Cosmological Non-Gaussian Signature Detection: Comparing Performance of Different Statistical Tests

J. Jin

Department of Statistics, Purdue University, 150 N. University Street, West Lafayette, IN 47907, USA
Email: jinj@stat.purdue.edu

J.-L. Starck

DAPNIA/SEDI-SAP, Service d'Astrophysique, CEA-Saclay, 91191 Gif-sur-Yvette Cedex, France
Email: jstarck@cea.fr

D. L. Donoho

Department of Statistics, Stanford University, Sequoia Hall, Stanford, CA 94305, USA
Email: donoho@stat.stanford.edu

N. Aghanim

IAS-CNRS, Université Paris Sud, Bâtiment 121, 91405 Orsay Cedex, France
Email: aghanim@ias.fr

Division of Theoretical Astronomy, National Astronomical Observatory of Japan, Osawa 2-21-1, Mitaka, Tokyo 181-8588, Japan

O. Forni

IAS-CNRS, Université Paris Sud, Bâtiment 121, 91405 Orsay Cedex, France
Email: olivier.forni@ias.fr

Currently, it appears that the best method for non-Gaussianity detection in the cosmic microwave background (CMB) consists in calculating the kurtosis of the wavelet coefficients. We know that wavelet-kurtosis outperforms other methods such as the bispectrum, the genus, ridgelet-kurtosis, and curvelet-kurtosis on an empirical basis, but relatively few studies have compared other transform-based statistics, such as extreme values, or more recent tools such as higher criticism (HC), or proposed "best possible" choices for such statistics. In this paper, we consider two models for transform-domain coefficients: (a) a power-law model, which seems suited to the wavelet coefficients of simulated cosmic strings, and (b) a sparse mixture model, which seems suitable for the curvelet coefficients of filamentary structure. For model (a), if power-law behavior holds with finite 8th moment, excess kurtosis is an asymptotically optimal detector, but if the 8th moment is not finite, a test based on extreme values is asymptotically optimal. For model (b), if the transform coefficients are very sparse, a recent test, higher criticism, is an optimal detector, but if they are dense, kurtosis is an optimal detector. Empirical wavelet coefficients of simulated cosmic strings have power-law character, infinite 8th moment, while curvelet coefficients of the simulated cosmic strings are not very sparse. In all cases, excess kurtosis seems to be an effective test in moderate-resolution imagery.

Keywords and phrases: cosmology, cosmological microwave background, non-Gaussianity detection, multiscale method, wavelet, curvelet.

1. INTRODUCTION

The cosmic microwave background (CMB), discovered in 1965 by Penzias and Wilson [1], is a relic of radiation emitted some 13 billion years ago, when the universe was about 370 000 years old. This radiation exhibits characteristic of an almost perfect blackbody at a temperature of 2.726 kelvins as measured by the FIRAS experiment on board COBE satellite [2]. The DMR experiment, again on board COBE, detected and measured angular small fluctuations of this temperature, at the level of a few tens of microkelvins, and at angular scale of about 10 degrees [3]. These so-called temperature anisotropies were predicted as the imprints of the initial density perturbations which gave rise to present large
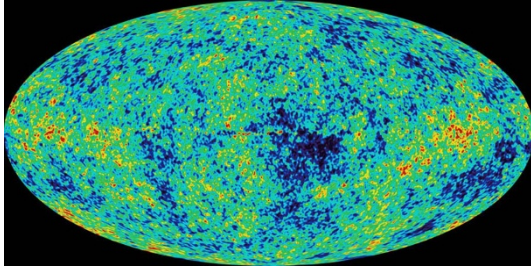
FIGURE 1: Courtesy of the WMAP team (reference to the website). All sky map of the CMB anisotropies measured by the WMAP satellite.

scale structures as galaxies and clusters of galaxies. This relation between the present-day universe and its initial conditions has made the CMB radiation one of the preferred tools of cosmologists to understand the history of the universe, the formation and evolution of the cosmic structures, and physical processes responsible for them and for their clustering.

As a consequence, the last several years have been a particularly exciting period for observational cosmology focusing on the CMB. With CMB balloon-borne and ground-based experiments such as TOCO [4], BOOMERanG [5], MAXIMA [6], DASI [7], and Archeops [8], a firm detection of the so-called "first peak" in the CMB anisotropy angular power spectrum at the degree scale was obtained. This detection has been very recently confirmed by the WMAP satellite [9, 10] (see Figure 1), which detected also the second and third peaks. WMAP satellite mapped the CMB temperature fluctuations with a resolution better than 15 arcminutes and a very good accuracy marking the starting point of a new era of precision cosmology that enables us to use the CMB anisotropy measurements to constrain the cosmological parameters and the underlying theoretical models.

In the framework of adiabatic cold dark matter models, the position, amplitude, and width of the first peak indeed provide strong evidence for the inflationary predictions of a flat universe and a scale-invariant primordial spectrum for the density perturbations. Furthermore, the presence of second and third peaks confirms the theoretical prediction of acoustic oscillations in the primeval plasma and sheds new light on various cosmological and inflationary parameters, in particular, the baryonic content of the universe. The accurate measurements of both the temperature anisotropies and polarised emission of the CMB will enable us in the very near future to break some of the degeneracies that are still affecting parameter estimation. It will also allow us to probe more directly the inflationary paradigm favored by the present observations.

Testing the inflationary paradigm can also be achieved through detailed study of the statistical nature of the CMB anisotropy distribution. In the simplest inflation models, the distribution of CMB temperature fluctuations should be Gaussian, and this Gaussian field is completely determined by its power spectrum. However, many models such as multifield inflation (e.g., [11] and the references therein), su-

per strings, or topological defects predict non-Gaussian contributions to the initial fluctuations [12, 13, 14]. The statistical properties of the CMB should discriminate models of the early universe. Nevertheless, secondary effects like the inverse Compton scattering, the Doppler effect, lensing, and others add their own contributions to the total non-Gaussianity.

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. Many approaches have been investigated: Minkowski functionals and the morphological statistics [15, 16], the bispectrum (3-point estimator in the Fourier domain) [17, 18, 19], the trispectrum (4-point estimator in the Fourier domain) [20], wavelet transforms [21, 22, 23, 24, 25, 26, 27], and the curvelet transform [27]. Different wavelet methods have been studied, such as the isotropic à trous algorithm [28] and the biorthogonal wavelet transform [29]. (The biorthogonal wavelet transform was found to be the most sensitive to non-Gaussianity [27].) In [27, 30], 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 characterised by a nonzero 4th moment).

Nevertheless, a major issue of the non-Gaussian studies in CMB remains our ability to disentangle all the sources of non-Gaussianity from one another. Recent progress has been made on the discrimination between different possible origins of non-Gaussianity. Namely, it was possible to separate the non-Gaussian signatures associated with topological defects (cosmic strings (CS)) from those due to Doppler effect of moving clusters of galaxies (both dominated by a Gaussian CMB field) by combining the excess kurtosis derived from both the wavelet and the curvelet transforms [27].

This success argues for us to construct a "toolkit" of well-understood and sensitive methods for probing different aspects of the non-Gaussian signatures.

In that spirit, the goal of the present study is to consider the advantages and limitations of detectors which apply kurtosis to transform coefficients of image data. We will study plausible models for transform coefficients of image data and compare the performance of tests based on kurtosis of transform coefficients to other types of statistical diagnostics.

At the center of our analysis are the following two facts:

(A) the wavelet/curvelet coefficients of CMB are Gaussian (we implicitly assume the most simple inflationary scenario);

(B) the wavelet/curvelet coefficients of topological defect and Doppler effect simulations are non-Gaussian.

We develop tests for non-Gaussianity for two models of statistical behavior of transform coefficients. The first, better suited for wavelet analysis, models transform coefficients of cosmic strings as following a power law. The second, theoretically better suited for curvelet coefficients, assumes that the

salient features of interest are actually filamentary (it can be residual strips due to a nonperfect calibration), which gives the curvelet coefficients a sparse structure.

We review some basic ideas from detection theory, such as likelihood ratio detectors, and explain why we prefer nonparametric detectors, valid across a broad range of assumptions.

In the power-law setting, we consider two kinds of nonparametric detectors. The first, based on kurtosis, is asymptotically optimal in the class of weakly dependent symmetric non-Gaussian contamination with finite 8th moments. The second, the Max, is shown to be asymptotically optimal in the class of weakly dependent symmetric non-Gaussian contamination with infinite 8th moment. While the evidence seems to be that wavelet coefficients of CS have about 6 existing moments, indicating a decisive advantage for extreme-value statistics, the performance of kurtosis-based tests and Max-based tests on moderate sample sizes (e.g., 64 K transform coefficients) does not follow the asymptotic theory; excess kurtosis works better at these sample sizes.

In the sparse-coefficients setting, we consider kurtosis, the Max, and a recent statistic called *higher criticism* (HC) [31]. Theoretical analysis suggests that curvelet coefficients of filamentary features should be sparse, with about $n^{1/4}$ substantial nonzero coefficients out of n coefficients in a subband; this level of sparsity would argue in favor of Max/HC. However, empirically, the curvelet coefficients of actual CS simulations are not very sparse. It turns out that kurtosis outperforms Max/HC in simulation.

Summarizing, the work reported here seems to show that for all transforms considered, the excess kurtosis outperforms alternative methods despite their strong theoretical motivation. A reanalysis of the theory supporting those methods shows that the case for kurtosis can also be justified theoretically based on observed statistical properties of the transform coefficients not used in the original theoretic analysis.

2. DETECTING FAINT NON-GAUSSIAN SIGNALS SUPERPOSED ON A GAUSSIAN SIGNAL

The superposition of a non-Gaussian signal with a Gaussian signal can be modeled as $Y = N + G$, where Y is the observed image, N is the non-Gaussian component, and G is the Gaussian component. We are interested in using transform coefficients to test whether $N \equiv 0$ or not.

