Astronomical Data Analysis and Sparsity: from Wavelets to Compressed Sensing

Jean-Luc Starck and Jerome Bobin,

***Abstract*—Wavelets have been used extensively for several years now in astronomy for many purposes, ranging from data filtering and deconvolution, to star and galaxy detection or cosmic ray removal. More recent sparse representations such ridgelets or curvelets have also been proposed for the detection of anisotropic features such cosmic strings in the cosmic microwave background. We review in this paper a range of methods based on sparsity that have been proposed for astronomical data analysis. We also discuss what is the impact of Compressed Sensing, the new sampling theory, in astronomy for collecting the data, transferring them to the earth or reconstructing an image from incomplete measurements.**

***Index Terms*—Astronomical data analysis, Wavelet, Curvelet, restoration, compressed sensing**

I. Introduction

The wavelet transform (WT) has been extensively used in astronomical data analysis during the last ten years. A quick search with ADS (NASA Astrophysics Data System, adswww.harvard.edu) shows that around 1000 papers contain the keyword "wavelet" in their abstract, and this holds for all astrophysical domains, from study of the sun through to CMB (Cosmic Microwave Background) analysis [29]. This broad success of the wavelet transform is due to the fact that astronomical data generally gives rise to complex hierarchical structures, often described as fractals. Using multiscale approaches such as the wavelet transform, an image can be decomposed into components at different scales, and the wavelet transform is therefore well-adapted to the study of astronomical data. Furthermore, since noise in the physical sciences is often not Gaussian, modeling in wavelet space of many kind of noise – Poisson noise, combination of Gaussian and Poisson noise components, non-stationary noise, and so on – has been a key motivation for the use of wavelets in astrophysics.

If wavelets represent well isotropic features, they are far from optimal for analyzing anisotropic objects. This has motived other constructions such the curvelet transform [9]. More generally, the best data decomposition is the one which leads to the sparsest representation, i.e. few coefficients have a large magnitude, while most of them are close to zero. Hence, for specific astronomical data set containing edges (planetary images, cosmic strings, etc), curvelets should be preferred to wavelets.

J.-L. Starck is with the Laboratoire AIM (UMR 7158), CEA/DSM-CNRS-Université Paris Diderot, IRFU, SEDI-SAP, Service d'Astrophysique, Centre de Saclay, F-91191 Gif-Sur-Yvette cedex, France.

J. Bobin is with the department of Applied and Computational Mathematics (ACM), California Institute of Technology, M/C 217-50, 1200 E.California, PASADENA CA-91125, USA.

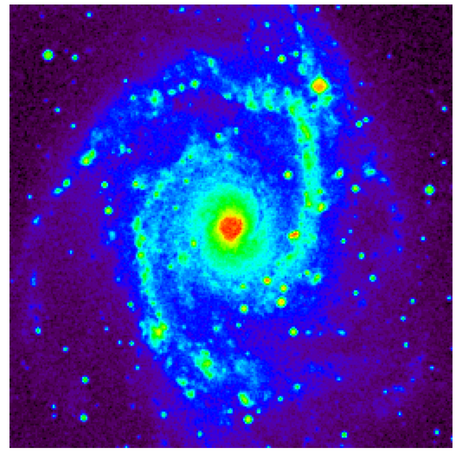

Fig. 1. Galaxy NGC 2997.

In this paper, we review a range of astronomical data analysis methods based on sparse representations. We first introduce the Undecimated Isotropic Wavelet Transform (UIWT) which is the most popular WT algorithm in astronomy. We show how the signal of interest can be detected in wavelet space using a noise modeling, allowing us to build the so-called *multiresolution support*. Then we present in III how this multiresolution support can used for restoration applications. In section IV, another representation, the curvelet transform, is introduced, which is well adapted to anisotropic structure analysis. Combined together, the wavelet and the curvelet transforms are very powerful to detect and discriminate very faint features. We give an example of application for cosmic string detection. Section V describes the compressed sensing theory which is strongly related to sparsity, and presents its impacts in astronomy, especially for spatial data compression.

II. The Isotropic Undecimated Wavelet Transform

The Isotropic undecimated wavelet transform (IUWT) [25] decomposes an $n \times n$ image c_0 into a coefficient set $W = \{w_1, \dots, w_J, c_J\}$, as a superposition of the form

$$c_0[k,l] = c_J[k,l] + \sum_{j=1}^{J} w_j[k,l],$$

where c_J is a coarse or smooth version of the original image c_0 and w_j represents 'the details of c_0' at scale 2^{-j} (see Starck et al.[30, 28] for more information). Thus, the algorithm outputs $J+1$ sub-band arrays of size $n \times n$. (The present indexing is such that $j = 1$ corresponds to the finest scale (high frequencies)).

