)\|_* \leq C_0 \|m\|_M \|f\|_* \tag{55}$$

for every $f \in L^2$.

Let us analyze a texture which appears in some SPOT images of the city of Toulouse. It concerns the periodic patterns of some roofs. These patterns are described by the following model. We begin with a real-valued function $\omega(x) = \omega(x_1, x_2) \in L^\infty(\mathbf{R}^2)$ which is α-periodic in $x_1 : \omega(x_1 + \alpha, x_2) = \omega(x_1, x_2)$. Here and in what follows, α is a small positive number. We also assume $\int_0^\alpha \omega(t, x_2)\,dt = 0$ identically in x_2. The latter assumption can be questioned and will be suppressed in a moment. This periodic function $\omega(x)$ is modeling the periodic pattern of the roof, but we have to delimit the position of the roof. This is taken care by the indicator function of the roof, or more generally by a function $m \in L^\infty$ such that ∇m is a Guy David measure. The model of the roof is given by the product $m(x)\omega(x)$. Then Bernstein's inequalities yield $\|\omega\|_* \leq \frac{\pi}{2}\alpha\|\omega\|_\infty$. Therefore the product $m(x)\omega(x)$ is a texture in a sense given by the following estimate:

Proposition 3 *With the preceding notations we have*

$$\|m(x)\omega(x)\|_* \leq C_0 \alpha \|\omega\|_\infty (\|m\|_\infty + \|\nabla m\|_D) \tag{56}$$

We now treat the full SPOT image $f(x)$ which is a sum

$$f(x) = f_0(x) + m(x)\omega(x) \tag{57}$$

where $f_0 \in L^2$ and $m(x)$, $\omega(x)$ are as above. We apply the ORF algorithm to f_0 and obtain

$$f_0 = u_0 + v_0 \tag{58}$$

We then keep the same value of the tuning parameter λ and we apply the ORF algorithm to f. Theorem 8 yields $f = u + v$ where $\|u - u_0\|_2 \leq C\sqrt{\alpha}$ and $\|v - (v_0 + \omega(x)m(x))\|_2 \leq C\sqrt{\alpha}$. Let us now compare two images f and f_0 of the city of Toulouse, the first one with the roof and the second one without it. The ORF algorithm splits f_0 into a sum $u_0 + v_0$ where v_0 is taking care of the texured components which were already present in the SPOT image. If we add the roof to f_0, the ORF algorihm will incorporate the full roof inside the texture as long as the mean value of the oscillating pattern is 0 (we here neglect an error term which is $O(\sqrt{\alpha})$). What would happen if ω is α periodic in x_1 but the mean value of ω is not zero ? We will see that the ORF algorithm does not incorporate the full roof into the texture component. Indeed we have $\omega = \omega_0 + \omega_1$ where ω_0 fulfills the vanishing mean condition and ω_1 does not depend on x_1. Then the product $\omega_1 m$ can be incorporated into f_0 and we have $f = f_0 + m\omega_1 + m\omega_0$. We can use the preceding discussion where f_0 and $m\omega_1$ are glued together. When the ORF algorithm is applied to the full image $f(x) = f_0(x) + \omega(x)m(x)$, it yields $f = u + v$ where $v(x) = v_0 + \omega_0(x)m(x) + R(x)$. This error term $R(x)$ is coming from the new 'object' $\omega_1 m$ which is now combined with f_0 and which does not stay inside u but spoils v, as explained by Corollary 3.