2.1. Hypothesis testing and likelihood ratio test

Transform coefficients of various kinds (Fourier, wavelet, etc.) have been used for detecting non-Gaussian behavior in numerous studies. Let $X_1, X_2, \ldots, X_n$ be the transform coefficients of Y; we model these as

$$X_i = \sqrt{1-\lambda} \cdot z_i + \sqrt{\lambda} \cdot w_i, \quad 0 < \lambda < 1, \tag{1}$$

where $\lambda > 0$ is a parameter, $z_i \stackrel{iid}{\sim} N(0,1)$ are the transform coefficients of the Gaussian component G, $w_i \stackrel{iid}{\sim} W$ are

the transform coefficients of the non-Gaussian component N, and W is some unknown symmetrical distribution. Here without loss of generality, we assume the standard deviation for both z_i and w_i is 1.

Phrased in statistical terms, the problem of detecting the existence of a non-Gaussian component is equivalent to discriminating between the hypotheses

$$H_0 : X_i = z_i, \tag{2}$$

$$H_1 : X_i = \sqrt{1-\lambda} \cdot z_i + \sqrt{\lambda} \cdot w_i, \quad 0 < \lambda < 1, \tag{3}$$

and $N \equiv 0$ is equivalent to $\lambda \equiv 0$. We call H_0 the *null hypothesis* and H_1 the *alternative hypothesis*.

When both W and λ are known, then the optimal test for problem (2)-(3) is simply the Neyman-Pearson likelihood ratio test (LRT) [32, page 74]. The size of $\lambda = \lambda_n$ for which reliable discrimination between H_0 and H_1 is possible can be derived using asymptotics. If we assume that the tail probability of W decays algebraically,

$$\lim_{x \to \infty} x^\alpha P\{|W| > x\} = C_\alpha, \quad C_\alpha \text{ is a constant} \tag{4}$$

(we say W has a power-law tail), and we calibrate λ to decay with n, so that increasing amounts of data are offset by increasingly hard challenges:

$$\lambda = \lambda_n = n^{-r}, \tag{5}$$

then there is a *threshold effect* for the detection problem (2)-(3). In fact, define

$$\rho_1^*(\alpha) = \begin{cases} \dfrac{2}{\alpha}, & \alpha \leq 8, \\[2mm] \dfrac{1}{4}, & \alpha > 8, \end{cases} \tag{6}$$

then, as $n \to \infty$, LRT is able to reliably detect for large n when $r < \rho_1^*(\alpha)$, and is unable to detect when $r > \rho_1^*(\alpha)$; this is proved in [33]. Since LRT is optimal, it is not possible for any statistic to reliably detect when $r > \rho_1^*(\alpha)$. We call the curve $r = \rho_1^*(\alpha)$ in the α-r plane the *detection boundary*; see Figure 2.

In fact, when $r < 1/4$, asymptotically LRT is able to reliably detect whenever W has a finite 8th moment, even without the assumption that W has a power-law tail. Of course, the case that W has an infinite 8th moment is more complicated, but if W has a power-law tail, then LRT is also able to reliably detect if $r < 2/\alpha$.

Despite its optimality, LRT is not a practical procedure. To apply LRT, one needs to specify the value of λ and the distribution of W, which seems unlikely to be available. We need nonparametric detectors, which can be implemented without any knowledge of λ or W, and depend on X_i's only. In the section below, we are going to introduce two nonparametric detectors: excess kurtosis and Max; later in Section 4.3, we will introduce a third nonparametric detector: higher criticism (HC).
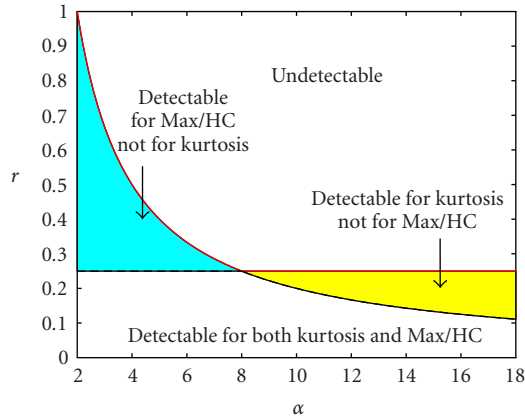
FIGURE 2: Detectable regions in the α-r plane. With (α, r) in the white region on the top or in the undetectable region, all methods completely fail for detection. With (α, r) in the white region on the bottom, both excess kurtosis and Max/HC are able to detect reliably. In the shaded region to the left, Max/HC is able to detect reliably, but excess kurtosis completely fails, and in the shaded region to the right, excess kurtosis is able to detect reliably, but Max/HC completely fails.

2.2. Excess kurtosis and Max

We pause to review the concept of p-value briefly. For a statistic T_n, the p-value is the probability of seeing equally extreme results under the null hypothesis:

$$p = P_{H_0}\{T_n \geq t_n(X_1, X_2, \ldots, X_n)\}; \tag{7}$$

here P_{H_0} refers to probability under H_0, and $t_n(X_1, X_2, \ldots, X_n)$ is the observed value of statistic T_n. Notice that the smaller the p-value, the stronger the evidence against the null hypothesis. A natural decision rule based on p-values rejects the null when $p < \alpha$ for some selected level α, and a convenient choice is $\alpha = 5\%$. When the null hypothesis is indeed true, the p-values for any statistic are distributed as uniform $U(0, 1)$. This implies that the p-values provide a common scale for comparing different statistics.

We now introduce two statistics for comparison.

Excess kurtosis (κ_n)

Excess kurtosis is a widely used statistic, based on the 4th moment. For any (symmetrical) random variable X, the kurtosis is

$$\kappa(X) = \frac{EX^4}{(EX^2)^2} - 3. \tag{8}$$

The kurtosis measures a kind of departure of X from Gaussianity, as $\kappa(z) = 0$.

Empirically, given n realizations of X, the excess kurtosis statistic is defined as

$$\kappa_n(X_1, X_2, \ldots, X_n) = \sqrt{\frac{n}{24}}\left[\frac{(1/n)\sum_i X_i^4}{((1/n)\sum_i X_i^2)^2} - 3\right]. \tag{9}$$

When the null is true, the excess kurtosis statistic is asymptotically normal:

$$\kappa_n(X_1, X_2, \ldots, X_n) \longrightarrow_w N(0, 1), \quad n \longrightarrow \infty, \tag{10}$$

thus for large n, the p-value of the excess kurtosis is approximately

$$\tilde{p} = \bar{\Phi}^{-1}(\kappa_n(X_1, X_2, \ldots, X_n)), \tag{11}$$

where $\bar{\Phi}(\cdot)$ is the survival function (upper-tail probability) of $N(0, 1)$.

It is proved in [33] that the excess kurtosis is asymptotically optimal for the hypothesis testing of (2)-(3) if

$$E[W^8] < \infty. \tag{12}$$

However, when $E[W^8] = \infty$, even though kurtosis is well defined ($E[W^4] < \infty$), there are situations in which LRT is able to reliably detect but excess kurtosis completely fails. In fact, by assuming (4)-(5) with an $\alpha < 8$, if (α, r) falls into the shaded region to the left of Figure 2, then LRT is able to reliably detect, however, excess kurtosis completely fails. This shows that in such cases, excess kurtosis is not optimal; see [33].

Max (M_n)

The largest (absolute) observation is a classical and frequently used nonparametric statistic:

$$M_n = \max(|X_1|, |X_2|, \ldots, |X_n|), \tag{13}$$

under the null hypothesis,

$$M_n \approx \sqrt{2\log n}, \tag{14}$$

and, moreover, by normalizing M_n with constants c_n and d_n, the resulting statistic converges to the Gumbel distribution E_v, whose cdf is $e^{-e^{-x}}$:

$$\frac{M_n - c_n}{d_n} \longrightarrow_w E_v, \tag{15}$$

where approximately

$$d_n = \frac{\sqrt{6}S_n}{\pi}, \qquad c_n = \bar{X} - 0.5772d_n; \tag{16}$$

here $\bar{X}$ and S_n are the sample mean and sample standard deviation of $\{X_i\}_{i=1}^n$, respectively. Thus a good approximation of the p-value for M_n is

$$\tilde{p} = \exp\left(-\exp\left(-\frac{M_n - c_n}{d_n}\right)\right). \tag{17}$$

We have tried the above experiment for $n = 244^2$, and found that taking $c_n = 4.2627$, $d_n = 0.2125$ gives a good approximation.

Assuming (4)-(5) and $\alpha < 8$, or $\lambda = n^{-r}$, and that W has a power-law tail with $\alpha < 8$, it is proved in [33] that Max is optimal for hypothesis testing (2)-(3). Recall if we further assume $1/4 < r < 2/\alpha$, then asymptotically excess kurtosis completely fails; however, Max is able to reliably detect and is competitive to LRT.

On the other hand, recall that excess kurtosis is optimal for the case $\alpha > 8$. In comparison, in this case, Max is not optimal. In fact, if we further assume $2/\alpha < r < 1/4$, then excess kurtosis is able to reliably detect, but Max will completely fail.

In Figure 2, we compared the detectable regions of the excess kurtosis and Max in the α-r plane.

To conclude this section, we mention an alternative way to approximate the p-values for any statistic T_n. This alternative way is important in case that an asymptotic (theoretic) approximation is poor for moderate large n, an example is the statistic HC_n^* we will introduce in Section 4.3; this alternative way is helpful even when the asymptotic approximation is accurate. Now the idea is that, under the null hypothesis, we simulate a large number ($N = 10^4$ or more) of T_n: $T_n^{(1)}, T_n^{(2)}, \ldots, T_n^{(N)}$, we then tabulate them. For the observed value $t_n(X_1, X_2, \ldots, X_n)$, the p-value will then be well approximated by

$$\frac{1}{N} \cdot \#\left\{k : T_n^{(k)} \geq t_n(X_1, X_2, \ldots, X_n)\right\}, \qquad (18)$$

and the larger the N, the better the approximation.

2.3. Heuristic approach

We have exhibited a phase-change phenomenon, where the asymptotically optimal test changes depending on power-law index α. In this section, we develop a heuristic analysis of detectability and phase change.