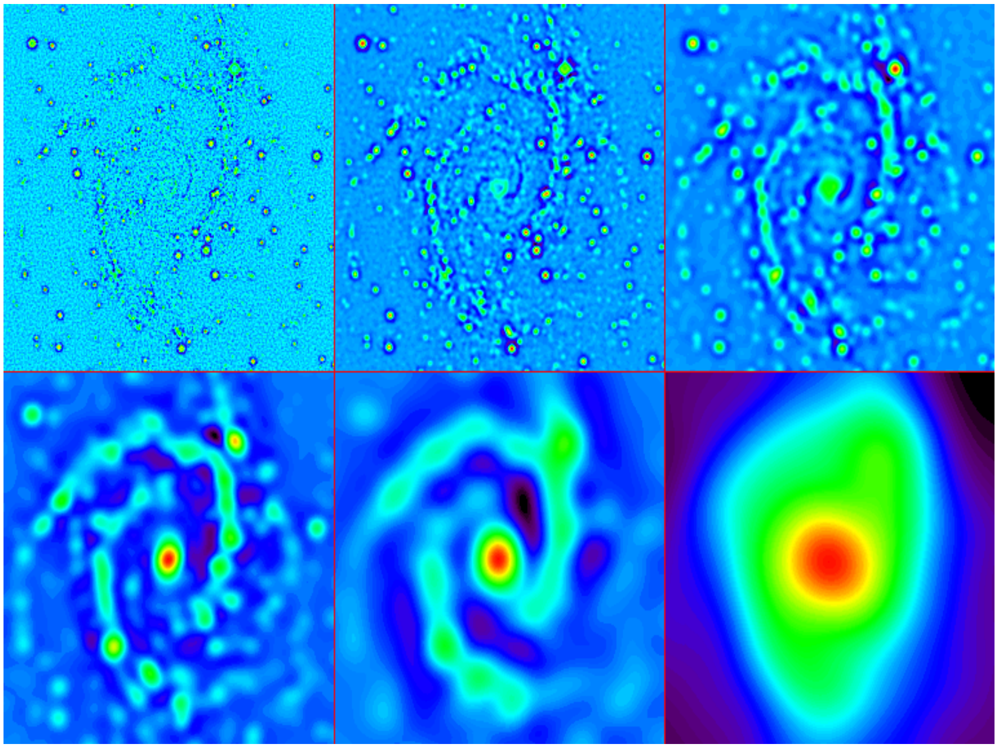

Fig. 2. Wavelet transform of NGC 2997 by the IUWT. The co-addition of these six images reproduces exactly the original image.

Hence, we have a *multi-scale pixel representation*, i.e. each pixel of the input image is associated with a set of pixels of the multi-scale transform. This wavelet transform is very well adapted to the detection of isotropic features, and this explains its success for astronomical image processing, where the data contain mostly isotropic or quasi-isotropic objects, such as stars, galaxies or galaxy clusters.

The decomposition is achieved using the filter bank $(h_{\mathrm{2D}}, g_{\mathrm{2D}} = \delta - h_{\mathrm{2D}}, \tilde{h}_{\mathrm{2D}} = \delta, \tilde{g}_{\mathrm{2D}} = \delta)$ where h_{2D} is the tensor product of two 1D filters h_{1D}. The passage from one resolution to the next one is obtained using the "à trous" algorithm [30]

$$
\begin{aligned}
c_{j+1}[k,l] &= \sum_m \sum_n h_{1D}[m] h_{1D}[n] c_j[k + 2^j m, l + 2^j n], \\
w_{j+1}[k,l] &= c_j[k,l] - c_{j+1}[k,l] \ , \qquad (1)
\end{aligned}
$$

where h_{1D} is typically a symmetric low-pass filter such as the B_3-Spline filter.

Fig. 2 shows IUWT of the galaxy NGC 2997 displayed in Fig. 1 . Five wavelet scales are shown and the final smoothed plane (lower right). The original image is given exactly by the sum of these six images.

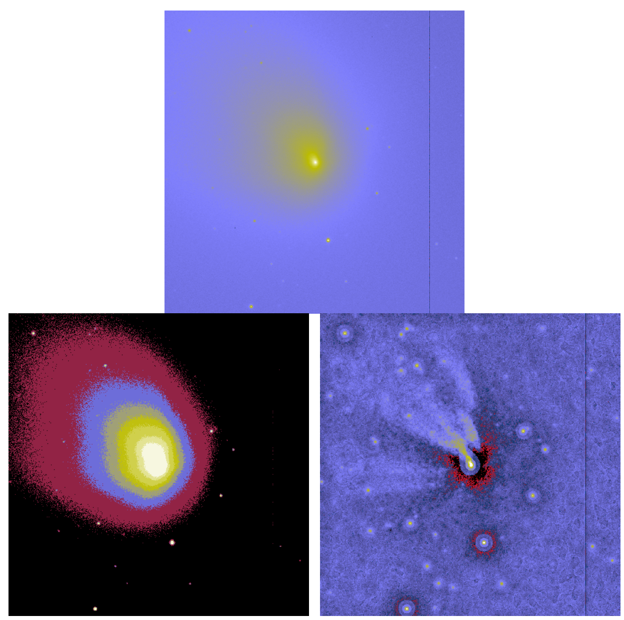

Fig. 3. Top left – Hale-Bopp Comet image. Bottom left – histogram equalization results. Bottom right – wavelet-log representations.

A. Example: Dynamic range compression using the IUWt

Since some features in an image may be hard to detect by the human eye due to low contrast, we often process the image before visualization. Histogram equalization is certainly one the most well-known methods for contrast enhancement. Images with a high dynamic range are also difficult to analyze. For example, astronomers generally visualize their images using a logarithmic look-up-table conversion.

Wavelets can be used to compress the dynamic range at all scales, and therefore allow us to clearly see some very faint features. For instance, the wavelet-log representation consists of replacing $w_j[k,l]$ by $\mathrm{sgn}(w_j[k,l]) \log(|w_j[k,l]|)$, leading to the alternative image

$$I_{k,l} = \log(c_{J,k,l}) + \sum_{j=1}^{J} \mathrm{sgn}(w_j[k,l]) \log(\mid w_j[k,l] \mid + \epsilon) \quad (2)$$

where ϵ is a small number (for example $\epsilon = 10^{-3}$). Fig. 3 shows a Hale-Bopp Comet image (logarithmic representation) (top left), its histogram equalization (middle row), and its wavelet-log representation (bottom). Jets clearly appear in the last representation of the Hale-Bopp Comet image.