12 A new algorithm by Stanley Osher and Luminita Vese

This new algorithm is named the Osher-Vese algorithm. It is a variant on the standard ORF algorithm. The notations of the preceding sections are kept: G denotes the dual space of $\dot{W}^{1,1}$ and the norm in G is $\|\cdot\|_*$. Here the given image does not need to be a function in L^2 but can also be a tempered distribution belonging to G. We aim at splitting this image into a sum $u + v$ where the two components u and v have the same meaning as in the standard

ORF algorithm: the objects which are present in the given image f should belong to u and the textured components and the noise are incorporated in v. The functional we want to minimize is now $\Theta(u) = \|u\|_{BV} + \lambda\|v\|_*$. An optimal decomposition exists but is no longer unique in some instances. A counter-example will be described below. The role of the tuning parameter λ is different, as Lemma 12 tells us. We begin with an obvious remark:

Lemma 11 *For every function $f \in BV$ we have $\|f\|_2^2 \leq \|f\|_*\|f\|_{BV}$ and*

$$\|f\| \leq \frac{1}{2\sqrt{\pi}}\|f\|_2 \leq \frac{1}{4\pi}\|f\|_{BV} \tag{59}$$

The second estimate follows from the isoperimetric inequality and a simple duality argument yields the first one. The following observation is a trivial consequence of Lemma 11.

Lemma 12 *If $0 < \lambda \leq 4\pi$, the Osher-Vese algorithm yields $u = 0$, $v = f$.*

This should be compared with the similar statement concerning the ORF algorithm: $\|f\|_* \leq \frac{1}{2\lambda} \Rightarrow u = 0, v = f$. The proof of Lemma 12 is straightforward. We have $\|f\|_* \leq \|u\|_* + \|f-u\|_* \leq \frac{1}{4\pi}\|u\|_{BV} + \|f-u\|_*$. This obviously implies $\lambda\|f\|_* \leq \frac{\lambda}{4\pi}\|u\|_{BV} + \lambda\|f-u\|_*$. If $\lambda < 4\pi$, the energy of the trivial decomposition of f given by $f = 0 + f$ is smaller than $\Theta(u)$. It implies $u = 0$. If $\lambda = 4\pi$, $u = 0$ is still a solution but this solution is not unique in general. A counter-example will be given after the following definition.

Definition 7 *A function $f \in BV$ is extremal if $\|f\|_2^2 = \|f\|_*\|f\|_{BV}$*

The only example of an extremal function we have in mind is the indicator function of a disc. Let us write $a = \|f\|_*, b = \|f\|_2$, and $c = \|f\|_{BV}$. If $\lambda = c/a$, then all the decompositions $f = \alpha f + (1-\alpha)f$, $\alpha \in (0,1)$ have the same energy. This shows that the Osher-Vese decomposition is not unique in general.

We now prove that in some cases the Osher-Vese algorithm performs much better than the ORF algorithm. The following theorem is a first step which should be completed by more mathematics and more numerical experiments. We would like to thank Gilles Jerôme for providing us with numerous examples. We have

Theorem 11 *Let $f \in BV$ be extremal and let us assume $\lambda > c/a$ where $a = \|f\|_*$ and $c = \|f\|_{BV}$. Then the Osher-Vese algorithm yields $u = f, v = 0$. Let us now treat a more involved situation where $f = g + h$ with $\|h\|_* \leq \epsilon$, g being extremal. Let $a = \|g\|_*$, $c = \|g\|_{BV}$, and let us assume $\lambda > c/a$. Then the Osher-Vese algorithm yields a decomposition $f = u + v$ where u is close to g in L^2. More precisely*

$$\|u - g\|_* \leq \frac{2\epsilon\lambda}{\lambda - c/a}, \quad \|u - g\|_2 \leq C(\lambda)\sqrt{\epsilon} \tag{60}$$

This never happens with the standard ORF algorithm, since some significant pieces of the objects are always incorporated into the v component. The constant $C(\lambda) = C(\lambda, g)$ will be explicit. The proof of the first statement in Theorem 11 is extremely simple. One writes $\|f\|_2^2 = \int fu\,dx + \int fv\,dx$. Lemma 3 yields $|\int fu\,dx| \leq a\|u\|_{BV}$ which implies $b^2 \leq a\|u\|_{BV} + c\|v\|_*$. Since $b^2 = ac$, we have $c \leq \|u\|_{BV} + \frac{c}{a}\|v\|_* < \|u\|_{BV} + \lambda\|v\|_*$ unless $v = 0$. The energy of the trivial decomposition $f = f + 0$ is minimal and this ends the proof.

