

Astronomical Data Analysis and Sparsity: From Wavelets to Compressed Sensing

Correct interpretation of astronomical images can usually be achieved by examining and analyzing a relatively small sample of those images.

By JEAN-LUC STARCK AND JÉRÔME 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 as ridgelets or curvelets have also been proposed for the detection of anisotropic features such as 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 the impact of compressed sensing, the new sampling theory, in astronomy for collecting the data, transferring them to earth or reconstructing an image from incomplete measurements.

KEYWORDS | Astronomical data analysis; compressed sensing; curvelet; restoration; wavelet

I. INTRODUCTION

The wavelet transform (WT) has been extensively used in astronomical data analysis during the last ten years. A quick search with the NASA Astrophysics Data System¹ 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

cosmic microwave background (CMB) 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 kinds of noise—Poisson noise, combination of Gaussian and Poisson noise components, nonstationary noise, and so on—has been a key motivation for the use of wavelets in astrophysics.

If wavelets represent isotropic features well, they are far from optimal for analyzing anisotropic objects. This has motivated other constructions such as the curvelet transform [9]. More generally, the best data decomposition is the one that 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 sets containing edges (planetary images, cosmic strings, etc.), curvelets should be preferred to wavelets.

In this paper, we review a range of astronomical data analysis methods based on sparse representations. We first introduce the isotropic undecimated wavelet transform (IUWT), which is the most popular WT algorithm in astronomy. We show how the signal of interest can be detected in wavelet space using noise modeling, allowing us to build the so-called *multiresolution support*. Then we present in Section III how this multiresolution support can be 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

Manuscript received March 13, 2009; revised May 29, 2009; accepted June 1, 2009. Date of publication October 30, 2009; date of current version May 19, 2010. This work was supported in part by the French National Agency for Research under Grant ANR-08-EMER-009-01.

J.-L. Starck is with the Laboratoire AIM (UMR 7158), CEA/DSM-CNRS-Université Paris Diderot, Centre de Saclay, F-91191 Gif-Sur-Yvette cedex, France (e-mail: jstarck@cea.fr).

J. Bobin is with the Department of Applied and Computational Mathematics, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: bobin@acm.caltech.edu).

Digital Object Identifier: 10.1109/JPROC.2009.2025663

¹adswww.harvard.edu.

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 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 [28] and [30] for more information). Thus, the algorithm outputs $J + 1$ subband arrays of size $n \times n$. (The present indexing is such that $j = 1$ corresponds to the finest scale or high frequencies.)

Hence, we have a *multiscale pixel representation*, i.e., each pixel of the input image is associated with a set of pixels of the multiscale transform. This wavelet transform is very well adapted to the detection of isotropic features. 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 ($h_{2D}, g_{2D} = \delta - h_{2D}$, $\tilde{h}_{2D} = \delta$, $\tilde{g}_{2D} = \delta$) where h_{2D} is the tensor product of two one-dimensional (1-D) filters h_{1D} and δ is the Dirac function. The passage from one resolution to the next one is obtained using the “à trous” algorithm [30]

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

where h_{1D} is typically a symmetric low-pass filter such as the B_3 spline filter $h_{1D} = \{(1/16), (1/4), (3/8), (1/4), (1/16)\}$.

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.

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 of the most well-known methods for contrast

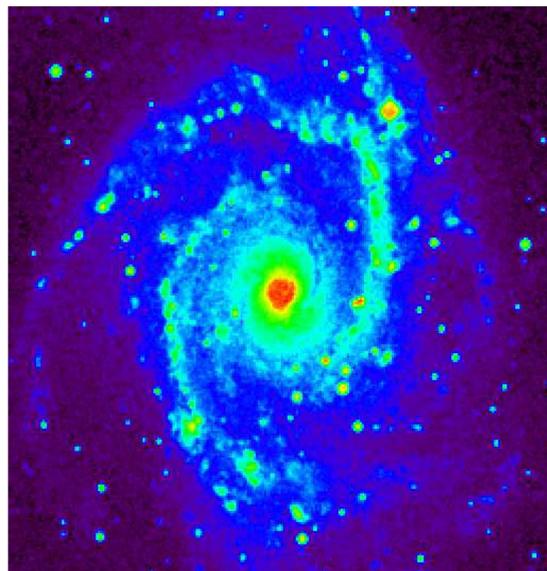


Fig. 1. Galaxy NGC 2997.