The detection property of Max follows from comparing the ranges of data. Recall that $X_i = \sqrt{1 - \lambda_n} \cdot z_i + \sqrt{\lambda_n} \cdot w_i$, the range of $\{z_i\}_{i=1}^n$ is roughly $(-\sqrt{2 \log n}, \sqrt{2 \log n})$, and the range of $\{\sqrt{\lambda_n} \cdot w_i\}_{i=1}^n$ is $\sqrt{\lambda_n} \cdot (-n^{1/\alpha}, n^{1/\alpha}) = (-n^{1/\alpha - r/2}, n^{1/\alpha - r/2})$; so, heuristically,

$$M_n \approx \max\left\{\sqrt{2 \log n}, n^{1/\alpha - r/2}\right\}; \qquad (19)$$

for large n, notice that

$$
\begin{aligned}
n^{1/\alpha - r/2} &\gg \sqrt{2 \log n}, &&\text{if } r < \frac{2}{\alpha}, \\
n^{1/\alpha - r/2} &\ll \sqrt{2 \log n}, &&\text{if } r > \frac{2}{\alpha},
\end{aligned}
\qquad (20)
$$

thus if and only if $r < 2/\alpha$, M_n for the alternative will differ significantly from M_n for the null, and so the criterion for detectability by Max is $r < 2/\alpha$.

Now we study detection by excess kurtosis. Heuristically,

$$
\begin{aligned}
\kappa_n &\approx \left(\frac{1}{\sqrt{24}}\right) \cdot \kappa\left(\sqrt{1 - \lambda_n} \cdot z_i + \sqrt{\lambda_n} \cdot w_i\right) \\
&= \left(\frac{1}{\sqrt{24}}\right) \cdot \sqrt{n} \cdot \lambda_n^2 \cdot \kappa(W) = O(n^{1/2 - 2r}),
\end{aligned}
\qquad (21)
$$

thus if and only if $r < 1/4$, κ_n for the alternative will differ significantly from κ_n under the null, and so the criterion for detectability by excess kurtosis is $r < 1/4$.

This analysis shows the reason for the phase change. In Figure 2, when the parameter (α, r) is in the shaded region to the left, for sufficiently large n, $n^{1/\alpha - r/2} \gg \sqrt{2 \log n}$ and the strongest evidence against the null is in the tails of the data set, which M_n is indeed using. However, when (α, r) moves from the shaded region to the left to the shaded region to the right, $n^{1/\alpha - r/2} \ll \sqrt{2 \log n}$, the tails no longer contain any important evidence against the null, instead, the central part of the data set contains the evidence. By symmetry, the 1st and the 3rd moments vanish, and the 2nd moment is 1 by the normalization; so the excess kurtosis is in fact the most promising candidate of detectors based on moments.

The heuristic analysis is the essence for theoretic proof as well as empirical experiment. Later in Section 3.4, we will have more discussions for comparing the excess kurtosis with Max down this vein.

3. WAVELET COEFFICIENTS OF COSMIC STRINGS

3.1. Simulated astrophysical signals

The temperature anisotropies of the CMB contain the contributions of both the primary cosmological signal, directly related to the initial density perturbations, and the secondary anisotropies. The latter are generated after matter-radiation decoupling [34]. They arise from the interaction of the CMB photons with the neutral or ionised matter along their path [35, 36, 37].

In the present study, we assume that the primary CMB anisotropies are dominated by the fluctuations generated in the simple single field inflationary cold-dark-matter model with a nonzero cosmological constant. The CMB anisotropies have therefore a Gaussian distribution. We allow for a contribution to the primary signal from topological defects, namely, cosmic strings (CS), as suggested in [38, 39]. See Figures 3 and 4.

We use for our simulations the cosmological parameters obtained from the WMAP satellite [10] and a normalization parameter $\sigma_8 = 0.9$. Finally, we obtain the so-called "simulated observed map," D, that contains the two previous astrophysical components. It is obtained from $D_\lambda = \sqrt{1 - \lambda}\text{CMB} + \sqrt{\lambda}\text{CS}$, where CMB and CS are, respectively, the CMB and the cosmic string simulated maps. $\lambda = 0.18$ is an upper limit constant derived by [38]. All the simulated maps have 500×500 pixels with a resolution of 1.5 arcminutes per pixel.

3.2. Evidence for $E[W^8] = \infty$

For the wavelet coefficients on the finest scale of the cosmic string map in Figure 3b, by throwing away all the coefficients related to pixels on the edge of the map, we have $n = 244^2$ coefficients; we then normalize these coefficients so that the empirical mean and standard deviation are 0 and 1, respectively; we denote the resulting dataset by $\{w_i\}_{i=1}^n$.
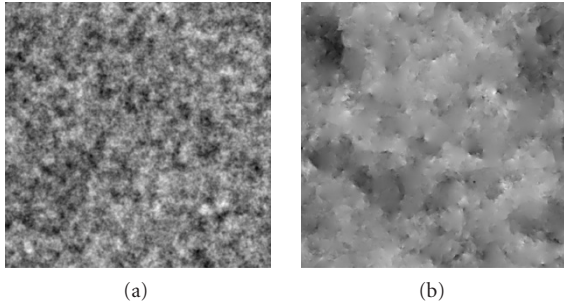
FIGURE 3: (a) Primary cosmic microwave background anisotropies and (b) simulated cosmic string map.

FIGURE 4: Simulated observation containing the CMB and the CS ($\lambda = 0.18$).

Assuming $\{w_i\}_{i=1}^n$ are independent samples from a distribution W, we have seen in Section 2 that whether excess kurtosis is better than Max depends on the finiteness of $E[W^8]$. We now analyze $\{w_i\}_{i=1}^n$ to learn about $E[W^8]$.

Let

$$\hat{m}_8^{(n)} = \frac{1}{n} \sum_{i=1}^n w_i^8 \tag{22}$$

be the empirical 8th moment of W using n samples. In theory, if $E[W^8] < \infty$, then $\hat{m}_8^{(n)} \to E[W^8]$ as $n \to \infty$. So one way to see if $E[W^8]$ is finite is to observe how $\hat{m}_8^{(n)}$ changes with n.

Technically, since we only have $n = 244^2$ samples, we can compare

$$\hat{m}_8^{(n/2^k)}, \quad k = 0, 1, 2, 3, 4; \tag{23}$$

if these values are roughly the same, then there is strong evidence for $E[W^8] < \infty$; otherwise, if they increase with sample size, that is evidence for $E[W^8] = \infty$. Here $m_8^{(n/2^k)}$ is an estimate of $E[W^8]$ using $n/2^k$ subsamples of $\{w_i\}_{i=1}^n$.

For $k = 1, 2, 3, 4$, to obtain $\hat{m}_8^{(n/2^k)}$, we randomly draw subsamples of size $n/2^k$ from $\{w_i\}_{i=1}^n$, and then take the average of the 8th power of this subsequence; we repeat this process 50 000 times, and we let $\hat{m}_8^{(n/2^k)}$ be the median of these 50 000 average values. Of course when $k = 0$, $\hat{m}_8^{(n/2^k)}$ is obtained from all n samples.

The results corresponding to the first wavelet band are summarized in Table 1. From the table, we have seen that $\hat{m}_8^{(n)}$ is significantly larger than $\hat{m}_8^{(n/8)}$ and $\hat{m}_8^{(n/16)}$; this supports that $E[W^8] = \infty$. Similar results were obtained from the other bands. In comparison, in Table 1, we also list the 4th, 5th, 6th, and 7th moments. It seems that the 4th, 5th, and 6th moments are finite, but the 7th and 8th moments are infinite.

3.3. Power-law tail of W

Typical models for heavy-tailed data include exponential tails and power-law tails. We now compare such models to the data on wavelet coefficients for W; the Gaussian model is also included as comparison.
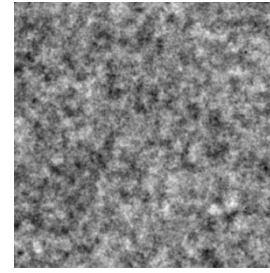
We sort the $|w_i|$'s in descending order, $|w|_{(1)} > |w|_{(2)} > \cdots > |w|_{(n)}$, and take the 50 largest samples $|w|_{(1)} > |w|_{(2)} > \cdots > |w|_{(50)}$. For a power-law tail with index α, we expect that for some constant C_α,

$$\log\left(\frac{i}{n}\right) \approx \log\left(C_\alpha\right) - \alpha \log\left(|w|_{(i)}\right), \quad 1 \le i \le 50, \tag{24}$$

so there is a strong linear relationship between $\log(i/n)$ and $\log(|w|_{(i)})$. Similarly, for the exponential model, we expect a strong linear relationship between $\log(i/n)$ and $|w|_{(i)}$, and for the Gaussian model, we expect a strong linear relationship between $\log(i/n)$ and $|w|_{(i)}^2$.

For each model, to measure whether the "linearity" is sufficient to explain the relationship between $\log(i/n)$ and $\log(|w|_{(i)})$ (or $|w|_{(i)}$, or $|w|_{(i)}^2$), we introduce the following z-score:

$$Z_i = \sqrt{n}\left[\frac{\hat{p}_i - i/n}{i/n(1 - i/n)}\right], \tag{25}$$

where $\hat{p}_i$ is the linear fit using each of the three models. If the resulting z-scores is random and have no specific trend, the model is appropriate; otherwise, the model may need improvement.

The results are summarized in Figure 5. The power-law tail model seems the most appropriate: the relationship between $\log(i/n)$ and $\log(|w|_{(i)})$ looks very close to linear, the z-score looks very small, and the range of z-scores much narrower than the other two. For the exponential model, the linearity is fine at the first glance, however, the z-score is *decreasing* with i, which implies that the tail is heavier than estimated. The Gaussian model fits much worse than exponential. To summarize, there is strong evidence that the tail follows a power law.