The proof of the second statement is similar. We first write

$$I = \int f(x)g(x)\,dx = \|g\|_2^2 + \int g(x)h(x)\,dx = I_1 + I_2 \tag{61}$$

We also have

$$I = \int u(x)g(x)\,dx + \int v(x)g(x)\,dx = I_3 + I_4 \tag{62}$$

Since g is extremal we have $I_1 = \|g\|_*\|g\|_{BV}$. Lemma 3 yields the following bounds $I_2 \geq -\|g\|_{BV}\|h\|_*$ and $I_3 + I_4 \leq \|g\|_*\|u\|_{BV} + \|g\|_{BV}\|h\|_*$. Therefore

$$\|g\|_*\|g\|_{BV} - \|g\|_{BV}\|h\|_* \leq \|g\|_*\|u\|_{BV} + \|g\|_{BV}\|h\|_* \tag{63}$$

It suffices to divide by $\|g\|_*$ to obtain

$$\|g\|_{BV} - \frac{c}{a}\|h\|_* \leq \|u\|_{BV} + \frac{c}{a}\|v\|_* \tag{64}$$

But we also have

$$\|u\|_{BV} + \lambda\|v\|_* \leq \|g\|_{BV} + \lambda\|h\|_* \tag{65}$$

since $f = u + v$ is assumed to be an optimal decomposition. Since $\lambda > c/a$, it suffices to combine (64) and (65) to obtain $\|v\|_* \le \frac{\lambda + c/a}{\lambda - c/a}\epsilon$. We have $u - g = h - v$ and $\|u - g\|_* \le 2\epsilon\frac{\lambda}{\lambda - c/a}$ by the triangle inequality. On the other hand $f = u + v$ is an optimal Osher-Vese decomposition and we have $\|u\|_{BV} \le \|g\|_{BV} + \lambda\epsilon$ which yields $\|u - g\|_{BV} \le 2\|g\|_{BV} + \lambda\epsilon$. Then $\int |u - g|^2\, dx \le \|u - g\|_{BV}\|u - g\|_* \le 2\epsilon\frac{\lambda}{\lambda - c/a}\|g\|_{BV} + \epsilon\lambda$. This ends the proof of the second assertion.

An example of an extremal function is the indicator function of a disc. It is clear that other examples should be studied. There are few extremal functions and that is why we instead propose what we call 'plain images'.

This section is aimed at proving that "plain images" have a trivial Osher-Vese decomposition. It explains the higher performances of the Osher-Vese algorithm, when compared with the ORF algorithm. We begin with defining plain images.

Definition 8 *A function $f \in BV$ is a plain image if there exists another function $g \in BV$ such that*

$$\int f(x)g(x)\, dx = \|f\|_{BV} \text{ and } \|g\|_* = 1 \tag{66}$$

In other words a function f is a plain image if there exists a second function $g \in BV$, $g \neq 0$, such that $\int f(x)g(x)\, dx = \|f\|_{BV}\|g\|_*$. For obtaining (66), it then suffices to replace g by αg for a suitable positive constant α.

If f is a plain image, the pair (f, g) is extremal in the sense of [] and we will say that g is dual to f. More precisely, let us apply the standard ORF algorithm to a given image $f \in BV$. If $\|f\|_* > \frac{1}{2\lambda}$, the ORF algorithm yields $f = u + v$ where $\|v\|_* = \frac{1}{2\lambda}$ and $\int u(x)v(x)\, dx = \|u\|_{BV}\|v\|_*$. Since $v = f - u \in BV$, it implies that u is a plain image. Conversely if u is a plain image and if v is dual to u, then $f = u + \frac{1}{2\lambda}v$ is the ORF decomposition of f. We can conclude:

Proposition 4 *Plain images coincide with sketches provided by the ORF algorithm applied to functions of bounded variation.*

Here are some other examples and counter-examples.