enhancement. Images with a high dynamic range are also difficult to analyze. For example, astronomers generally visualize their images using a logarithmic lookup-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 $\text{sgn}(w_j[k, l]) \log(|w_j[k, l]|)$, leading to the alternative image

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

where ϵ is a small number (for example, $\epsilon = 10^{-3}$). Fig. 3 shows a Hale–Bopp Comet image (logarithmic representation) (top), its histogram equalization (bottom left), and its wavelet-log representation (bottom right). 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 follows in many cases a Gaussian distribution, a Poisson distribution, or a combination of both. It is important to detect the wavelet coefficients that are “significant,” i.e., the wavelet coefficients that have an absolute value too large to be due to noise. We defined the multiresolution support M of an image Y by

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

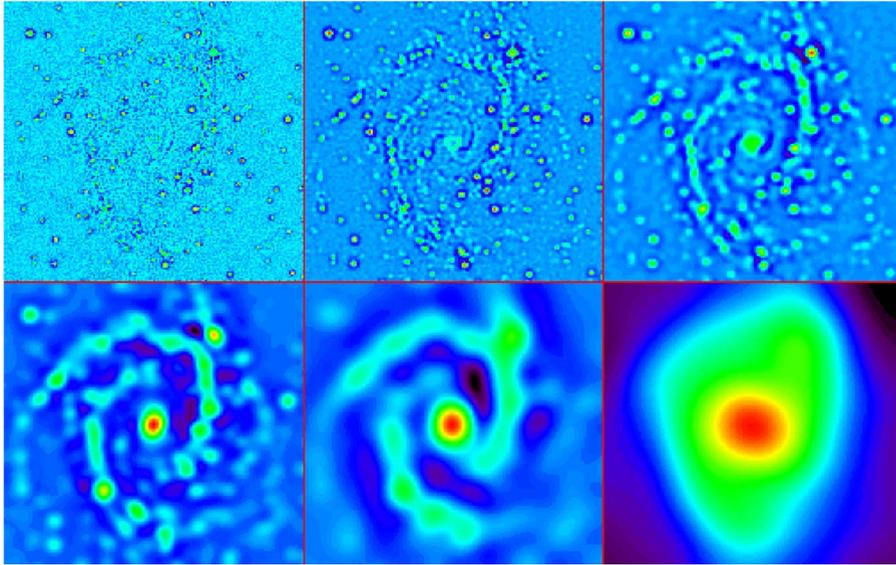


Fig. 2. Wavelet transform of NGC 2997 by the IUWT. The coaddition of these six images reproduces exactly the original image.

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 three and five. The value of three 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

$$\begin{aligned} \text{if } |w_j[k, l]| \geq t_j \text{ then } w_j[k, l] \text{ is significant} \\ \text{if } |w_j[k, l]| < t_j \text{ then } w_j[k, l] \text{ is not significant.} \end{aligned} \quad (4)$$

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] $\mathcal{A}(Y) = 2\sqrt{I + (3/8)}$ 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 readout noise.

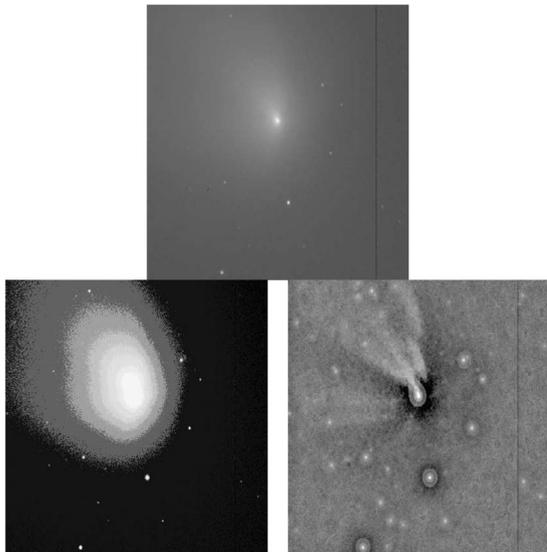


Fig. 3. (Top) Hale-Bopp Comet image. (Bottom left) Histogram equalization results. (Bottom right) Wavelet-log representations.

