

True cosmic microwave background power spectrum estimation

P. Paykari¹*, J.-L. Starck¹, and M. J. Fadili²

¹ Laboratoire AIM, UMR CEA-CNRS-Paris 7, Irfu, SAp/SEDI, Service d'Astrophysique, CEA Saclay, F-91191 GIF-SUR-YVETTE CEDEX, France.

² GREYC CNRS UMR 6072, ENSICAEN, 6 Bd du Maréchal Juin, 14050 Caen Cedex, France

March 6, 2012

ABSTRACT

Context.

Aims. The cosmic microwave background (CMB) power spectrum is a powerful cosmological probe as it entails almost the entire statistical information of CMB perturbations. Having access to only one sky, the CMB power spectrum measured by our experiments is only a realization of the true underlying angular power spectrum. We aim to recover the true underlying CMB power spectrum from the one realization that we have without knowing the cosmological parameters.

Methods. The sparsity of the CMB power spectrum is first investigated in two dictionaries; discrete cosine transform (DCT) and wavelet transform (WT). The CMB power spectrum can be recovered with very few coefficients in these two dictionaries and hence is very compressible.

Results. We studied the performance of these dictionaries in smoothing a set of simulated power spectra. Based on this, we developed a technique that estimates the true underlying CMB power spectrum from data, i.e., without a need to know the cosmological parameters.

Conclusions. This smooth estimated spectrum can be used to simulate CMB maps with similar properties as the true CMB simulations with the correct cosmological parameters. This allows us to perform Monte Carlo simulations in a given project without having to know the cosmological parameters. The developed IDL code, **TOUSI**, for theoretical power spectrum using sparse estimation, will be released with the next version of ISAP.

Key words. Keywords should be given

Cosmology : Cosmic Microwave Background, Methods : Data Analysis, Methods : Statistical

1. Introduction

Measurements of the cosmic microwave background (CMB) anisotropies are powerful cosmological probes. In the currently favored cosmological model, with the nearly Gaussian-distributed curvature perturbations, almost the entire statistical information is contained in the CMB angular power spectrum. The observed quantity on the sky is generally the CMB temperature anisotropy $\Theta(\vec{p})$ in direction \vec{p} , which is described as $T(\vec{p}) = T_{CMB}[1 + \Theta(\vec{p})]$. This field is expanded on the spherical harmonic functions as

$$\Theta(\vec{p}) = \sum_{\ell=0}^{+\infty} \sum_{m=-\ell}^{\ell} a[\ell, m] Y_{\ell m}(\vec{p}), \quad (1)$$

$$\text{where } a[\ell, m] = \int_{\mathbb{S}^2} \Theta(\vec{p}) Y_{\ell m}^*(\vec{p}) d\vec{p}, \quad (2)$$

where $\mathbb{S}^2 \subset \mathbb{R}^3$ is the unit sphere, ℓ is the multipole moment, which is related to the angular size on the sky as $\ell \sim 180^\circ/\theta$ and m is the phase ranging from $-\ell$ to ℓ . The $a[\ell, m]$ are the spherical harmonic coefficients of the (noise-free) observed sky. For a Gaussian random field, the mean and covariance are sufficient statistics, meaning that they

carry the whole statistical information of the field. If the random field has zero mean, $\mathbb{E}(a_{00}) = 0$ and the expansion can be started at $\ell = 2$, neglecting the dipole terms, i.e., $\ell = 1$ ¹. For $\ell \geq 2$, the triangular array $(a[\ell, m])_{\ell, m}$ represents zero-mean, complex-valued random coefficients, with variance

$$\mathbb{E}(|a[\ell, m]|^2) = C[\ell] > 0, \quad (3)$$

where $C[\ell]$ is the CMB angular power spectrum, which only depends on ℓ because of the isotropy assumption. Therefore, from Equation 3, an unbiased estimator of $C[\ell]$ is given by the empirical power spectrum

$$\hat{C}[\ell] = \frac{1}{2\ell + 1} \sum_m |a[\ell, m]|^2. \quad (4)$$

Furthermore, because the random field is stationary, the spherical harmonic coefficients are uncorrelated,

$$\mathbb{E}(a[\ell, m] a^*[\ell', m']) = \delta_{\ell\ell'} \delta_{mm'} C[\ell]. \quad (5)$$