B. Signal detection in the wavelet space

Observed data Y in the physical sciences are generally corrupted by noise, which is often additive and which follows in many cases a Gaussian distribution, a Poisson distribution, or a combination of both. It is important to detect the wavelet coefficients which are "significant", i.e. the wavelet coefficients which have an absolute value too large to be due to noise. We defined the multiresolution M of an image Y by:

$$M_j[k,l] = \begin{cases} 1 & \text{if } w_j[k,l] \text{ is significant} \\ 0 & \text{if } w_j[k,l] \text{ is not significant} \end{cases} \quad (3)$$

where $w_j[k,l]$ is the wavelet coefficient of Y at scale j and at position (k,l). We need now to determine when a wavelet coefficient is significant. For Gaussian noise, it is easy to derive an estimation of the noise standard deviation σ_j at scale j from the noise standard deviation, which can be evaluated with good accuracy in an automated way [27]. To detect the significant wavelet coefficients, it suffices to compare the wavelet coefficients $w_j[k,l]$ to a threshold level t_j. t_j is generally taken equal to $K\sigma_j$, and K is chosen between 3 and 5. The value of 3 corresponds to a probability of false detection of 0.27%. If $w_j[k,l]$ is small, then it is not significant and could be due to noise. If $w_j[k,l]$ is large, it is significant:

$$\begin{array}{ll} \text{if } \mid w_j[k,l] \mid \geq t_j & \text{then } w_j[k,l] \text{ is significant} \\ \text{if } \mid w_j[k,l] \mid < t_j & \text{then } w_j[k,l] \text{ is not significant} \end{array} \quad (4)$$

When the noise is not Gaussian, other strategies may be used:

- **Poisson noise:** if the noise in the data Y is Poisson, the transformation [3] $\mathcal{A}(Y) = 2\sqrt{I + \frac{3}{8}}$ acts as if the data arose from a Gaussian white noise model, with $\sigma = 1$, under the assumption that the mean value of I is sufficiently large. However, this transform has some limits and it has been shown that it cannot be applied for data with less than 20 photons per pixel. So for X-ray or gamma ray data, other solutions have to be chosen, which manage the case of a reduced number of events or photons under assumptions of Poisson statistics
- **Gaussian + Poisson noise:** the generalization of variance stabilization [18] is:

  $$\mathcal{G}((Y[k,l]) = \frac{2}{\alpha}\sqrt{\alpha Y[k,l] + \frac{3}{8}\alpha^2 + \sigma^2 - \alpha g}$$

  where α is the gain of the detector, and g and σ are the mean and the standard deviation of the read-out noise.
- **Poisson noise with few events using the MS-VST** For images with very few photons, one solution consists in using the Multi-Scale Variance Stabilization Transform (MSVST) [32]. The MSVST combines both the Anscombe transform and the IUWT in order to produce *stabilized* wavelet coefficients, i.e. coefficients corrupted by a Gaussian noise with a standard deviation equal to 1. In this framework, wavelet cofficients are now calculated by:

  $$\begin{array}{c} \text{IUWT} \\ + \\ \text{MS-VST} \end{array} \begin{cases} c_j = \sum_m \sum_n h_{1D}[m] h_{1D}[n] \\ \quad\quad c_{j-1}[k + 2^{j-1}m, l + 2^{j-1}n] \\ w_j = \mathcal{A}_{j-1}(c_{j-1}) - \mathcal{A}_j(c_j) \end{cases} \quad (5)$$

  where $\mathcal{A}_j$ is the VST operator at scale j defined by:

  $$\mathcal{A}_j(c_j) = b^{(j)} \sqrt{|c_j + e^{(j)}|} \quad (6)$$

  where the variance stabilization constants $b^{(j)}$ and $e^{(j)}$ only depends on the filter h_{1D} and the scale level j. They can all be pre-computed once for any given h [32]. The multiresolution support is computed from the MSVST coefficients, considering a Gaussian noise with a standard deviation equal to 1. This stabilization procedure is also invertible as we have:

  $$c_0 = \mathcal{A}_0^{-1} \left[\mathcal{A}_J(a_J) + \sum_{j=1}^{J} c_j \right] \quad (7)$$

For other kind of noise (correlated noise, non stationary noise, etc), other solutions have been proposed to derive the multiresolution support [29]. In next section, we show how the multiresolution support can be used for denoising and deconvolution.

III. Restoration using the Wavelet Transform

A. Denoising

The most used filtering method is the hard thresholding, which consists of setting to 0 all wavelet coefficients of Y which have an absolute value lower than a threshold t_j

$$\tilde{w}_j[k,l] = \begin{cases} w_j[k,l] & \text{if } \mid w_j[k,l] \mid > t_j \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

More generally, for a given sparse representation (wavelet, curvelet, etc) with its associate fast transform $\mathcal{T}_w$ and fast reconstruction $\mathcal{R}_w$, we can derive a hard thresholding denoising solution X from the data Y, by first estimating the multiresolution support M using a given noise model, and then calculating:

$$X = \mathcal{R}_w M \mathcal{T}_w Y. \quad (9)$$

We transform the data, multiply the coefficients by the support and reconstruct the solution.

The solution can however be improved considering the following optimization problem $\min_X \parallel M(\mathcal{T}_w Y - \mathcal{T}_w X) \parallel_2^2$ where M is the multiresolution support of Y. A solution can be obtained using the Landweber iterative scheme [22, 30]:

$$X^{n+1} = X^n + \mathcal{R}_w M \left[\mathcal{T}_w Y - \mathcal{T}_w X^n \right] \quad (10)$$

If the solution is known to be positive, the positivity constraint can be introduced using the following equation:

$$X^{n+1} = P_+ \left(X^n + \mathcal{R}_w M \left[\mathcal{T}_w Y - \mathcal{T}_w X^n \right] \right) \quad (11)$$

where P_+ is the projection on the cone of non-negative images.

This algorithm allows us to constraint the residual to have a zero value inside the multiresolution support [30]. For astronomical image filtering, iterating improves significantly the results, especially for the photometry (i.e. the integrated number of photons in a given object).