Now we estimate the index α for the power-law tail. A widely used method for estimating α is the Hill estimator [40]:

$$\hat{\alpha}_H^{(l)} = \frac{l+1}{\sum_{i=1}^l i \log\left(|w|_{(i)}/|w|_{(i+1)}\right)}, \tag{26}$$

where l is the number of (the largest) $|w|_{(i)}$ to include for estimation. In our situation, $l = 50$ and

$$\hat{\alpha} = \hat{\alpha}_H^{(50)} = 6.134; \tag{27}$$

TABLE 1: Empirical estimate 4th, 5th, 6th, 7th, and 8th moments calculated using a subsamples of size $n/2^k$ of $\{|w_i|\}_{i=1}^n$, with $k = 0, 1, 2, 3, 4$. The table suggests that the 4th, 5th, and 6th moments are finite, but the 7th and 8th moments are infinite.

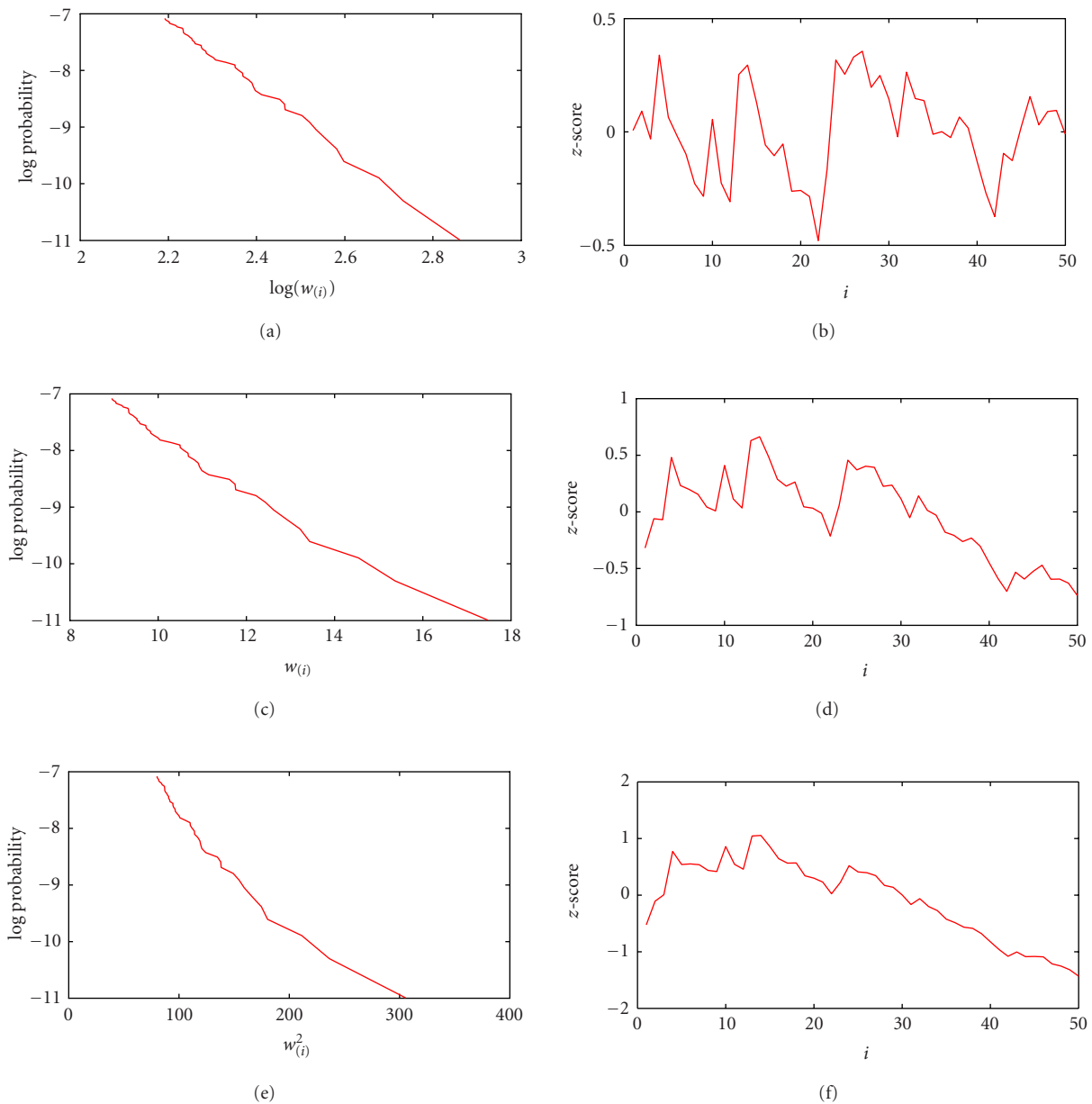| Size of subsample	4th moment	5th moment	6th moment	7th moment	8th moment
n	30.0826	262.6756	2.7390×10^3	3.2494×10^4	4.2430×10^5
$n/2$	29.7100	256.3815	2.6219×10^3	2.9697×10^4	3.7376×10^5
$n/2^2$	29.6708	250.0520	2.4333×10^3	2.6237×10^4	3.0239×10^5
$n/2^3$	29.4082	246.3888	2.3158×10^3	2.4013×10^4	2.3956×10^5
$n/2^4$	27.8039	221.9756	1.9615×10^3	1.9239×10^4	1.8785×10^5

FIGURE 5: Plots of log probability $\log(i/n)$ versus (a) $\log(|w|_{(i)})$, (c) $|w|_{(i)}$, and (e) $|w|_{(i)}^2$ for $1 \leq i \leq 50$, corresponding to the power-law/exponential/Gaussian models we introduced in Section 3; w are the wavelet coefficients of the finest scale (i.e., highest frequencies). Normalized z-score as defined in (25) for (b) the power-law, (d) exponential, and (c) Gaussian models again for $1 \leq i \leq 50$.

TABLE 2: Table of α values for which the different wavelet bands of the CS map.

Multiscale method	Alpha
Biorthogonal wavelet	
Scale 1, horizontal	6.13
Scale 1, vertical	4.84
Scale 1, diagonal	4.27
Scale 2, horizontal	5.15
Scale 2, vertical	4.19
Scale 2, diagonal	3.83
Scale 3, horizontal	4.94
Scale 3, vertical	4.99
Scale 3, diagonal	4.51
Scale 4, horizontal	3.26
Scale 4, vertical	3.37
Scale 4, diagonal	3.76

we also found that the standard deviation of this estimate $\approx$ 0.9. Table 2 gives estimates of α for each band of the wavelet transform. This shows that α is likely to be only slightly less than 8: this means the performance of excess kurtosis and Max might be very close empirically.

3.4. Comparison of excess Kurtosis and Max with simulation

To test the results in Section 3.3, we now perform a small simulation experiment. A complete cycle includes the following steps ($n = 244^2$ and $\{w_i\}_{i=1}^n$ are the same as in Section 3.3).

(1) Let λ range from 0 to 0.1 with increment 0.0025.
(2) Draw $(z_1, z_2, \ldots, z_n)$ independently from $N(0,1)$ to represent the transform coefficients for CMB.
(3) For each λ, let

$$X_i = X_i^{(\lambda)} = \sqrt{1-\lambda}z_i + \sqrt{\lambda}w_i, \quad \lambda = 0, 0.0025, \ldots, 0.1, \tag{28}$$

represent the transform coefficients for CMB + CS.
(4) Apply detectors κ_n, M_n to the $X_i^{(\lambda)}$'s; and obtain the p-values.

We repeated steps (3)-(4) independently 500 times.

Based on these simulations, first, we have estimated the probability of detection under various λ, for each detector:

$$\text{Fraction of detections}$$
$$= \frac{\text{number of cycles with a } p\text{-value} \leq 0.05}{500}. \tag{29}$$

Results are summarized in Figure 6.

Second, we pick out those simulated values for $\lambda = 0.05$ alone, and plot the ROC curves for each detector. The ROC curve is a standard way to evaluate detectors [41]; the x-axis gives the fraction of false alarms (the fraction of detections

when the null is true (i.e., $\lambda = 0$)); the y-axis gives the corresponding fraction of true detections). Results are shown in Figure 6. The figure suggests that the excess kurtosis is slightly better than M_n. We also show an adaptive test, HC_n in two forms (HC_n^* and HC_n^+); these will be described later.

We now interpret. As our analysis predicts that W has a power-law tail with $E[W^8] = \infty$, it is surprising that excess kurtosis still performs better than Max.

In Section 2.3, we compared excess kurtosis and Max in a heuristic way; here we will continue that discussion, using now empirical results. Notice that for the data set $(w_1, w_2, \ldots, w_n)$, the largest (absolute) observation is

$$M = M_n = 17.48, \tag{30}$$

and the excess kurtosis is

$$\kappa = \kappa_n = \frac{1}{n}\left[\sum_i w_i^4\right] - 3 = 27.08. \tag{31}$$

In the asymptotic analysis of Section 2.3, we assumed $\kappa(W)$ is a constant. However, for $n = 244^2$, we get a very large excess kurtosis $27.08 \approx n^{0.3}$; this will make excess kurtosis very favorable in the current situation.

Now, in order for M_n to work successfully, we have to take λ to be large enough that

$$\sqrt{\lambda}M > \sqrt{2\log n}, \tag{32}$$

so $\lambda > 0.072$. The p-value of M_n is then

$$\exp\left(-\exp\left(-\frac{\sqrt{\lambda}M - 4.2627}{0.2125}\right)\right), \tag{33}$$

moreover, the p-value for excess kurtosis is, heuristically,

$$\bar{\Phi}^{-1}(\sqrt{n}\lambda^2\kappa); \tag{34}$$

setting them to be equal, we can solve κ in terms of M:

$$\kappa = \kappa_0(M). \tag{35}$$

The curve $\kappa = \kappa_0(M)$ separates the M-κ plane into 2 regions: the region above the curve is favorable to the excess kurtosis, and the region below the curve is favorable to Max; see Figure 7. In the current situation, the point $(M, \kappa) = (17.48, 27.08)$ falls far above the curve; this explains why excess kurtosis is better than Max for the current data set.