Since they are Gaussian, they are also independent. The angular power spectrum depends on the cosmological parameters through an angular transfer function $T_\ell(k)$ as

$$C[\ell] = 4\pi \int \frac{dk}{k} T_\ell^2(k) P(k), \quad (6)$$

¹ The dipole anisotropy is dominated by the Earth's motion in space and it is hence ignored.

* paniez.paykari@cea.fr

where k defines the scale and $P(k)$ is the primordial matter power spectrum.

Making accurate measurements of this power spectrum has been one of the main goals of cosmology in the past two decades. We have seen a range of ground- and balloon-based experiments, such as Acbar (Reichardt et al. 2009) and CBI (Readhead et al. 2004), as well as satellite experiments, such as WMAP (Bennett et al. 2003) and the recently launched satellite Planck². All these experiments produce a temperature map of the sky from which the CMB power spectrum is obtained. The estimation of the power spectrum from CMB experiments is important because this spectrum is a way to estimate the cosmological parameters that describe the Universe. General methods for extracting this spectrum from a N_{pix} -map, with nonuniform coverage and correlated noise are quite expensive and time-consuming especially for Planck, where we will be dealing with $N_{pix} = 5 \times 10^7$. All these experiments estimate the CMB angular power spectrum from a sky map, which is a *realization* of the underlying true power spectrum; no matter how much the experiments improve, we are still limited to an accuracy within the cosmic variance. This means that even if we had a perfect experiment (i.e. with zero instrumental noise), we would not be able to recover a perfect power spectrum, because of the cosmic variance limit.

In this paper we investigate the possibility of estimating the true underlying power spectrum from a *realized* spectrum; an estimation of the true power spectrum without a need to know the cosmological parameters. For this we exploit the sparsity properties of the CMB power spectrum, and capitalize on it to propose an estimator of the theoretical power spectrum.

The idea of sparsity in different dictionaries has been used previously. For example, Mukherjee & Wang (2004) used wavelets to estimate the level of non-Gaussianity in the first year of the WMAP data. In Mukherjee & Wang (2003) wavelets were used to estimate the primordial power spectrum. Faÿ et al. (2008) have used wavelets to estimate the power spectrum from a CMB map, as an alternative to the MASTER method of Hivon et al. (2002). There have also been previous attempts to smooth the CMB power spectrum, for instance spline-fitting, where the smoothed power spectrum was used for visual aids (Oh et al. 1999). Also, Aghamousa et al. (2012) have smoothed WMAP 1-3-5-7 power spectra to investigate their evolution.

We used the sparsity of the CMB power spectrum as a key ingredient to estimate the theoretical power spectrum without having to know the cosmological parameters; this estimate will not belong to a set of possible theoretical power spectra (i.e., all $C[\ell]$ that can be obtained by CAMB³ by varying the cosmological parameters). Instead, this estimation should be useful for other applications, such as:

- Monte Carlo: we may need to conduct Monte Carlo simulations in some applications without assuming the cosmological parameters.
- Wiener filtering: Wiener filtering is often used to filter the CMB map and requires the theoretical power

spectrum as an input. We may not wish to assume any cosmology at this stage of the processing.

- Some estimators (weak lensing, integrated SachsWolfe (ISW), etc.) require the theoretical power spectrum to be known. Using a data-based estimate of the theoretical $C[\ell]$ could be an interesting alternative, or at least a good first guess in an iterative scheme where the theoretical $C[\ell]$ is required for determining the cosmological parameters.