B. Deconvolution

In a deconvolution problem, $Y = HX + N$, when the sensor is linear, H is the block Toeplitz matrix. Similarly to the denoising problem, the solution can be obtained minimizing $\min_X \parallel M\mathcal{T}_w(Y - HX) \parallel_2^2$ under a positivity constraint, leading to the Landweber iterative scheme [22, 30]:

$$X^{n+1} = P_+ \left(X^n + H^t \mathcal{R}_w M \mathcal{T}_w \left[Y - HX^n \right] \right) \quad (12)$$

Only coefficients that belong to the multiresolution support are kept, while the others are set to zero [22]. At each iteration, the multiresolution support M can be updated by selecting new coefficients in the wavelet transform of the residual which have an absolute value larger than a given threshold.

Example

A simulated Hubble Space Telescope image of a distant cluster of galaxies is shown in Fig. 4, middle. The simulated data are shown in Fig. 4, left. Wavelet deconvolution solution is shown Fig. 4, right. The method is stable for any kind of point spread function, and any kind of noise modeling can be considered.

C. Inpainting

Missing data are a standard problem in astronomy. They can be due to bad pixels, or image area we consider as problematic due to calibration or observational problems. These masked area lead to many difficulties for post-processing, especially to estimate statistical information such the power spectrum or the bispectrum. The inpainting technique consists in filling the gaps. The classical image inpainting problem can be defined as follows. Let X be the ideal complete image, Y the observed incomplete image and L the binary mask (i.e. $L[k,l] = 1$ if we have information at pixel (k,l), $L[k,l] = 0$ otherwise). In short, we have: $Y = LX$. Inpainting consists in recovering X knowing Y and L.

Noting $||z||_0$ the l_0 pseudo-norm, i.e. the number of non-zero entries in z and $||z||$ the classical l_2 norm (i.e. $||z||^2 = \sum_k (z_k)^2$), we thus want to minimize:

$$\min_X \|\Phi^T X\|_0 \quad \text{subject to} \quad \parallel Y - LX \parallel_{\ell_2} \leq \sigma, \quad (13)$$

where σ stands for the noise standard deviation in the noisy case. It has also been shown that if X is sparse enough, the l_0 pseudo-norm can also be replaced by the convex l_1 norm (i.e. $||z||_1 = \sum_k |z_k|$) [14]. The solution of such an optimization task can be obtained through an iterative thresholding algorithm called MCA [15, 16] :

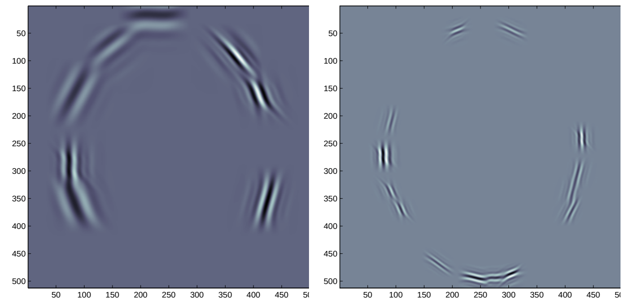

Fig. 6. A few first generation curvelets.

$$X^{n+1} = \Delta_{\Phi,\lambda_n}(X^n + Y - LX^n) \quad (14)$$

where the nonlinear operator $\Delta_{\Phi,\lambda}(Z)$ consists in:

- decomposing the signal Z on the dictionary Φ to derive the coefficients $\alpha = \Phi^T Z$.
- threshold the coefficients: $\tilde{\alpha} = \rho(\alpha, \lambda)$, where the thresholding operator ρ can either be a hard thresholding (i.e. $\rho(\alpha_i, \lambda) = \alpha_i$ if $|\alpha_i| > \lambda$ and 0 otherwise) or a soft thresholding (i.e. $\rho(\alpha_i, \lambda) = \mathrm{sign}(\alpha_i)\max(0, |\alpha_i| - \lambda)$). The hard thresholding corresponds to the l_0 optimization problem while the soft-threshold solves that for l_1.
- reconstruct $\tilde{Z}$ from the thresholds coefficients $\tilde{\alpha}$.

The threshold parameter λ_n decreases with the iteration number and it plays a role similar to the cooling parameter of the simulated annealing techniques, i.e. it allows the solution to escape from local minima. More details relative to this optimization problem can be found in [12, 16]. For many dictionaries such as wavelets or Fourier, fast operators exist to decompose the signal so that the iteration of eq. 14 is very fast. It requires only to perform at each iteration a forward transform, a thresholding of the coefficients and an inverse transform.

Example: The experiment was conducted on a simulated weak lensing mass map masked by a typical mask patterns (see Fig. 5). The left panel shows the simulated mass map and the middle panel show the masked map. The result of the inpainting method is shown in the right panel. We note that the gaps are undistinguishable by eye. More interesting, it has been shown that, using the inpainted map, we can reach an accuracy of about 1% for the power spectrum and 3% for the bispectrum [19].

IV. From Wavelet to Curvelet

The 2D curvelet transform [9] was developed in an attempt to overcome some limitations inherent in former multiscale methods e.g. the 2D wavelet, when handling smooth images with edges *i.e.* singularities along smooth curves. Basically, the curvelet dictionary is a multiscale pyramid of localized directional functions with anisotropic support obeying a specific parabolic scaling such that at scale 2^{-j}, its length is $2^{-j/2}$ and its width is 2^{-j}. This is motivated by the parabolic scaling property of smooth curves. Other properties of the