- *Poisson noise with few events using the MS-VST*: For images with very few photons, one solution consists in using the multiscale 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 one. In this framework, wavelet coefficients are now calculated by

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

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 depend on the filter h_{1D} and the scale level j . They can all be precomputed once for any given h_{1D} [32]. The multiresolution support is computed from the MSVST coefficients, considering a Gaussian noise with a standard deviation equal to one. This stabilization procedure is also invertible, as we have

$$c_0 = \mathcal{A}_0^{-1} \left[\mathcal{A}_j(a_j) + \sum_{j=1}^J w_j \right]. \quad (7)$$

For other kinds of noise (correlated, nonstationary, etc.), other solutions have been proposed to derive the multiresolution support [29]. In the 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 zero all wavelet coefficients of Y that have an absolute value lower than a threshold t_j

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

More generally, for a given sparse representation (wavelet, curvelet, etc.) with its associated 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 $\min_X \|M(\mathcal{T}_w Y - \mathcal{T}_w X)\|_2^2$, where M is the multiresolution support of Y . A solution can be obtained using the Landweber iterative scheme [22], [30]

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

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

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

where P_+ is the projection on the cone of nonnegative images.

This algorithm allows us to constrain 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 $\min_X \|M \mathcal{T}_w (Y - HX)\|_2^2$ under a positivity constraint, leading to the Landweber iterative scheme [22], [30]

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

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 that 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(b). The simulated data are shown in Fig. 4(a) and the wavelet

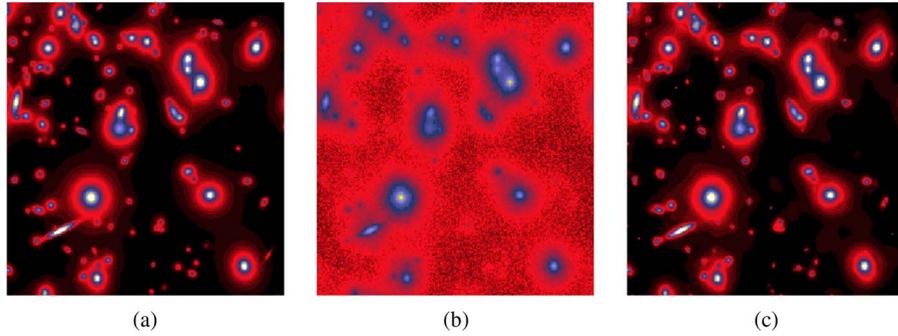


Fig. 4. Simulated Hubble Space Telescope image of a distant cluster of galaxies. (a) Original, unaberrated, and noise-free. (b) Input, aberrated, noise added. (c) Wavelet restoration wavelet.

deconvolution solution Fig. 4(c). 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 to image area we consider as problematic due to calibration or observational problems. These masked areas lead to many difficulties for postprocessing, especially to estimate statistical information such as 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 .

Denoting $\|z\|_0$ the l_0 pseudonorm, i.e., the number of nonzero entries in z , and $\|z\|$ the classical l_2 norm (i.e., $\|z\|^2 = \sum_k (z_k)^2$), we thus want to minimize

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

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 pseudonorm 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]

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

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

- 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 zero otherwise) or a soft thresholding (i.e., $\rho(\alpha_i, \lambda) = \text{sign}(\alpha_i) \max(0, |\alpha_i| - \lambda)$). 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 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] and [16]. For many dictionaries such as wavelets or Fourier, fast operators exist to decompose the signal so that the iteration of (14) is very fast. It requires only performing 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 pattern (see Fig. 5). Fig. 5(a) shows the simulated mass map and (b) shows the masked map. The result of the inpainting method is shown in (c). We note that the gaps are undistinguishable by the eye. More interestingly, 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 two-dimensional (2-D) curvelet transform [9] was developed in an attempt to overcome some limitations inherent in former multiscale methods, e.g., the 2-D 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 $2^{-j/2}$