3.5. Experiments on Wavelet coefficients

3.5.1. CMB + CS

We study the relative sensitivity of the different wavelet-based statistical methods when the signals are added to a dominant Gaussian noise, that is, the primary CMB.
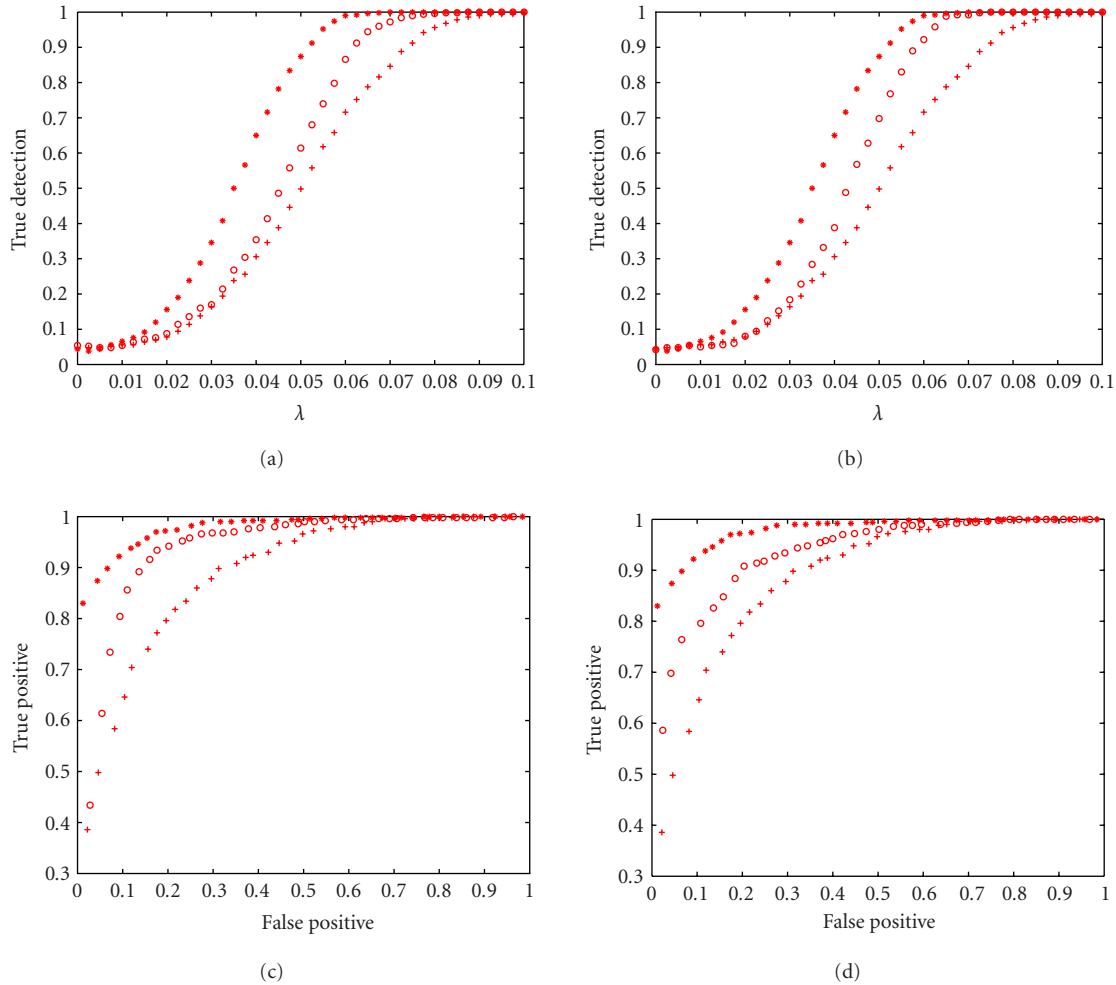
FIGURE 6: (a) The fraction of detection for the excess kurtosis, HC^*, and Max; the x-axis is the corresponding λ; (b) the fraction of detection for kurtosis, HC^+, and Max. (c) ROC curves for the excess kurtosis, HC^*, and Max; (d) ROC curves for the excess kurtosis, HC^+, and Max.

We ran 5000 simulations by adding the 100 CMB realisations to the CS ($D(\lambda, i) = \sqrt{1 - \lambda}\text{CMB}_i + \sqrt{\lambda}\text{CS}$, $i = 1 \cdots 100$), using 50 different values for λ, ranging linearly between 0 and 0.18. Then we applied the biorthogonal wavelet transform, using the standard 7/9 filter [42] to these 5000 maps. On each band b of the wavelet transform, for each dataset $D(\lambda, i)$, we calculate the kurtosis value $K_{D(b,\lambda,i)}$. In order to calibrate and compare the departures from a Gaussian distribution, we have simulated for each image $D(\lambda, i)$ a Gaussian random field $G(\lambda, i)$ which has the same power spectrum as $D(\lambda, i)$, and we derive its kurtosis values $K_{G(b,\lambda,i)}$. For a given band b and a given λ, we derive for each kurtosis $K_{D(b,\lambda,i)}$ its p-value $p_K(b, \lambda, i)$ under the null hypothesis (i.e., no CS) using the distribution of $K_{G(b,\lambda,*)}$. The mean p-value $\bar{p}_K(b, \lambda)$ is obtained by taking the mean of $p_K(b, \lambda, *)$. For a given band b, the curve $\bar{p}_K(b, \lambda)$ versus λ shows the sensitivity of the method to detect CS. Then we do the same operation, but replacing the kurtosis by HC and Max. Figure 8 shows the mean p-value versus λ for the nine finest scale subbands of the wavelet

transform. The first three subbands correspond to the finest scale (high frequencies) in the three directions, respectively, horizontal, vertical, and diagonal. Bands 4, 5, and 6 correspond to the second resolution level and bands 7, 8, 9 to the third. Results are clearly in favor of the excess kurtosis.

The same experiments have been repeated, but replacing the biorthogonal wavelet transform by the undecimated isotropic à trous wavelet transform. Results are similarly in favor of the excess kurtosis. Table 3 gives the λ values (multiplied 100) for which the CS are detected at a 95% confidence level. Only bands where this level is achieved are given. The smaller the λ, the better the sensibility of the method to the detect the CS. These results show that the excess kurtosis outperforms clearly HC and Max, whatever the chosen multiscale transform and the analyzed scale.

No method is able to detect the CS at a 95% confidence level after the second scale in these simulations. In practice, the presence of noise makes the detection even more difficult, especially in the finest scales.
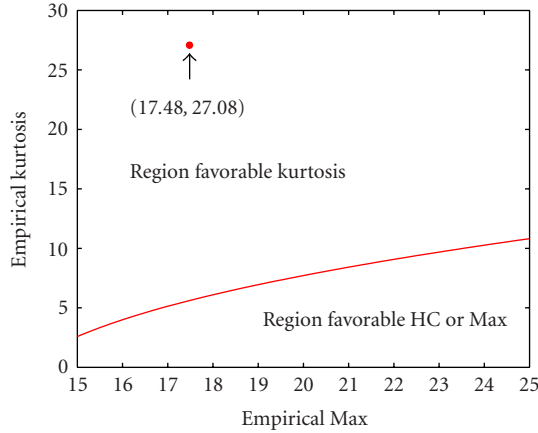
FIGURE 7: The M-κ plane and the curve $\kappa = \kappa_0(M)$, where M is the largest (absolute) observation of w_i's, and κ is the empirical excess kurtosis of w_i's, where w_i's are the wavelet coefficients of the simulated cosmic string. Heuristically, if (M, κ) falls above the curve, excess kurtosis will perform better than Max. The red bullet represents the points of $(M, \kappa) = (17.48, 27.08)$ for the current data set w_i's, which is far above the curve.

3.5.2. CMB + SZ

We now consider a totally different contamination. Here, we take into account the secondary anisotropies due to the kinetic Sunyaev-Zel'dovich (SZ) effect [35]. The SZ effect represents the Compton scattering of CMB photons by the free electrons of the ionised and hot intracluster gas. When the galaxy cluster moves with respect to the CMB rest frame, the Doppler shift induces additional anisotropies; this is the so-called kinetic SZ (KSZ) effect. The kinetic SZ maps are simulated following Aghanim et al. [43] and the simulated observed map D is obtained from $D_\lambda = $ CMB $+ \lambda$KSZ, where CMB and KSZ are, respectively, the CMB and the kinetic SZ simulated maps. We ran 5000 simulations by adding the 100 CMB realisations to the KSZ ($D(\lambda, i) = $ CMB$_i + \lambda$KSZ, $i = 1 \cdots 100$), using 50 different values for λ, ranging linearly between 0 and 1. The p-values are calculated just as in the previous section.

Table 4 gives the λ values for which SZ is detected at a 95% confidence level for the three multiscale transforms. Only bands where this level is achieved are given. Results are again in favor of the Kurtosis.

4. CURVELET COEFFICIENTS OF FILAMENTS

Curvelet analysis was proposed by Candès and Donoho (1999) [44] as a means to efficiently represent edges in images; Donoho and Flesia (2001) [45] showed that it could also be used to describe non-Gaussian statistics in natural images. It has also been used for a variety of image processing tasks [46, 47, 48]. We now consider the use of curvelet analysis for detection of non-Gaussian cosmological structures which are filamentary.

4.1. Curvelet coefficients of filaments

Suppose we have an image I which contains within it a single filament, that is, a smooth curve of appreciable length L. We analyse it using the curvelet frame. Applying analysis techniques described carefully in [49], we can make precise the following claim: *at scale $s = 2^{-j}$, there will be about $O(L2^{j/2})$ significant coefficients caused by this filamentary feature, and they will all be of roughly similar size. The remaining $O(4^j)$ coefficients at that scale will be much smaller, basically zero in comparison.*

The pattern continues in this way if there is a collection of m filaments of individual lengths L_i and total length $L = L_1 + \cdots + L_m$. Then we expect roughly $O(L2^{j/2})$ substantial coefficients at level j, out of 4^j total.