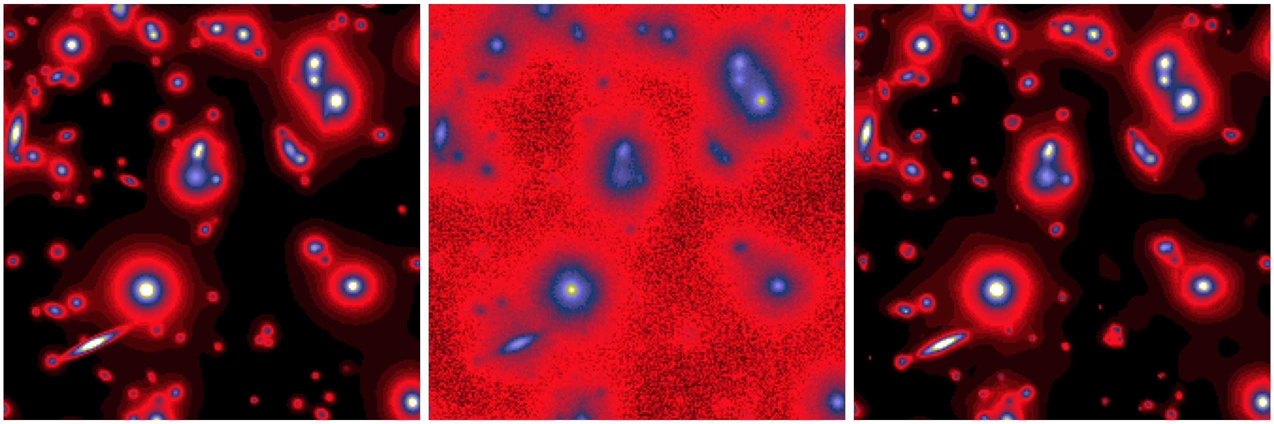

Fig. 4. Simulated Hubble Space Telescope image of a distant cluster of galaxies. Left: original, unaberrated and noise-free. middle: input, aberrated, noise added. Right, wavelet restoration wavelet.

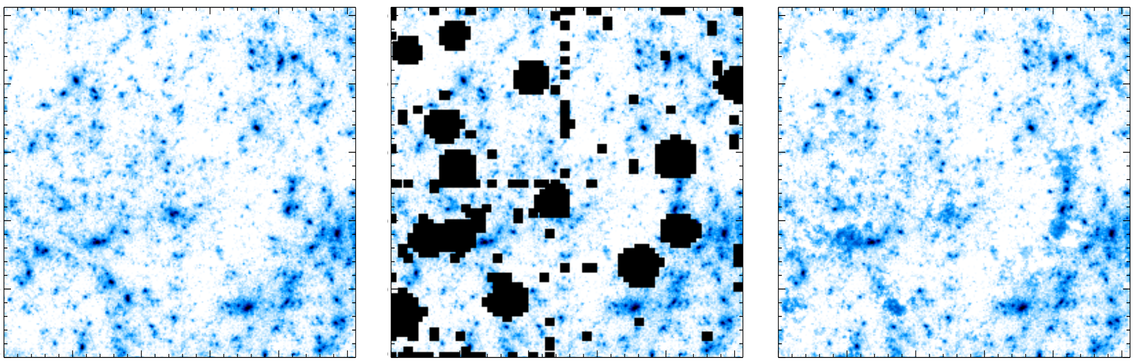

Fig. 5. Left panel, simulated weak lensing mass map, middle panel, simulated mass map with the mask pattern of CFHTLS data, right panels, inpainted mass map. The region shown is 1° x 1°.

curvelet transform as well as decisive optimality results in approximation theory are reported in [8]. Notably, curvelets provide optimally sparse representations of manifolds which are smooth away from edge singularities along smooth curves. Several digital curvelet transforms [23, 7] have been proposed which attempt to preserve the essential properties of the continuous curvelet transform and several papers report on their successful application in astrophysical experiments [24, 21, 26].

Fig. 6 shows a few curvelets at different scales, orientations and locations.

Application to the detection of cosmic strings

Some applications require the use of sophisticated statistical tools in order to detect a very faint signals, embedded in noise. An interesting case is the detection of non-Gaussian signatures in Cosmic Microwave Background (CMB). which is of great interest for cosmologists. Indeed, the non-Gaussian signatures in the CMB can be related to very fundamental questions such as the global topology of the universe [20], superstring theory, topological defects such as cosmic strings [6], and multi-field inflation [4]. The non-Gaussian signatures can, however, have a different but still cosmological origin. They can be associated with the Sunyaev-Zel'dovich (SZ) effect [31] (inverse Compton effect) of the hot and ionized intra-cluster gas of galaxy clusters [1], with the gravitational lensing by large scale structures, or with the reionization of the universe [1]. They may also be simply due to foreground emission, or to non-Gaussian instrumental noise and systematics .

All these sources of non-Gaussian signatures might have different origins and thus different statistical and morphological characteristics. It is therefore not surprising that a large number of studies have recently been devoted to the subject of the detection of non-Gaussian signatures. In [2, 21], it was shown that the wavelet transform was a very powerful tool to detect the non-Gaussian signatures. Indeed, the excess kurtosis (4th moment) of the wavelet coefficients outperformed all the other methods (when the signal is characterized by a non-zero 4th moment).

Finally, a major issue of the non-Gaussian studies in CMB remains our ability to disentangle all the sources of non-Gaussianity from one another. It has been shown it was possible to separate the non-Gaussian signatures associated with topological defects (cosmic strings) from those due to the Doppler effect of moving clusters of galaxies (i.e. the kinetic Sunyaev-Zel'dovich effect), both dominated by a Gaussian CMB field, by combining the excess kurtosis derived from both the wavelet and the curvelet transforms [21].

The wavelet transform is suited to spherical-like sources of non-Gaussianity, and a curvelet transform is suited to structures representing sharp and elongated structures such as cosmic strings. The combination of these transforms highlights the presence of the cosmic strings in a mixture CMB+SZ+CS. Such a combination gives information about the nature of the non-Gaussian signals. The sensitivity of each transform to a particular shape makes it a very strong discriminating tool [21, 17].