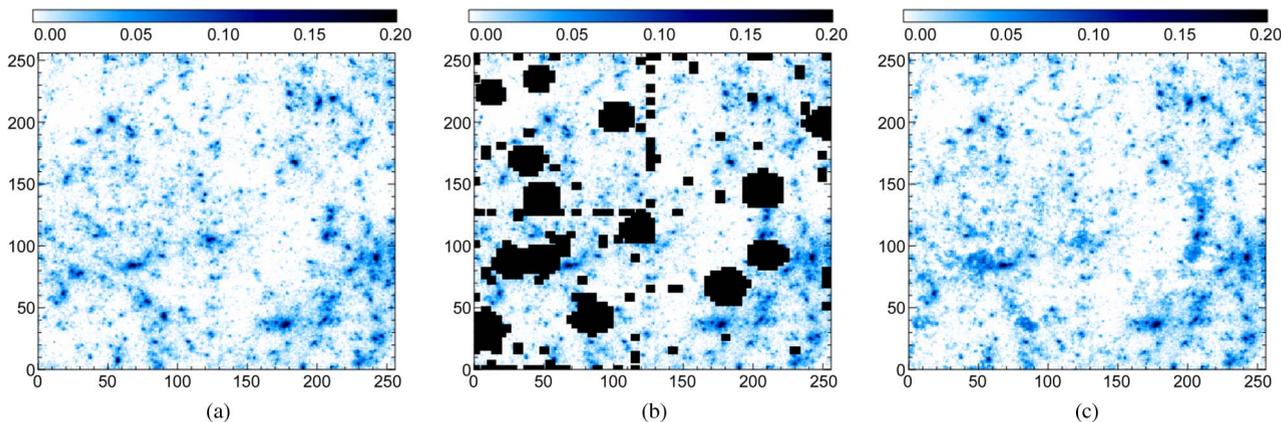


Fig. 5. (a) Simulated weak lensing mass map, (b) simulated mass map with a standard mask pattern, and (c) inpainted mass map. The region shown is $1^\circ \times 1^\circ$.

and its width is 2^{-j} . This is motivated by the parabolic scaling property of smooth curves. Other properties of the curvelet transform as well as decisive optimality results in approximation theory are reported in [8]. Notably, curvelets provide optimally sparse representations of manifolds that are smooth away from edge singularities along smooth curves. Several digital curvelet transforms [7], [23] have been proposed that attempt to preserve the essential properties of the continuous curvelet transform, and several papers report on their successful application in astrophysical experiments [21], [24], [26].

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

A. Application to the Detection of Cosmic Strings

Some applications require the use of sophisticated statistical tools in order to detect a very faint signal, embedded in noise. An interesting case is the detection of non-Gaussian signatures in 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 multifield 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 intracluster 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] and [21], it was shown that the wavelet transform

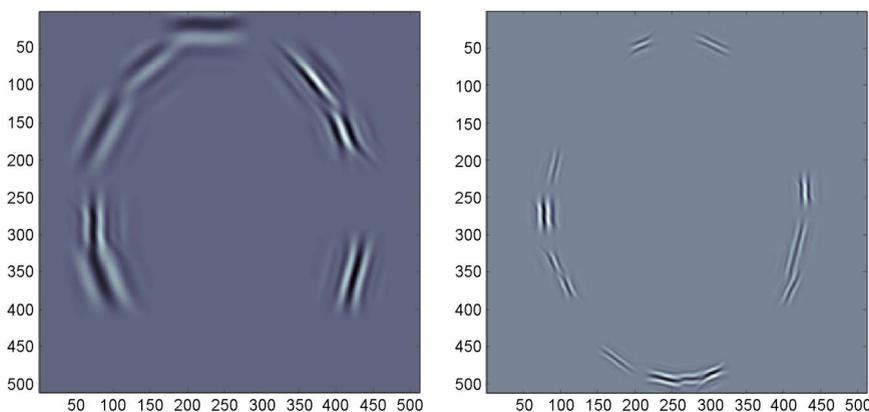


Fig. 6. A few first-generation curvelets. Backprojections of a few curvelet coefficients at different positions and scales.

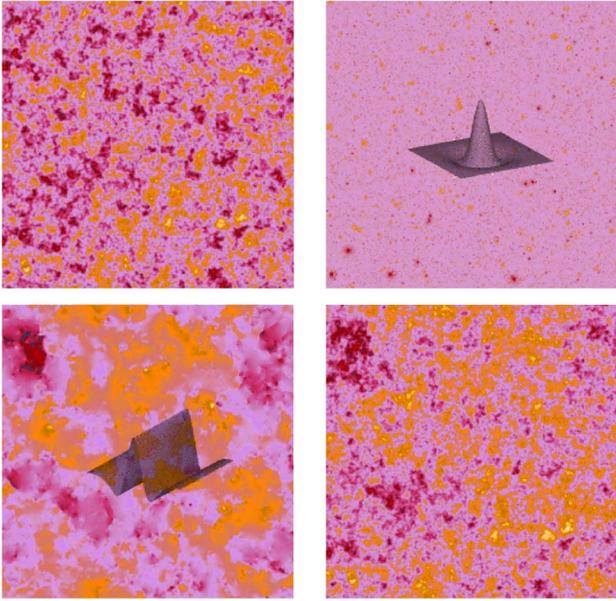


Fig. 7. (Top) (left) Primary CMB anisotropies and (right) kinetic Sunyaev-Zel'dovich fluctuations. (Bottom) (left) Cosmic string simulated map and (right) simulated observation containing the previous three components. The wavelet function is overlotted on the SZ map, and the curvelet function is overlotted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