This suggests a rough model for the analysis of non-Gaussian random images which contain apparent "edge-like" phenomena. If we identify the edges with filaments, then we expect to see, at a scale with n coefficients, about $Ln^{1/4}$ nonzero coefficients. Assuming all the edges are equally "pronounced," this suggests that we view the curvelet coefficients of I at a given scale as consisting of a fraction $\epsilon = L/n^{3/4}$ nonzeros and the remainder zero. Under this model, the curvelet coefficients of a superposition of a Gaussian random image should behave like

$$X_i = (1 - \epsilon)N(0, 1) + \frac{\epsilon}{2}N(-\mu, 1) + \frac{\epsilon}{2}N(\mu, 1), \quad (36)$$

where ϵ is the fraction of large curvelet coefficients corresponding to filaments, and μ is the amplitude of these coefficients of the non-Gaussian component N.

The problem of detecting the existence of such a non-Gaussian mixture is equivalent to discriminating between the hypotheses

$$H_0 : X_i \overset{iid}{\sim} N(0, 1), \quad (37)$$

$$H_1^{(n)} : X_i = (1 - \epsilon_n)N(0, 1) + \frac{\epsilon_n}{2}N(-\mu_n, 1) + \frac{\epsilon_n}{2}N(\mu_n, 1), \quad (38)$$

and $N \equiv 0$ is equivalent to $\epsilon_n \equiv 0$.

4.2. Optimal detection of sparse mixtures

When both ϵ and μ are known, the optimal test for problem (37)-(38) is simply the Neyman-Pearson likelihood ratio test (LRT), [32, page 74]. Asymptotic analysis shows the following [50, 51].

Suppose we let $\epsilon_n = n^{-\beta}$ for some exponent $\beta \in (1/2, 1)$, and

$$\mu_n = \sqrt{2s \log(n)}, \quad 0 < s < 1. \quad (39)$$

There is a *threshold effect*: setting

$$\rho_2^*(\beta) = \begin{cases} \beta - \frac{1}{2}, & \frac{1}{2} < \beta \le \frac{3}{4}, \\ \left(1 - \sqrt{1 - \beta}\right)^2, & \frac{3}{4} < \beta < 1, \end{cases} \quad (40)$$
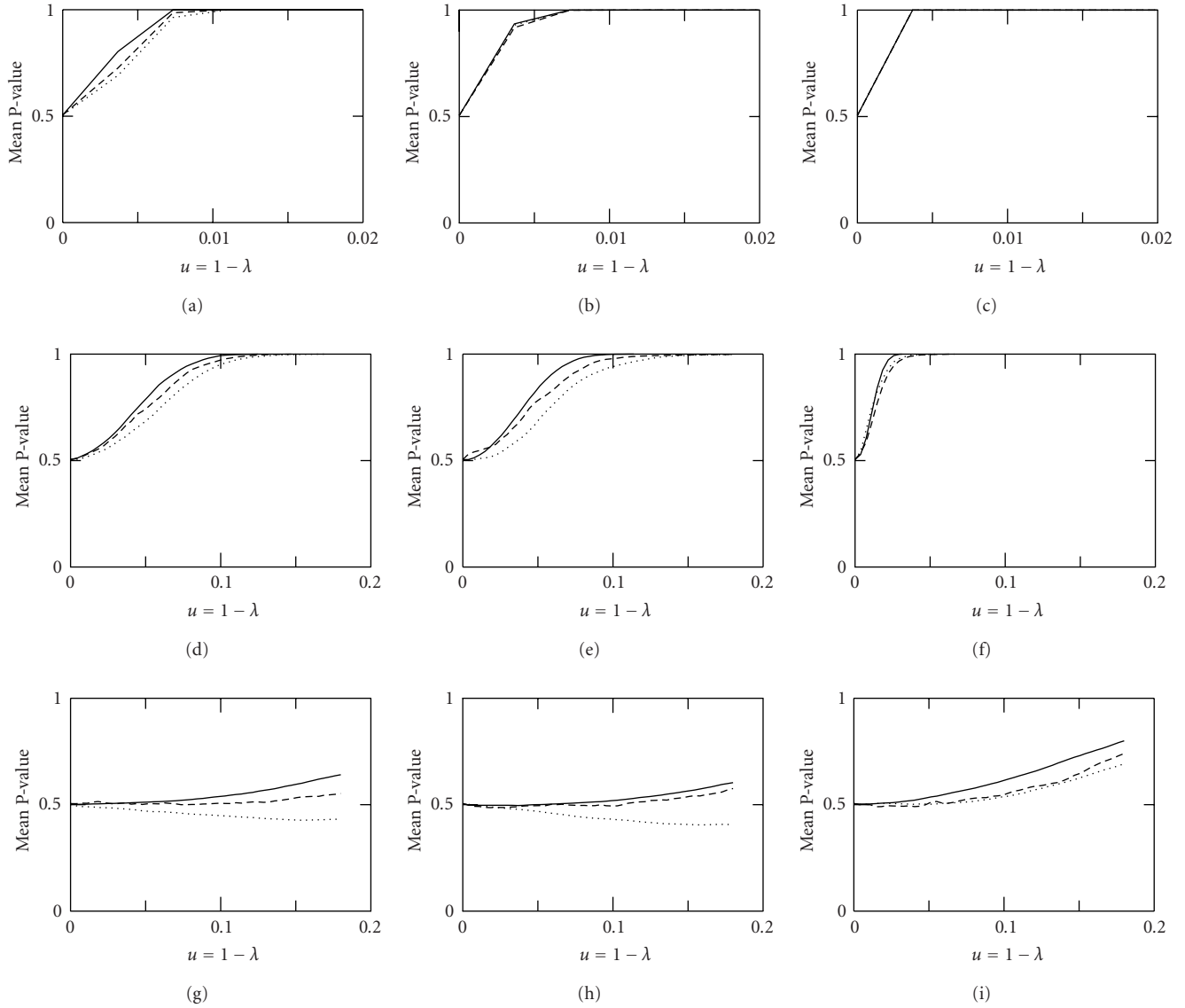
FIGURE 8: For the nine first bands of the wavelet transform, the mean p-value versus λ. The solid, dashed, and dotted lines correspond, respectively, to the excess kurtosis, the HC, and Max ((a) band 1, (b) band 2, (c) band 3, (d) band 4, (e) band 5, (f) band 6, (g) band 7, (h) band 8, and (i) band 9).

then, as $n \to \infty$, LRT is able to reliably detect for large n when $s > \rho_2^*(\beta)$, and is unable to detect when $s < \rho_2^*(\beta)$ [31, 50, 51]. Since LRT is optimal, it is not possible for any statistic to reliably detect when $s < \rho_2^*(\alpha)$. We call the curve $s = \rho_2^*(\beta)$ in the β-s plane the *detection boundary*; see Figure 9.

We also remark that if the sparsity parameter $\beta < 1/2$, it is possible to discriminate merely using the value of the empirical variance of the observations or some other simple moments, and so there is no need for advanced theoretical approaches.

4.3. Adaptive testing using higher criticism

The higher criticism statistic (HC), was proposed in [31], where it was proved to be asymptotically optimal in detecting (37)-(38).

To define HC, first we convert the individual X_i's into p-values for individual z-tests. Let $p_i = P\{N(0, 1) > X_i\}$ be the ith p-value, and let $p_{(i)}$ denote the p-values *sorted in increasing order*; the higher criticism statistic is defined as

$$HC_n^* = \max_i \left| \frac{\sqrt{n}[i/n - p_{(i)}]}{\sqrt{p_{(i)}(1 - p_{(i)})}} \right|, \tag{41}$$

or in a modified form:

$$HC_n^+ = \max_{\{i:1/n \le p_{(i)} \le 1-1/n\}} \left| \frac{\sqrt{n}[i/n - p_{(i)}]}{\sqrt{p_{(i)}(1 - p_{(i)})}} \right|; \tag{42}$$

TABLE 3: Table of λ values (multiplied by 100) for CS detections at 95% confidence.

Multiscale method	Excess kurtosis	HC	Max
Biorthogonal wavelet			
Scale 1, horizontal	0.73	0.73	0.73
Scale 1, vertical	0.73	0.73	0.73
Scale 1, diagonal	0.38	0.38	0.38
Scale 2, horizontal	8.01	9.18	8.81
Scale 2, vertical	6.98	8.44	10.65
Scale 2, diagonal	2.20	2.94	2.57
À trous wavelet transform			
Scale 1	1.47	1.47	1.47
Scale 2	9.91	12.85	16.53
Curvelet			
Scale 1, band 1	1.47	2.20	3.30
Scale 1, band 2	13.59	16.90	—
Scale 2, band 1	11.38	14.32	—

TABLE 4: Table of λ values for which the SZ detections at 95% confidence.

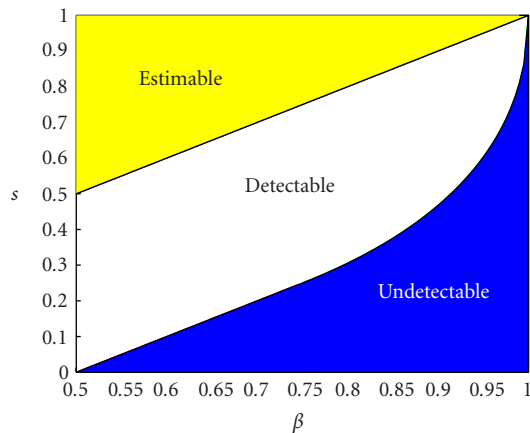| Multiscale method	Excess kurtosis	HC	Max
Biorthogonal wavelet			
Scale 1, horizontal	0.30	0.32	—
Scale 1, vertical	0.32	0.32	—
Scale 1, diagonal	0.06	0.06	0.24
Scale 2, horizontal	—	—	—
Scale 2, vertical	—	—	—
Scale 2, diagonal	0.65	0.71	—
À trous wavelet transform			
Scale 1	0.41	0.47	—
Curvelet			
Scale 1, band 1	0.59	0.69	0.83

FIGURE 9: The detection boundary separates the square in the β-s plane into the detectable region and the undetectable region. When (β, s) falls into the estimable region, it is possible not only to reliably detect the presence of the signals, but also estimate them.

we let HC_n refer either to HC_n^* or HC_n^+ whenever there is no confusion. The above definition is slightly different from [31], but the ideas are essentially the same.