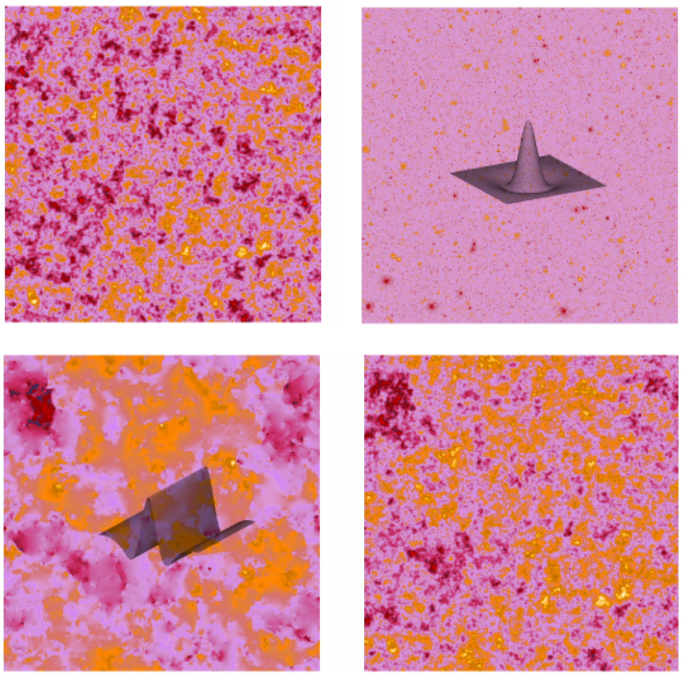

Fig. 7. Top, primary Cosmic Microwave Background anisotropies (left) and kinetic Sunyaev-Zel'dovich fluctuations (right). Bottom, cosmic string simulated map (left) and simulated observation containing the previous three components (right). The wavelet function is overplotted on the Sunyaev-Zel'dovich map and the curvelet function is overplotted on the cosmic string map.

In order to illustrate this, we show in Fig. 7 a set of simulated maps. Primary CMB, kinetic SZ and cosmic string maps are shown respectively in Fig. 7 top left, top right and bottom left. The "simulated observed map", containing the three previous components, is displayed in Fig. 7 bottom right. The primary CMB anisotropies dominate all the signals except at very high multipoles (very small angular scales). The wavelet function is overplotted on the kinetic Sunyaev-Zel'dovich map and the curvelet function is overplotted on cosmic string map.

V. Compressed Sensing

A. Compressed Sensing in a nutshell

Compressed sensing (CS) [10, 13] is a new sampling/compression theory based on the revelation that one can exploit sparsity or compressibility when acquiring signals of general interest, and that one can design nonadaptive sampling techniques that condense the information in a compressible signal into a small amount of data. The gist of Compressed Sensing (CS) relies on two fundamental properties :

1) *Compressibility of the data :* The signal X is said to be *compressible* if it exists a dictionary Φ where the coefficients $\alpha = \Phi^T X$, obtained after decomposing X on Φ, are sparsely distributed.
2) *Acquiring incoherent measurements :* In the Compressed Sensing framework, the signal X is not acquired directly; one then acquires a signal X by collecting data of the form $Y = AX + \eta$: A is an $m \times n$ (with $m < n$) "sampling" or measurement matrix, and η is a noise term. Assuming X to be sparse, the incoherence of A and Φ (*e.g.* the Fourier basis and the Dirac basis) entails that the information carried by X is diluted in all the measurements Y. Combining the incoherence of A and Φ with the sparsity of X in Φ makes the decoding problem tractable.

In the following, we choose the measurement matrix A to be a submatrix of an orthogonal matrix Θ : the resulting measurement matrix is denoted Θ_Λ and obtained by picking a set of columns of Θ indexed by Λ; Θ_Λ is obtained by subsampling the transformed signal ΘX. In practice, when Θ admits a fast implicit transform (*i.e.* discrete Fourier transform, Hadamard transform, noiselet transform), the compression step is very fast and made reliable for on-board satellite implementation.

A standard approach in CS attempts to reconstruct X by solving

$$\min_\alpha \|\alpha\|_{\ell_1} \text{ s. t. } \|Y - \Theta_\Lambda \Phi \alpha\|_{\ell_2} < \epsilon \tag{15}$$

where ϵ^2 is an estimated upper bound on the noise power.

B. Compressed sensing for the Herschel data

The Herschel/PACS mission of the European Space Agency (ESA) is facing with a strenuous compression dilemma : it needs a compression rate equal to $\rho = 1/N$ with $N = 6$. A first approach has been proposed which consists in averaging $N = 6$ consecutive images of a raster scan and transmitting the final average image. Nevertheless, doing so with high speed raster scanning leads to a dramatic loss in resolution. In [5], we emphasized on the redundancy of raster scan data : 2 consecutive images are almost the same images up to a small shift δ. Then, jointly compressing/decompressing consecutive images of the same raster scan has been put forward to alleviate the Herchel/PACS compression dilemma. The problem then consists in recovering a single image X from N compressed and shifted noisy versions of X :

$$\forall i \in \{1, \cdots, N\}; \quad X_i = \mathcal{T}_{\delta_i}(X) + \eta_i \tag{16}$$

where $\mathcal{T}_{\delta_i}$ is an operator that shifts the original image X with a shift δ_i. The term η_i models instrumental noise or model imperfections. According to the compressed sensing framework, each signal is projected onto the subspace ranged by Θ. Each compressed observation is then obtained as follows :

$$\forall i \in \{1, \cdots, N\}; \quad Y_i = \Theta_{i\Lambda_i} X_i \tag{17}$$

where the sets $\{\Lambda_i\}$ are such that the union of all the measurement matrices $[\Theta_{\Lambda_1}, \cdots, \Theta_{\Lambda_1}]$ span $\mathbb{R}^n$. In practice, the subsets Λ_i are disjoint and have a cardinality $m = \lfloor n/N \rfloor$. When there is no shift between consecutive images, these conditions guarantee that the signal X can be reconstructed univocally from $\{Y_i\}_{i=1,\cdots,N}$, up to noise. The decoding step amounts to seeking the signal x as follows :

$$\min_{\alpha} \|\alpha\|_{\ell_1} \text{ s. t. } \sum_{i=1}^{N} \|Y_i - \Theta_{\Lambda_i} \Phi \alpha\|_{\ell_2} < \sqrt{N}\epsilon \qquad (18)$$

The solution of this optimization problem can be found via an iterative thresholding algorithm (see [5]) :

$$X^{n+1} = \Delta_{\Phi,\lambda_n}\left(X^n + \mu_\Theta \sum_{i=1}^{N} \Theta_{\Lambda_i}^T \left(Y_i - \Theta_{\Lambda_i} X^n\right)\right) \qquad (19)$$

where the nonlinear operator $\Delta_{\Phi,\lambda}(Z)$ is defined in Equation 14 and the step-size $\mu_{\Theta_\Lambda} < 2/\sum_i \|\Theta_{\Lambda_i}^T \Theta_{\Lambda_i}\|_2$. Similarly to the MCA algorithm, the threshold λ_n decreases with the iteration number towards the final value : λ_f; a typical value is $\lambda_f = 2 - 3\sigma$. This algorithm has been shown to be very efficient for solving the problem in Equation 15 in [5].