was a very powerful tool to detect the non-Gaussian signatures. Indeed, the excess kurtosis (fourth moment) of the wavelet coefficients outperformed all the other methods (when the signal is characterized by a nonzero fourth moment).

Lastly, 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 that 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 SZ 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 [17], [21].

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 the top left, top

right, and bottom left of Fig. 7. The “simulated observed map,” containing the three previous components, is displayed in the bottom right of Fig. 7. The primary CMB anisotropies dominate all the signals except at very high multipoles (very small angular scales). The wavelet function is overlotted on the kinetic Sunyaev-Zel'dovich map, and the curvelet function is overlotted on a 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 CS relies on two fundamental properties.

- 1) *Compressibility of the data:* The signal X is said to be *compressible* if there 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 CS 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 the number of measurements m smaller than the number of samples n in X : $m < n$, and A is a random matrix) “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 onboard satellite implementation.

A standard approach in CS attempts to reconstruct X by solving

$$\min_{\alpha} \|\alpha\|_{\ell_1} \text{ such that } \|Y - \Theta_\Lambda \Phi \alpha\|_{\ell_2} < \epsilon \quad (15)$$

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)² is faced with a strenuous compression dilemma: it needs a compression rate equal to $\rho = 1/N$ with $N = 6$. A first approach has been proposed that 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 the redundancy of raster scan data: two consecutive images are almost the same images up to a small shift d . Then, jointly compressing/decompressing consecutive images of the same raster scan has been put forward to alleviate the Herschel/PACS compression dilemma. The problem then consists in recovering a single image X from N compressed and shifted noisy versions of X

$$\forall i \in \{1, \dots, N\}; \quad X_i = \mathcal{S}_{d_i}(X) + \eta_i \quad (16)$$

where \mathcal{S}_{d_i} is an operator that shifts the original image X with a shift d_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:

$$\forall i \in \{1, \dots, N\}; \quad Y_i = \Theta_{\Lambda_i} X_i \quad (17)$$

where the sets $\{\Lambda_i\}$ are such that the union of all the measurement matrices $[\Theta_{\Lambda_1}, \dots, \Theta_{\Lambda_N}]$ spans \mathbb{R}^n . In practice, the subsets Λ_i are disjoint and have a cardinality $m = \lfloor n/N \rfloor$, where m is the coefficients we transfer, n is the number of pixels of each observed image, and N is number of images (here $N = 6$). When there is no shift between consecutive images, these conditions guarantee that the signal X can be reconstructed uniquely from $\{Y_i\}_{i=1, \dots, N}$, up to noise. The decoding step amounts to seeking the signal x as follows:

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

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

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

²The Photodetector Array Camera and Spectrometer (PACS) is one of the three instruments onboard ESA's Herschel Space Observatory. Herschel is a space telescope observing in the far-infrared and submillimeter wavelength region. It was launched on May 14, 2009.

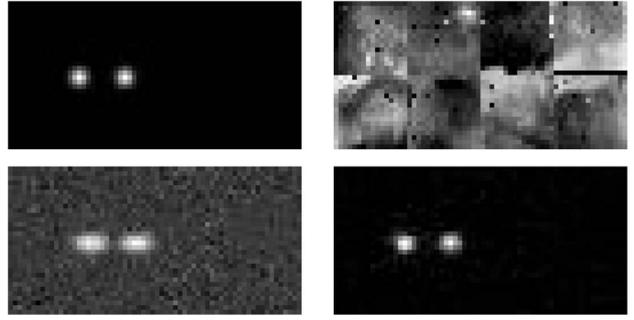


Fig. 8. (Top left) Original image. (Top right) Example of noisy map. (Bottom left) Mean of the six noisy images (see text for more details). (Bottom right) Reconstruction from noiselet-based CS projections. The iterative algorithm has been used with 100 iterations.