With an appropriate normalization sequence:

$$a_n = \sqrt{2 \log \log n},$$
$$b_n = 2 \log \log n + 0.5 \log \log \log n - 0.5 \log(4\pi), \quad (43)$$

the distribution of HC_n converges to the Gumbel distribution E_v^4, whose cdf is $\exp(-4 \exp(-x))$ [52]:

$$a_n HC_n - b_n \longrightarrow_w E_v^4, \quad (44)$$

so the p-values of HC_n are approximately

$$\exp\left(-4 \exp\left(-[a_n HC_n - b_n]\right)\right). \quad (45)$$

For moderately large n, in general, the approximation in (45) is accurate for the HC_n^+, but not for HC_n^*. For $n = 244^2$, taking $a_n = 2.2536$ and $b_n = 3.9407$ in (45) gives a good approximation for the p-value of HC_n^+.
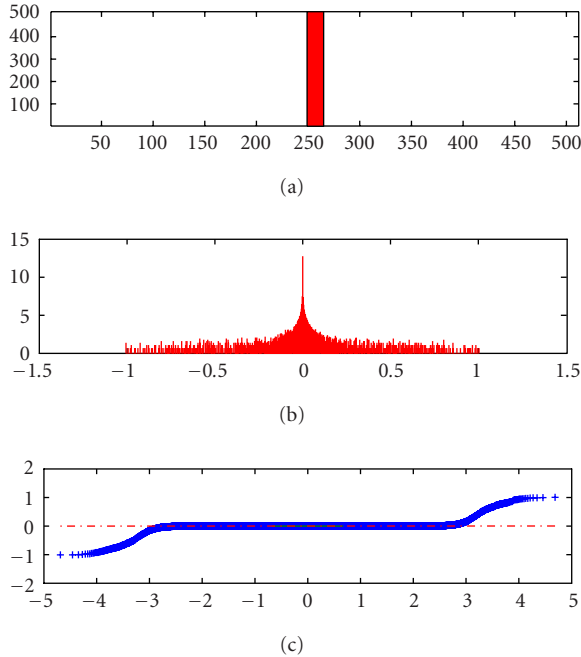
(a)

(b)

(c)

FIGURE 10: (a) The image of the bar, (b) the log-histogram of the curvelet coefficients of the bar, and (c) the qq-plot of the curvelet coefficients versus normal distribution.

(a)

(b)

(c)

FIGURE 11: For the curvelet coefficients v_i's of the simulated CS map in Figure 3, (a) the log-histogram of v_i's, (b) the qq-plot of v_i's versus normal distribution, and (c) the qq-plot of $\text{sign}(v_i)|v_i|^{0.815}$ versus double exponential.

A brief remark comparing Max and HC. Max only takes into account the few largest observations, HC takes into account those outliers, but also moderate large observations; as a result, in general, HC is better than Max, especially when we have unusually many moderately large observations. However, when the actual evidence lies in the middle of the distribution, both HC and Max will be very weak.

4.4. Curvelet coefficients of cosmic strings

In Section 3, we studied wavelet coefficients of simulated cosmic strings. We now study the curvelet coefficients of the same simulated maps.

We now discuss empirical properties of curvelet coefficients of (simulated) cosmic strings. This was first deployed on a test image showing a simple "bar" extending vertically across the image. The result, seen in Figure 10, shows the image, the histogram of the curvelet coefficients at the next-to-finest scale, and the qq-plot against the normal distribution. The display matches in general terms the sparsity model of Section 4. That display also shows the result of superposing Gaussian noise on the image; the curvelet coefficients clearly have the general appearance of a mixture of normals with sparse fractions at nonzero mean, just as in the model.

We also applied the curvelet transform to the simulated cosmic string data. Figure 11 shows the results, which suggest that the coefficients do not match the simple sparsity model. Extensive modelling efforts, not reported here, show that the curvelet coefficients transformed by $|v|^{0.815}$ have an exponential distribution.
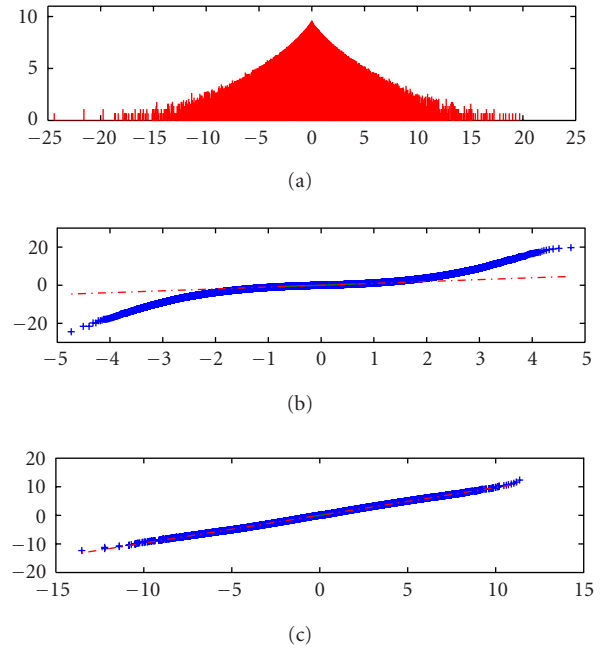
This discrepancy from the sparsity model has two explanations. First, cosmic string images contain (to the naked eye) both point-like features and curve-like features. Because curvelets are not specially adapted to sparsifying point-like features, the coefficients contain extra information not expressible by our geometric model. Second, cosmic string images might contain filamentary features at a range of length scales and a range of density contrasts. If those contrasts exhibit substantial amplitude variation, the simple mixture model must be replaced by something more complex. In any event, the curvelet coefficients of cosmic strings do not have the simple structure proposed in Section 4.

When applying various detectors of non-Gaussian behavior to curvelet coefficients, as in the simulation of Section 3.5, we find that, despite the theoretical ideas backing the use of HC as an optimal test for sparse non-Gaussian phenomena, the kurtosis consistently has better performance. The results are included in Tables 3 and 4.

Note that, although the curvelet coefficients are not as sensitive detectors as wavelets in this setting, that can be an advantage, since they are relatively immune to point-like features such as SZ contaimination. Hence, they are more specific to CS as opposed to SZ effects.

5. CONCLUSION

The kurtosis of the wavelet coefficients is very often used in astronomy for the detection non-Gaussianities in the CMB.

It has been shown [27] that it is also possible to separate the non-Gaussian signatures associated with cosmic strings from those due to SZ effect by combining the excess kurtosis derived from these both the curvelet and the wavelet transform. We have studied in this paper several other transform-based statistics, the Max and the higher criticism, and we have compared them theoretically and experimentally to the kurtosis. We have shown that kurtosis is asymptotically optimal in the class of weakly dependent symmetric non-Gaussian contamination with finite 8th moments, while HC and Max are asymptotically optimal in the class of weakly dependent symmetric non-Gaussian contamination with infinite 8th moment. Hence, depending on the nature of the non-Gaussianity, a statistic is better than another one. This is a motivation for using several statistics rather than a single one, for analysing CMB data. Finally, we have studied in details the case of cosmic string contaminations on simulated maps. Our experiment results show clearly that kurtosis outperforms Max/HC.

ACKNOWLEDGMENT

REFERENCES

[1] A. A. Penzias and R. W. Wilson, "Measurement of the flux density of CAS A at 4080 Mc/s," *Astrophysical Journal*, vol. 142, no. 1, pp. 1149–1155, 1965.

[2] D. J. Fixsen, E. S. Cheng, D. A. Cottingham, et al., "A balloon-borne millimeter-wave telescope for cosmic microwave background anisotropy measurements," *Astrophysical Journal*, vol. 470, pp. 63–63, 1996.

[3] G. F. Smoot, C. L. Bennett, A. Kogut, et al., "Structure in the COBE differential microwave radiometer first-year maps," *Astrophysical Journal Letters*, vol. 396, no. 1, pp. L1–L5, 1992.

[4] A. D. Miller, R. Caldwell, M. J. Devlin, et al., "A measurement of the angular power spectrum of the cosmic microwave background from $l = 100$ to 400," *Astrophysical Journal Letters*, vol. 524, no. 1, pp. L1–L4, 1999.

[5] P. de Bernardis, P. A. R. Ade, J. J. Bock, et al., "A flat universe from high-resolution maps of the cosmic microwave background radiation," *Nature*, vol. 404, no. 6781, pp. 955–959, 2000.

[6] S. Hanany, P. A. R. Ade, A. Balbi, et al., "MAXIMA-1: a measurement of the cosmic microwave background anisotropy on angular scales of $10' - 5°$," *Astrophysical Journal Letters*, vol. 545, no. 1, pp. L5–L9, 2000.

[7] N. W. Halverson, E. M. Leitch, C. Pryke, et al., "Degree angular scale interferometer first results: a measurement of the cosmic microwave background angular power spectrum," *Astrophysical Journal*, vol. 568, no. 1, pp. 38–45, 2002.

[8] A. Benoît, P. A. R. Ade, A. Amblard, et al., "The cosmic microwave background anisotropy power spectrum measured by Archeops," *Astronomy & Astrophysics*, vol. 399, no. 3, pp. L19–L23, 2003.

[9] C. L. Bennett, R. S. Hill, G. Hinshaw, et al., "First-year Wilkinson microwave anisotropy probe (WMAP) observations: foreground emission," *Astrophysical Journal Supplement Series*, vol. 148, no. 1, pp. 97–117, 2003.

[10] C. L. Bennett, M. Halpern, G. Hinshaw, et al., "First-year Wilkinson microwave anisotropy probe (WMAP) observations: preliminary maps and basic results," *Astrophysical Journal Supplement Series*, vol. 148, no. 1, pp. 1–27, 2003.