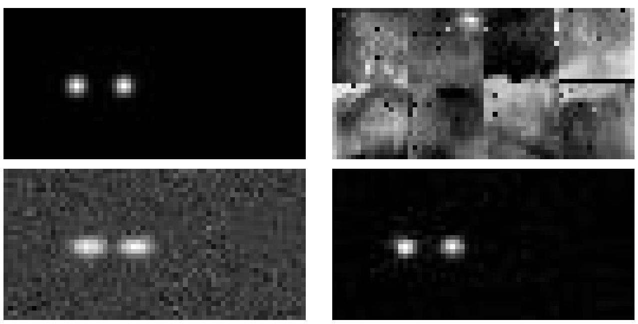

Fig. 8. Top left : Original image. Top right : First input noisy map (out of 6). The PACS data already contains approximately Gaussian noise. Bottom left : Mean of the 6 input images. Bottom right : Reconstruction from noiselet-based CS projections. The iterative algorithm has been used with 100 iterations.

a) Illustration: We compare two approaches to solve the Herschel/PACS compression problem : i) transmitting the average of 6 consecutive images (MO6), ii) compressing 6 consecutive images of a raster scan and decompressing using Compressed Sensing. Real Herschel/PACS data are complex : the original datum X is contaminated with a slowly varying "flat field" component c_f. In a short sequence of 6 consecutive images, the flat field component is almost fixed. In this context, the data $\{x_i\}_{i=0,\cdots,1}$ can then be modeled as follows :

$$X_i = \mathcal{T}_{\delta_i}(X) + \eta_i + c_f \qquad (20)$$

If c_f is known (which will be the case in the forthcoming experiments), $\mathcal{T}_{\delta_i}\left(X^{(n)}\right)$ is replaced by $\mathcal{T}_{\delta_i}\left(X^{(n)}\right) + c_f$ in Equation 19. The data have been designed by adding realistic pointwise sources to real calibration measurements performed in mid-2007. In the following experiment, the sparsifying dictionary is Φ is an undecimated wavelet tight frame and the measurement matrices are submatrices of the noiselet basis [11].

The top-left picture of Figure 8 features the original signal X. In the top-right panel of Figure 8. The "flat field" component overwhelms the useful part of the data so that x has at best a level that is 30 times lower than the "flat field" component. The MO6 solution (*resp.* the CS-based solution) is shown on the left (*resp.* right) and at the bottom of Figure 8. We showed in [5] that Compressed Sensing provides a resolution enhancement that can reach 30% of the FWHM of the instrument's PSF for a wide range of signal intensities (*i.e.* flux of X).

This experiment illustrates the reliability of the CS-based compression to deal with real-world data compression. The efficiency of Compressed Sensing applied to the Herschel/PACS data compression relies also on the redundancy of the data : consecutive images of a raster scan are fairly shifted versions of a reference image. The good perfomances of CS is obtained by merging the information of consecutive images. The same **data fusion** scheme could be used to reconstruct with high accuracy wide sky areas from full raster scans.

VI. Conclusion

By establishing a direct link between sampling and sparsity, compressed sensing had a huge impact in many scientific fields, especially in astronomy. We have seen that CS could offer an elegant solution to the Herschel data transfer problem. By emphasing so rigorously the importance of sparsity, compressed sensing has also shed light on all work related to sparse data representation (such as the wavelet transform, curvelet transform, etc.). Indeed, a signal is generally not sparse in direct space (i.e. pixel space), but it can be very sparse after being decomposed on a specific set of functions. For inverse problems, compressed sensing gives a strong theoretical support for methods which seek a sparse solution, since such a solution may be (under appropriate conditions) the exact one. Similar results are hardly accessible with other regularization methods. This explain why wavelets and curvelets are so successful for astronomical image denoising, deconvolution and inpainting.

Acknowledgment

This work was partially supported by the French National Agency for Research (ANR -08-EMER-009-01).

References

[1] N. Aghanim and O. Forni. Searching for the non-Gaussian signature of the CMB secondary anisotropies. Astronomy and Astrophysics, 347:409–418, July 1999.

[2] N. Aghanim, M. Kunz, P. G. Castro, and O. Forni. Non-Gaussianity: Comparing wavelet and Fourier based methods. Astronomy and Astrophysics, 406:797–816, August 2003.

[3] F.J. Anscombe. The transformation of Poisson, binomial and negative-binomial data. Biometrika, 15:246–254, 1948.

[4] F. Bernardeau and J. Uzan. Non-Gaussianity in multifield inflation. Physical Review D, 66:103506–+, November 2002.

[5] J. Bobin, J.-L. Starck, and R. Ottensamer. Compressed Sensing in Astronomy. ArXiv e-prints, 802, February 2008.

[6] F. R. Bouchet, D. P. Bennett, and A. Stebbins. Patterns of the cosmic microwave background from evolving string networks. Nature, 335:410, 1988.

[7] E. Candès, L. Demanet, D. Donoho, and L. Ying. Fast discrete curvelet transforms. SIAM Multiscale Model. Simul., 5/3:861–899, 2006.