Paper content

Section 2 introduces the concepts of sparse representation and its applications to the CMB power spectrum. In Section 3 we explain how sparsity is used to propose an estimator of the theoretical power spectrum. Experimental results are described and discussed in Section 4, and Section 5 is devoted to a discussion and comparison of the moving average estimator. In Section 6, conclusions are drawn and potential perspectives are stated.

2. Sparsity of the CMB power spectrum

2.1. A brief tour of sparsity

A signal $X = (X[1], \dots, X[N])$ considered as a vector in \mathbb{R}^N , is said to be sparse if most of its entries are equal to zero. If k number of the N samples are not equal to zero, where $k \ll N$, then the signal is said to be k -sparse. If only a few of the entries have high values and the rest are zero or close to zero, the signal is said to be weakly sparse (or compressible). With a slight abuse of terminology we call compressible signals sparse. Generally, signals are not sparse in direct space, but can be sparsified by transforming them to another domain. For example, $\sin(x)$ is 1-sparse in the Fourier domain, while it is clearly not sparse in the original one. In the so-called sparsity synthesis model, a signal can be represented as the linear expansion

$$X = \Phi\alpha = \sum_{i=1}^T \phi_i \alpha[i] \quad , \quad (7)$$

where $\alpha[i]$ are the synthesis coefficients of X , $\Phi = (\phi_1, \dots, \phi_T)$ is the dictionary, and ϕ_i are called the atoms (elementary waveforms) of the dictionary Φ . In the language of linear algebra, the dictionary Φ is a $N \times T$ matrix whose columns are the normalized atoms, supposed here to be normalized to a unit ℓ_2 -norm, i.e. $\forall i \in [1, T], \|\phi_i\|_2^2 = \sum_{n=1}^N |\phi_i[n]|^2 = 1$ ⁴. A function can be decomposed in many dictionaries, but the best dictionary is that with the sparsest (most economical) representation of the signal. In practice, it is convenient to use dictionaries with fast implicit transform (such as Fourier transform, wavelet transform, etc.) which allow us to directly obtain the coefficients and reconstruct the signal from these coefficients using fast algorithms that run in linear or almost linear time (unlike matrix-vector multiplications). The Fourier, wavelet, and discrete cosine transforms are certainly the most well known dictionaries. A comprehensive account of sparsity and its applications can be found in the monograph Starck et al. (2010).

⁴ The l_p -norm of a vector X , $p \geq 1$, is defined as $\|X\|_p = (\sum_i |X[i]|^p)^{1/p}$, with the usual adaptation $\|X\|_\infty = \max_i X[i]$.

² <http://www.esa.int/SPECIALS/Planck/index.html>

³ CAMB solves the Boltzmann equations for a cosmological model set out by the given cosmological parameters.

2.2. Which dictionary for the theoretical CMB power spectrum?

We investigated the sparsity of the CMB power spectrum in two different dictionaries, both having a fast implicit transform: the wavelet transform (WT) and the discrete cosine transform (DCT).

Figure 1 shows an angular power spectrum (calculated with CAMB (Lewis et al. 2000) with WMAP7 (Larson et al. 2010) parameters) along with the DCT- and WT-reconstructed power spectra with a varying fraction of the largest transform coefficients retained in the reconstruction. The inner plots show the difference between the actual power spectrum and the reconstructed ones. Clearly, with only few coefficients the shape of the power spectrum is correctly reconstructed in both dictionaries. The height and the position of the peaks and troughs are very important here because the estimation of the cosmological parameters heavily relies on these characteristics of the power spectrum. The best domain would be the one with the sparsest representation and yet the most accurate representation of the power spectrum. Let $C[\ell]^{(M)}$ be its best M -term approximation, i.e., obtained by reconstructing from the M -largest (in magnitude) coefficients of $C[\ell]$ in a given domain. To compare the WT and DCT dictionaries, we plot the resulting non-linear approximation (NLA) error curve in Figure 2, which shows the reconstruction error E_M as a function of M , the number of retained coefficients

$$E_M = \frac{\|C[\ell] - C[\ell]^{(M)}\|_2}{\|C[\ell]\|_2} \times 100. \quad (8)$$

As M increases we approach the complete reconstruction, and the error reaches zero when all coefficients have been used. Usually the domain with the steepest E_M curve is the sparsest domain. In this case, though, both dictionaries show a very similar behavior. There is only a small window in the coefficients for which DCT performs better than WT. However, DCT flattens after using $\sim 1\%$ of the coefficients and does not improve the reconstruction until many coefficients have been used.