Let Ω be a bounded domain with a $\mathcal{C}^3$-boundary. Then the indicator function χ_Ω of Ω is a plain image. Indeed we consider the normal vector $\nu(x)$ at $x \in \partial\Omega$ and extend it to a $\mathcal{C}^2$-vector field $H(x)$ such that $\|H\|_\infty = 1$. We then define $g=$ div$H$ and we have $\|g\|_* \leq 1$. But

$$\int \chi g \, dx = -\int \nabla\chi \cdot H \, dx = \mathcal{H}^1(\partial\Omega) \tag{67}$$

It implies $\|g\|_* = 1$ and f is a plain image. Similarly a piecewise constant function with jump discontinuities across $\mathcal{C}^3$-boundaries is a plain image.

Here are two counter-examples. The indicator function of a polygon cannot be a plain image, as it is proved in []. A second counter-example is given by $f(x) = \exp(-|x|^2)$. Then $\int fg \, dx = \|f\|_{BV}$ and $\|g\|_* = 1$ imply $g(x) = |x|^{-1}$. This g cannot belong to BV. However the radial function φ defined by $\varphi(x) = 1$ when $|x| \leq 1$, $\varphi(x) = |x|^{-2}$ if $|x| \geq 1$, is a plain image. We now have:

Theorem 12 *Let f be a plain image and let g be dual to f. If $\lambda > \lambda_0 = \|g\|_{BV}$, then the corresponding Osher-Vese decomposition of f is the trivial decomposition given by $u = f$, $v = 0$. Let us now treat a more involved situation where $f = g + h$ with $\|h\|_* \leq \epsilon$, g being a plain image. Let $\tilde{g}$ be dual to g and $C_0 = \|\tilde{g}\|_{BV}$. If $\lambda > C_0$, then the Osher-Vese algorithm yields a decomposition $f = u + v$ where u is close to g in L^2. More precisely*

$$\|u - g\|_* \leq \frac{2\epsilon\lambda}{\lambda - C_0}, \quad \|u - g\|_2 \leq C(\lambda)\sqrt{\epsilon} \tag{68}$$

The proof is the same as in Theorem 11. We argue by contradiction. Let $f = u + v$ be an optimal decomposition minimizing $\mathcal{K}(u)$. Keeping the notations of Definition ??? we have

$$\|f\|_{BV} = \int fg \, dx = \int ug \, dx + \int vg \, dx = I_1 + I_2 \qquad ()$$

But $|I_1| \leq \|u\|_{BV}$ and $|I_2| \leq \|v\|_* \|g\|_{BV}$ by Lemma 3. It implies $\|f\|_{BV} \leq \|u\|_{BV} + \|v\|_* \|g\|_{BV} < \|u\|_{BV} + \lambda\|v\|_*$ unless $v = 0$. The trivial decomposition $f = f + 0$ is winning.

References

[1] E. Adelson, R. Hingorani, and E. Simoncelli. *Orthogonal pyramid transform for image coding.* Proceedings SPIE Conference on Visual Communications and Image Processing II, vol. 845, October 27-29 (1987) pp. 50-58.

[2] S.-I. Amari and J.-F. Cardoso. *Blind source separation -semiparametric statistical approach.* IEEE Trans. on Sig. Proc. **45** (11): 2692-2700, November 1997, Special issue on neural networks.

[3] L. Ambrosio. *Corso introduttivo alla teoria geometrica della Misura ed alle Superfici Minime.* Scuola Normale Superiore Lecture Notes, Pisa, 1997.

[4] A. Antoniadis and G. Oppenheim, eds. *Wavelets and Statistics.* Lecture Notes in Statistics, 103, XV^{eme} Rencontres Franco-Belges, Springer-Verlag (1995)

[5] G. Aubert and J-F. Aujol. *Modeling very oscillatory signals. Applications to image processing.* INRIA, no. 4878, (2003)

[6] J-F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Image decomposition: Application to textured images and SAR images. INRIA no. 4704, (2003)

[7] G. Bourdaud and M.E.D. Kateb. *Fonctions qui opèrent sur les espaces de Besov.* Math. Ann. 303 (1995) 653-675.