[8] E. J. Candès and D. L. Donoho. Curvelets – a surprisingly effective nonadaptive representation for objects with edges. In A. Cohen, C. Rabut, and L.L. Schumaker, editors, Curve

and Surface Fitting: Saint-Malo 1999, Nashville, TN, 1999. Vanderbilt University Press.

[9] E.J. Candès and D. Donoho. Ridgelets: the key to high dimensional intermittency? Philosophical Transactions of the Royal Society of London A, 357:2495–2509, 1999.

[10] Emmanuel Candès, Justin Romberg, , and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Trans. on Information Theory, 52(2):489–509, 2006.

[11] R. Coifman, F. Geshwind, and Y. Meyer. Noiselets. Appl. Comput. Harmon. Anal., 10(1):27–44, 2001.

[12] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. SIAM Journal on Multiscale Modeling and Simulation, 4(4):1168–1200, 2005.

[13] D. Donoho. Compressed sensing. IEEE Trans. on Information Theory, 52(4):1289–1306, 2006.

[14] D.L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. IEEE Transactions on Information Theory, 47:2845–2862, 2001.

[15] M. Elad, J.-L. Starck, P. Querre, and D.L. Donoho. Simultaneous Cartoon and Texture Image Inpainting using Morphological Component Analysis (MCA). J. on Applied and Computational Harmonic Analysis, 19(3):340–358, 2005.

[16] M.J. Fadili, J.-L Starck, and F. Murtagh. Inpainting and zooming using sparse representations. The Computer Journal, 2006. submitted.

[17] J. Jin, J.-L. Starck, D.L. Donoho, N. Aghanim, and O. Forni. Cosmological non-gaussian signatures detection: Comparison of statistical tests. Eurasip Journal, 15:2470–2485, 2005.

[18] F. Murtagh, J.-L. Starck, and A. Bijaoui. Image restoration with noise suppression using a multiresolution support. Astronomy and Astrophysics, Supplement Series, 112:179–189, 1995.

[19] S. Pires, J. . Starck, A. Amara, R. Teyssier, A. Refregier, and J. Fadili. FASTLens (FAst STatistics for weak Lensing) : Fast method for Weak Lensing Statistics and map making. AA, 2009. in press.

[20] A. Riazuelo, J.-P. Uzan, R. Lehoucq, and J. Weeks. Simulating cosmic microwave background maps in multi-connected spaces. astro-ph/0212223, 2002.

[21] J.-L. Starck, N. Aghanim, and O. Forni. Detecting cosmological non-gaussian signatures by multi-scale methods. Astronomy and Astrophysics, 416:9–17, 2004.

[22] J.-L. Starck, A. Bijaoui, and F. Murtagh. Multiresolution support applied to image filtering and deconvolution. CVGIP: Graphical Models and Image Processing, 57:420–431, 1995.

[23] J.-L. Starck, E. Candès, and D.L. Donoho. The curvelet transform for image denoising. IEEE Transactions on Image Processing, 11(6):131–141, 2002.

[24] J.-L. Starck, E. Candes, and D.L. Donoho. Astronomical image representation by the curvelet tansform. AA, 398:785–800, 2003.

[25] J.-L. Starck, M.J. Fadili, and F. Murtagh. The Undecimated Wavelet Decomposition and its Reconstruction. IEEE Transactions on Image Processing, 16(2):297–309, 2007.

[26] J.-L. Starck, Y. Moudden, P. Abrial, and M. Nguyen. Wavelets, ridgelets and curvelets on the sphere. Astronomy and Astrophysics, 446:1191–1204, 2006.

[27] J.-L. Starck and F. Murtagh. Automatic noise estimation from the multiresolution support. Publications of the Astronomical Society of the Pacific, 110:193–199, 1998.

[28] J.-L. Starck and F. Murtagh. Astronomical Image and Data Analysis. Springer-Verlag, 2002.

[29] J.-L. Starck and F. Murtagh. Astronomical Image and Data Analysis. Astronomical image and data analysis, by J.-L. Starck and F. Murtagh. Astronomy and astrophysics library. Berlin: Springer, 2006, 2006.

[30] J.-L. Starck, F. Murtagh, and A. Bijaoui. Image Processing and Data Analysis: The Multiscale Approach. Cambridge University Press, 1998.

[31] R. A. Sunyaev and I. B. Zeldovich. Microwave background radiation as a probe of the contemporary structure and history of the universe. Annual Review of Astronomy and Astrophysics, 18:537–560, 1980.

[32] B. Zhang, M.J. Fadili, and J.-L. Starck. Wavelets, ridgelets and curvelets for poisson noise removal. IEEE Transactions on Image Processing, 17(7):1093–1108, 2008.

Jean-Luc Starck Jean-Luc Starck has a Ph.D from University Nice-Sophia Antipolis and an Habilitation from University Paris XI. He was a visitor at the European Southern Observatory (ESO) in 1993, at UCLA in 2004 and at Stanford's statistics department in 2000 and 2005. He has been a Researcher at CEA since 1994. His research interests include image processing, statistical methods in astrophysics and cosmology. He is an expert in multiscale methods such wavelets and curvelets, He is leader of the project Multiresolution at CEA and he is a core team member of the PLANCK ESA project. He has published more than 100 papers in different areas in scientific journals. He is also author of two books entitled Image Processing and Data Analysis: the Multiscale Approach (Cambridge University Press, 1998), and Astronomical Image and Data Analysis (Springer, 2nd edition, 2006).

Jerome Bobin Jrme Bobin graduated from the Ecole Normale Superieure (ENS) de Cachan, France, in 2005 and received the M.Sc. degree in signal and image processing from ENS Cachan and Universit Paris XI, Orsay, France. He received the Agrgation de Physique in 2004. Since 2005, he is pursuing his Ph.D with J-L.Starck at the CEA. His research interests include statistics, information theory, multiscale methods and sparse representations in signal and image processing.