[11] F. Bernardeau and J. Uzan, "Non-Gaussianity in multifield inflation," *Physical Review D*, vol. 66, no. 10, 103506, 14 pages, 2002.

[12] X. Luo, "The angular bispectrum of the cosmic microwave background," *Astrophysical Journal Letters*, vol. 427, no. 2, pp. L71–L74, 1994.

[13] A. H. Jaffe, "Quasilinear evolution of compensated cosmological perturbations: the nonlinear σ model," *Physical Review D*, vol. 49, no. 8, pp. 3893–3909, 1994.

[14] A. Gangui, F. Lucchin, S. Matarrese, and S. Mollerach, "The three-point correlation function of the cosmic microwave background in inflationary models," *Astrophysical Journal*, vol. 430, no. 2, pp. 447–457, 1994.

[15] D. Novikov, J. Schmalzing, and V. F. Mukhanov, "On non-Gaussianity in the cosmic microwave background," *Astronomy & Astrophysics*, vol. 364, pp. 17–25, 2000.

[16] S. F. Shandarin, "Testing non-Gaussianity in cosmic microwave background maps by morphological statistics," *Monthly Notices of the Royal Astronomical Society*, vol. 331, no. 4, pp. 865–874, 2002.

[17] B. C. Bromley and M. Tegmark, "Is the cosmic microwave background really non-Gaussian?" *Astrophysical Journal Letters*, vol. 524, no. 2, pp. L79–L82, 1999.

[18] L. Verde, L. Wang, A. F. Heavens, and M. Kamionkowski, "Large-scale structure, the cosmic microwave background and primordial non-Gaussianity," *Monthly Notices of the Royal Astronomical Society*, vol. 313, no. 1, pp. 141–147, 2000.

[19] N. G. Phillips and A. Kogut, "Statistical power, the bispectrum, and the search for non-Gaussianity in the cosmic microwave background anisotropy," *Astrophysical Journal*, vol. 548, no. 2, pp. 540–549, 2001.

[20] M. Kunz, A. J. Banday, P. G. Castro, P. G. Ferreira, and K. M. Górski, "The trispectrum of the 4 year COBE DMR data," *Astrophysical Journal Letters*, vol. 563, no. 2, pp. L99–L102, 2001.

[21] N. Aghanim and O. Forni, "Searching for the non-Gaussian signature of the CMB secondary anisotropies," *Astronomy & Astrophysics*, vol. 347, pp. 409–418, 1999.

[22] O. Forni and N. Aghanim, "Searching for non-Gaussianity: statistical tests," *Astronomy and Astrophysics Supplement Series*, vol. 137, no. 3, pp. 553–567, 1999.

[23] M. P. Hobson, A. W. Jones, and A. N. Lasenby, "Wavelet analysis and the detection of non-Gaussianity in the cosmic microwave background," *Monthly Notices of the Royal Astronomical Society*, vol. 309, no. 1, pp. 125–140, 1999.

[24] R. B. Barreiro and M. P. Hobson, "The discriminating power of wavelets to detect non-Gaussianity in the cosmic microwave background," *Monthly Notices of the Royal Astronomical Society*, vol. 327, no. 3, pp. 813–828, 2001.

[25] L. Cayón, J. L. Sanz, E. Martínez-González, et al., "Spherical mexican hat wavelet: an application to detect non-Gaussianity in the COBE-DMR maps," *Monthly Notices of the Royal Astronomical Society*, vol. 326, no. 4, pp. 1243–1248, 2001.

[26] J. Jewell, "A statistical characterization of galactic dust emission as a non-Gaussian foreground of the cosmic microwave background," *Astrophysical Journal*, vol. 557, no. 2, pp. 700–713, 2001.

[27] J.-L. Starck, N. Aghanim, and O. Forni, "Detection and discrimination of cosmological non-Gaussian signatures by multi-scale methods," *Astronomy & Astrophysics*, vol. 416, no. 1, pp. 9–17, 2004.

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

[29] S. G. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, Calif, USA, 1998.

[30] N. Aghanim, M. Kunz, P. G. Castro, and O. Forni, "Non-Gaussianity: comparing wavelet and fourier based methods," *Astronomy & Astrophysics*, vol. 406, no. 3, pp. 797–816, 2003.

[31] D. L. Donoho and J. Jin, "Higher criticism for detecting sparse heterogeneous mixtures," *Annals of Statistics*, vol. 32, no. 3, pp. 962–994, 2004.

[32] E. L. Lehmann, *Testing Statistical Hypotheses*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1986.

[33] D. L. Donoho and J. Jin, "Optimality of excess kurtosis for detecting a non-Gaussian component in high-dimensional random vectors," Tech. Rep., Stanford University, Stanford, Calif, USA, 2004.

[34] M. White and J. D. Cohn, "TACMB-1: The theory of anisotropies in the cosmic microwave background," *American Journal of Physics*, vol. 70, no. 2, pp. 106–118, 2002.

[35] 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*, vol. 18, pp. 537–560, 1980.

[36] J. P. Ostriker and E. T. Vishniac, "Generation of microwave background fluctuations from nonlinear perturbations at the ERA of galaxy formation," *Astrophysical Journal*, vol. 306, pp. L51–L54, 1986.

[37] E. T. Vishniac, "Reionization and small-scale fluctuations in the microwave background," *Astrophysical Journal*, vol. 322, pp. 597–604, 1987.

[38] F. R. Bouchet, P. Peter, A. Riazuelo, and M. Sakellariadou, "Evidence against or for topological defects in the BOOMERanG data?" *Physical Review D*, vol. 65, no. 2, 021301(R), 4 pages, 2002.

[39] F. R. Bouchet, D. P. Bennett, and A. Stebbins, "Patterns of the cosmic microwave background from evolving string networks," *Nature*, vol. 335, no. 6189, pp. 410–414, 1988.

[40] B. M. Hill, "A simple general approach to inference about the tail of a distribution," *Annals of Statistics*, vol. 3, no. 5, pp. 1163–1174, 1975.

[41] C. E. Metz, "Basic principles of ROC analysis," *Seminars in Nuclear Medicine*, vol. 8, no. 4, pp. 283–298, 1978.

[42] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Processing*, vol. 1, no. 2, pp. 205–220, 1992.

[43] N. Aghanim, K. M. Górski, and J.-L. Puget, "How accurately can the SZ effect measure peculiar cluster velocities and bulk flows?" *Astronomy & Astrophysics*, vol. 374, no. 1, pp. 1–12, 2001.

[44] 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., Vanderbilt University Press, Nashville, Tenn, USA, 1999.

[45] D. L. Donoho and A. G. Flesia, "Can recent developments in harmonic analysis explain the recent findings in natural scene statistics?" *Network: Computation in Neural Systems*, vol. 12, no. 3, pp. 371–393, 2001.

[46] E. J. Candès and D. L. Donoho, "Edge-preserving denoising in linear inverse problems: optimality of curvelet frames," *Annals of Statistics*, vol. 30, no. 3, pp. 784–842, 2002.

[47] J.-L. Starck, E. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Processing*, vol. 11, no. 6, pp. 670–684, 2002.

[48] J.-L. Starck, F. Murtagh, E. J. Candès, and D. L. Donoho, "Gray and color image contrast enhancement by the curvelet transform," *IEEE Trans. Image Processing*, vol. 12, no. 6, pp. 706–717, 2003.

[49] E. J. Candès and D. L. Donoho, "Frames of curvelets and optimal representations of objects with piecewise c^2 singularities," *Communications On Pure and Applied Mathematics*, vol. 57, no. 2, pp. 219–266, 2004.

[50] Yu. I. Ingster, "Minimax detection of a signal for l^n-balls," *Mathematical Methods of Statistics*, vol. 7, no. 4, pp. 401–428, 1999.

[51] J. Jin, "Detecting a target in very noisy data from multiple looks," *IMS Lecture Notes Monograph*, vol. 45, pp. 1–32, 2004.

[52] G. R. Shorack and J. A. Wellner, *Empirical Processes with Applications to Statistics*, John Wiley & Sons, New York, NY, USA, 1986.

J. Jin received the B.S. and M.S. degrees in mathematics from Peking University, China, the M.S. degree in applied mathematics from the University of California at Los Angeles (UCLA), California, and the Ph.D. degree in statistics from Stanford University, Stanford, California, where D. L. Donoho served as his adviser. He is an Assistant Professor of statistics at Purdue University, Indiana. His research interests are in the area of large-scale multiple hypothesis testing, statistical estimation, and their applications to protein mass spectroscopy and astronomy.

J.-L. Starck has a Ph.D. degree from University Nice-Sophia Antipolis and a 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, and statistical methods in astrophysics and cosmology. 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, 2002).

D. L. Donoho is Anne T. and Robert M. Bass Professor in the humanities and sciences at Stanford University. He received his A.B. degree in statistics from Princeton University where his thesis adviser was John W. Tukey and his Ph.D. degree in statistics from Harvard University where his thesis adviser was Peter J. Huber. He is a Member of the US National Academy of Sciences and of the American Academy of Arts and Sciences.

N. Aghanim is a Cosmologist at the Institut d'Astrophysique Spatiale in Orsay (France). Her main research interests are cosmic microwave background (CMB) and large-scale structure. She has been working during the last ten years on the statistical characterisation of CMB temperature anisotropies through power spectrum analyses and higher-order moments of wavelet coefficients. She naturally got involved and interested in signal processing techniques in order to improve the detection of low signal to noise such as those associated with secondary anisotropies and separate them from the primary signal.

O. Forni is a Planetologist at the Institut d'Astrophysique Spatiale in Orsay (France). His main research activities deal with the evolution of the planets and satellites of the solar system. Recently he has been working on the statistical properties of the cosmic microwave background (CMB) and of the secondaries anisotropies by means of multiscale transforms analysis. He also got involved in component separation techniques in order to improve the detection of low power signatures and to analyse hyperspectral infrared data on Mars.