where the nonlinear operator $\Delta_{\Phi, \lambda}(Z)$ is defined in (14) and the step-size $\mu_{\Theta} < 2 / \sum_i \|\Theta_{\Lambda_i}^T \Theta_{\Lambda_i}\|_2$. 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 [5, (15)].

a) *Illustration:* We compare two approaches to solve the Herschel/PACS compression problem: i) transmitting the average of six consecutive images (MO6) and ii) compressing six consecutive images of a raster scan and decompressing using CS. 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 six consecutive images, the flat-field component is almost fixed. In this context, the data $\{x_i\}_{i=0, \dots, 1}$ can then be modeled as follows:

$$X_i = \mathcal{S}_{d_i}(X) + \eta_i + c_f. \quad (20)$$

If c_f is known, $\mathcal{S}_{d_i}(X^{(n)})$ is replaced by $\mathcal{S}_{d_i}(X^{(n)}) + c_f$ in (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 Φ is an undecimated wavelet tight frame and the measurement matrices are submatrices of the noiselet basis [11].

The top-left picture of Fig. 8 features the original signal X . In the top-right panel of Fig. 8, we can see a simulated observed image of X . The flat-field component overwhelms the useful part of the data so that X_i has at best a level that is 30 times lower than the flat-field component. The MO6 solution (respectively, the CS-based solution) is shown on the left (respectively, right) and at the bottom of Fig. 8. We showed in [5] that compressed sensing provides a resolution enhancement that can reach 30% of the full width at half-maximum of the instrument's

point spread function 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 performance 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 emphasizing so rigorously the

importance of sparsity, compressed sensing also has 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 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 that 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 explains why wavelets and curvelets are so successful for astronomical image denoising, deconvolution, and inpainting. ■