[8] G. Bourdaud, *Fonctions qui opèrent sur les espaces de Besov et de Triebel.* Ann. Inst. Henri Poincaré, Analyse non linéaire, 10 (1993) 413-422.

[9] P.J.Burt and E.H.Adelson. *The Laplacian pyramid as a compact code.* IEEE Trans. Commun. 31(4):532-540, April 1983.

[10] Andrea Braides. *Approximation of free discontinuity problems.* Lecture Notes in Mathematics no. 1694, Springer (1991).

[11] Emmanuel Candès and David Donoho. *Ridgelets: a key to higher-dimensional intermittency?* Phil. Trans. R. Soc. Lond. A (1999) no. 1760, 2495-2509.

[12] E. Candès and D. Donoho. *New tight frames of curvelets and optimal representation of objects with piecewise $\mathcal{C}^2$ singularities.* Communications on Pure and Applied Mathematics, Vol. LVII, (2004) 0219-0226.

[13] Albert Cohen. *Numerical analysis of wavelet methods.* Handbook of numerical analysis, P.G.Ciarlet and J.L.Lions eds. (1999).

[14] J.-F. Cardoso. *Blind signal separation: statistical principles.* Proceedings of the IEEE Special issue on blind identification and estimation **9** (10): 2009-2025, October 1998.

[15] René Carmona, Wen-Liang Hwang & Bruno Torrésani. *Practical time-frequency analysis* Academic Press (1998).

[16] Albert Cohen and Robert Ryan. *Wavelets and multiscale signal processing.* Chapman and Hall, London (1995).

[17] Albert Cohen, Ronald DeVore, Pencho Petrushev and Hong Xu. *Non-linear approximation and the space $BV(R^2)$.* American Journal of Mathematics, 121 (1999) 587-628.

[18] Albert Cohen, Wolfgang Dahmen, Ingrid Daubechies and Ronald DeVore. *Harmonic Analysis of the space BV.* (September 25, 2000)

[19] Ronald R. Coifman. *Adapted multiresolution analysis, Computation, Signal processing, and Operator theory.* Proceedings of the ICM, Kyoto (1990) pp. 879-888, Springer-Verlag.

[20] R. R. Coifman and M. V. Wickerhauser. *Entropy-based algorithms for best-basis selection.* IEEE Trans. Info. Theory, March 1992.

[21] Ingrid Daubechies. *Ten lectures on wavelets.* SIAM Philadelphia (1992).

[22] Guy David. *Singular sets of minimizers for the Mumford-Shah functional.* Birkhäuser (2004).

[23] Ron DeVore. *Nonlinear approximation.* Acta Numerica (1998) pp. 51-150.

[24] Ron DeVore, Björn Jawerth and Vasil Popov. *Compression of wavelet decompositions.* American Journal of Mathematics 114 (1992) pp. 737-785.

[25] Ron DeVore and Bradley J. Lucier.*Fast wavelet techniques for near optimal image compression.* 1992 IEEE Military Communications Conference October 11-14, (1992).

[26] David Donoho. *Nonlinear Solution of Linear Inverse Problems by Wavelet-Vaguelette Decomposition.* Applied and Computational Harmonic Analysis 2, 101-126 (1995).

[27] David Donoho. *Abstract statistical estimation and modern harmonic analysis.* Proceedings of the ICM, Zürich (1994) pp. 997-1005 Birkäuser.

[28] David Donoho, Ron DeVore, Ingrid Daubechies and Martin Vetterli. *Data compression and harmonic analysis.* IEEE Trans. on Information Theory, Vol. 44, No 6, October 1998, 2435-2476.

[29] David Donoho and Iain Johnstone. *Wavelet shrinkage: Asymptopia?* J.R.Statist. Soc. B (1995) 57 N 2 pp. 301-369.