Acknowledgment

The authors wish to thank J. Fadili for useful comments. 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," *Astron. Astrophys.*, vol. 347, pp. 409–418, Jul. 1999.
- [2] N. Aghanim, M. Kunz, P. G. Castro, and O. Forni, "Non-Gaussianity: Comparing wavelet and Fourier based methods," *Astron. Astrophys.*, vol. 406, pp. 797–816, Aug. 2003.
- [3] F. J. Anscombe, "The transformation of Poisson, binomial and negative-binomial data," *Biometrika*, vol. 15, pp. 246–254, 1948.
- [4] F. Bernardau and J. Uzan, "Non-Gaussianity in multifield inflation," *Phys. Rev. D*, vol. 66, p. 103 506, Nov. 2002.
- [5] J. Bobin, J.-L. Starck, and R. Ottensamer, "Compressed sensing in astronomy," in *IEEE J. Sel. Topics Signal Process.—Special Issue on Signal Processing for Astronomical and Space Research Application*, vol. 2, no. 5, pp. 718–726, 2008.
- [6] F. R. Bouchet, D. P. Bennett, and A. Stebbins, "Patterns of the cosmic microwave background from evolving string networks," *Nature*, vol. 335, p. 410, 1988.
- [7] E. Candès, L. Demanet, D. Donoho, and L. Ying, "Fast discrete curvelet transforms," *SIAM Multiscale Model. Simul.*, vol. 5, no. 3, pp. 861–899, 2006.
- [8] E. J. Candès and D. L. Donoho, "Curvelets—A surprisingly effective nonadaptive representation for objects with edges," in *Curve and Surface Fitting: Saint-Malo 1999*, A. Cohen, C. Rabut, and L. L. Schumaker, Eds. Nashville, TN: Vanderbilt Univ. Press, 1999.
- [9] E. J. Candès and D. Donoho, "Ridgelets: The key to high dimensional intermittency?" *Phil. Trans. Royal Soc. Lond. A*, vol. 357, pp. 2495–2509, 1999.
- [10] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [11] R. Coifman, F. Geshwind, and Y. Meyer, "Noiselets," *Appl. Comput. Harmon. Anal.*, vol. 10, no. 1, pp. 27–44, 2001.
- [12] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM J. Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, 2005.
- [13] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [14] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, pp. 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. Appl. Comput. Harmon. Anal.*, vol. 19, no. 3, pp. 340–358, 2005.
- [16] M. J. Fadili, J.-L. Starck, and F. Murtagh, "Inpainting and zooming using sparse representations," *Comput. J.*, vol. 52, no. 1, pp. 64–79, 2006.
- [17] J. Jin, J.-L. Starck, D. L. Donoho, N. Aghanim, and O. Forni, "Cosmological non-Gaussian signatures detection: Comparison of statistical tests," *EURASIP J.*, vol. 15, pp. 2470–2485, 2005.
- [18] F. Murtagh, J.-L. Starck, and A. Bijaoui, "Image restoration with noise suppression using a multiresolution support," *Astron. Astrophys. (Suppl. Series)*, vol. 112, pp. 179–189, 1995.
- [19] S. Pires, J.-L. 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," *Mon. Not. R. Astron. Soc.*, vol. 395, no. 3, pp. 1265–1279.
- [20] A. Riazuelo, J.-P. Uzan, R. Lehoucq, and J. Weeks, "Simulating cosmic microwave background maps in multi-connected spaces," *Phys. Rev. D*, vol. 69, 2002.
- [21] J.-L. Starck, N. Aghanim, and O. Forni, "Detecting cosmological non-Gaussian signatures by multi-scale methods," *Astron. Astrophys.*, vol. 416, pp. 9–17, 2004.
- [22] J.-L. Starck, A. Bijaoui, and F. Murtagh, "Multiresolution support applied to image filtering and deconvolution," *Graph. Models Image Process.*, vol. 57, pp. 420–431, 1995.
- [23] J.-L. Starck, E. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 131–141, 2002.
- [24] J.-L. Starck, E. Candès, and D. L. Donoho, "Astronomical image representation by the curvelet transform," *Astron. Astrophys.*, vol. 398, pp. 785–800, 2003.
- [25] J.-L. Starck, M. J. Fadili, and F. Murtagh, "The undecimated wavelet decomposition and its reconstruction," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 297–309, 2007.
- [26] J.-L. Starck, Y. Moudden, P. Abrial, and M. Nguyen, "Wavelets, ridgelets and curvelets on the sphere," *Astron. Astrophys.*, vol. 446, pp. 1191–1204, 2006.
- [27] J.-L. Starck and F. Murtagh, "Automatic noise estimation from the multiresolution support," *Pub. Astron. Soc. Pacific*, vol. 110, pp. 193–199, 1998.
- [28] J.-L. Starck and F. Murtagh, *Astronomical Image and Data Analysis*. Berlin, Germany: Springer-Verlag, 2002.
- [29] J.-L. Starck and F. Murtagh, *Astronomical Image and Data Analysis*. New York: Springer, 2006.
- [30] J.-L. Starck, F. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis: The Multiscale Approach*. Cambridge, U.K.: Cambridge Univ. 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," *Annu. Rev. Astron. Astrophys.*, vol. 18, pp. 537–560, 1980.
- [32] B. Zhang, M. J. Fadili, and J.-L. Starck, "Wavelets, ridgelets and curvelets for Poisson noise removal," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1093–1108, 2008.

ABOUT THE AUTHORS

Jean-Luc Starck received the Ph.D. degree from University Nice-Sophia Antipolis, France, and the Habilitation degree from University Paris XI, France.

He was a Visitor with the European Southern Observatory (ESO) in 1993; the University of California, Los Angeles, in 2004; and the Statistics Department, Stanford University, in 2000 and 2005. He has been a Researcher with CEA, Université Paris Diderot, France, since 1994. His research interests include image processing, statistical methods in astrophysics, and cosmology. He is an expert in multiscale methods (wavelets, curvelets, etc.), Leader of the project Multiresolution at CEA, and a core Team Member of the PLANCK ESA project. He has published more than 100 papers in different areas in scientific journals. He is the author of *Image Processing and Data Analysis: The Multiscale Approach* (Cambridge, U.K.: Cambridge University Press, 1998) and *Astronomical Image and Data Analysis* (2nd ed.) (New York: Springer, 2006).



Jérôme Bobin graduated from the Ecole Normale Supérieure (ENS) de Cachan, France, in 2005. He received the M.Sc. degree in signal and image processing from ENS Cachan and Université Paris XI, Orsay, France. He received the Agrégation de Physique degree from in 2004. He received the Ph.D. degree from CEA-Université Paris Diderot, France, and he is now at Caltech in the Department of Applied and Computational Mathematics.

His research interests include statistics, information theory, multiscale methods and sparse representations in signal and image processing.