[30] Dennis Gabor. *Theory of communication.* J. IEE, **93** (1946) pp. 429-457. 339-342.

[31] Hubel D. *Discurso de investidura de Doctor Honoris Causa.* Universidad Autonoma de Madrid. (1997)

[32] Stéphane Jaffard, Yves Meyer, and Robert D. Ryan. *Wavelets: Tools for Science and Technology.* SIAM (2001)

[33] Gilles Jérôme. *Ph.D.*

[34] G. Kanizsa. *La Grammaire du Voir.* Ed. Diderot, Arts & Sciences (1997)

[35] Stéphane Mallat. *A Wavelet Tour of Signal Processing.* Academic Press (1998).

[36] D. Marr. *Vision, A computational investigation into the human representation and processing of visual information.* W.H. Freeman and Co. (1982)

[37] Y. Meyer. *Oscillating patterns in image processing and in some nonlinear evolution equations.* (Lewis Memorial Lectures) AMS (2001)

[38] François G. Meyer. *Multi-layered image representation: application to image compression.* Notes de cours du Centre Emile Borel (Sept. 14-Dec. 18, 1998).

[39] Jean-Michel Morel and Sergio Solimini. *Variational methods in image segmentation.* Birkháuser, Boston (1995).

[40] David Mumford. *Book review on* [38]. Bulletin of the American Mathematical Society, Vol. 33, n.2, April 1996.

[41] David Mumford and Jayant Shah. *Boundary detection by minimizing functionals.* Proc. IEEE Conf. Comp. Vis. Pattern Recognition, 1985.

[42] D. Mumford and J. Shah. *Optimal representations by piecewise smooth functions and associated variational problems.* Comm. Pure Applied Mathematics, 42 (5) (1989) 577-685.

[43] Stanley Osher and Leonid Rudin. *Total variation based image restoration with free local constraints.* In Proc. IEEE ICIP, vol I, pages 31-35, Austin (Texas) USA, Nov.1994.

[44] Stanley Osher, Leonid Rudin and Emad Fatemi. *Nonlinear total variation based noise removal algorithms.* Physica D, 60 (1992) 259-268.

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

[46] P. Oswald. *On the boundedness of the mapping* $f \mapsto |f|$ *in Besov spaces.* Comment. Univ. Carolinae 33 (1992) 57-66.

[47] Jaak Peetre. *New thoughts on Besov Spaces.* Duke Univ. Math. Series (1976).

[48] Naoki Saito. Local Feature Extraction and its applications using a Library of Bases. in 'Topics in Analysis and its applications' (R. R. Coifman Ed.) World Scientific Pub. Co. (1999).

[49] Naoki Saito, B.M. Larson and B. Benichou. *Sparsity vs. statistical independence from a best basis view-point.* To appear in Wavelet Applications in Signal and Image Processing VIII (A. Aldroubi, A.F.Laine, and M.A. Unser eds) Proc. SPIE 4119 (2000), Invited Paper.

[50] Naoki Saito. *Least statistically-dependent basis and its application to image modeling.* Wavelet Applications in Signal and Image Processing VI (A. Aldroubi, A.F.Laine, and M.A. Unser eds) Proc. SPIE 3458 pp. 24-37 (1998).

[51] L. Vese and S. Osher. *Modeling textures with total variation minimization and oscillating patterns in image processing.* Journal of Scientific Computing 19 (2003) 553-572.

[52] L. Vese and S. Osher. ıImage denoising and decomposition with total variation minimization and oscillatory functions. Special issue on Mathematics and Image Analysis, Journal of Mathematical Imaging and Vision, 20, (2004) 7-18.

[53] Ying Nian Wu and Song Chun Zhu. *from local features to global perception.* Neurocomputing, 26-27, July 1999.

Ali Haddad & Yves Meyer
CMLA
Ecole normale supérieure de Cachan
94235 CACHAN Cedex
France
ymeyer@cmla.ens-cachan.fr