Both dictionaries seem to suffer from boundary problems at low and high ℓ s. This can be solved for high ℓ s because one can always perform the reconstruction beyond the desired ℓ . For low ℓ s it can be solved by different means, for instance by extrapolating the spectrum. Note that the boundary problems are more severe in the DCT domain than in the WT; this is because DCT atoms are not compactly supported.

Next we investigated the sparsity of a set of *realized* spectra in the two dictionaries. We simulated 100 maps from the theoretical power spectrum used above and estimated their power spectra using Equation 4. As before, we decomposed each realization in the DCT and WT dictionaries and reconstructed keeping increasing fractions of the highest coefficients. At this stage, it is important to note that because we are dealing with the empirical power spectrum, we are no longer in an approximation setting but instead in an estimation one. Indeed, the empirical power spectrum can be considered a noisy version of the true one. Intuitively, reconstructing from a very small fraction of high coefficients will reject most of the noise (low estimator variance) but at the price of retaining only a small fraction of the true spectrum coefficients (strong bias). The opposite

is true when a large proportion of coefficients is kept in the reconstruction. Therefore, there is a threshold value that will entail a bias-variance trade-off, hence minimizing the estimation risk. This is exactly the idea underlying thresholding estimators in sparsifying domains.

This discussion is clearly illustrated by the inner plots of Figure 3, which shows the normalized mean-square error (NMSE) defined as

$$\text{NMSE}_M = \frac{\|C[\ell] - \hat{C}[\ell]^{(M)}\|_2^2}{\|C[\ell]\|_2^2} \times 100, \quad (9)$$

as a function of the fraction of coefficients used in the reconstruction. The error is large when only a few coefficients are used. As more coefficients are included, one starts to recover the main (i.e., the general shape of the spectrum) features of the power spectrum. With more coefficients, more noise enters the estimation and the error increases again. The NMSE curve shows a clear minimum at which the underlying true power spectrum is best recovered.

Despite the differences in the performance of the two dictionaries, the minima of the NMSE have about the same proportions of the coefficients. This is because the NMSE reflects a *global* behavior. On the one hand, although the DCT can recover the features of the spectrum correctly, it is less smooth than WT. On the other hand, the WT cannot reconstruct the proper shape of the power spectrum, but provides a smoother estimate.

Figures 4 and 5 show the same results for an average over the 100 realizations. Evidently, on average DCT performs very well in recovering the features of the CMB spectrum. Indeed, the minimum of NMSE curve is at a lower proportion of coefficients for DCT than WT. This is because the small *noisy* features of the DCT-reconstructed spectra cancel out in the averaging, while the wrong recovery of the shape of the spectrum by WT does not. In a nutshell, DCT seems to perform better in reconstructing the true underlying CMB power spectrum from its realizations.

To summarize, from the above discussion, we conclude the following:

- the CMB power spectrum is very sparse in both DCT and WT dictionaries, although their sparsifying capabilities are different;
- DCT recovers global features of spectrum (i.e., the peaks and troughs), while WT recovers localized features;
- WT recovers more localized (noisy) features than the global ones for realizations, while the DCT concentrates on the global features.

In the next section, these complementary capabilities of the DCT and WT transforms will be combined to propose a versatile way for adaptively estimating the theoretical power spectrum from a single realization of it.

3. Sparse reconstruction of the theoretical power spectrum

We begin with the simple model where the observed signal Y is contaminated by a zero-mean white Gaussian noise, $Y = X + \varepsilon$, where X is the signal of interest and

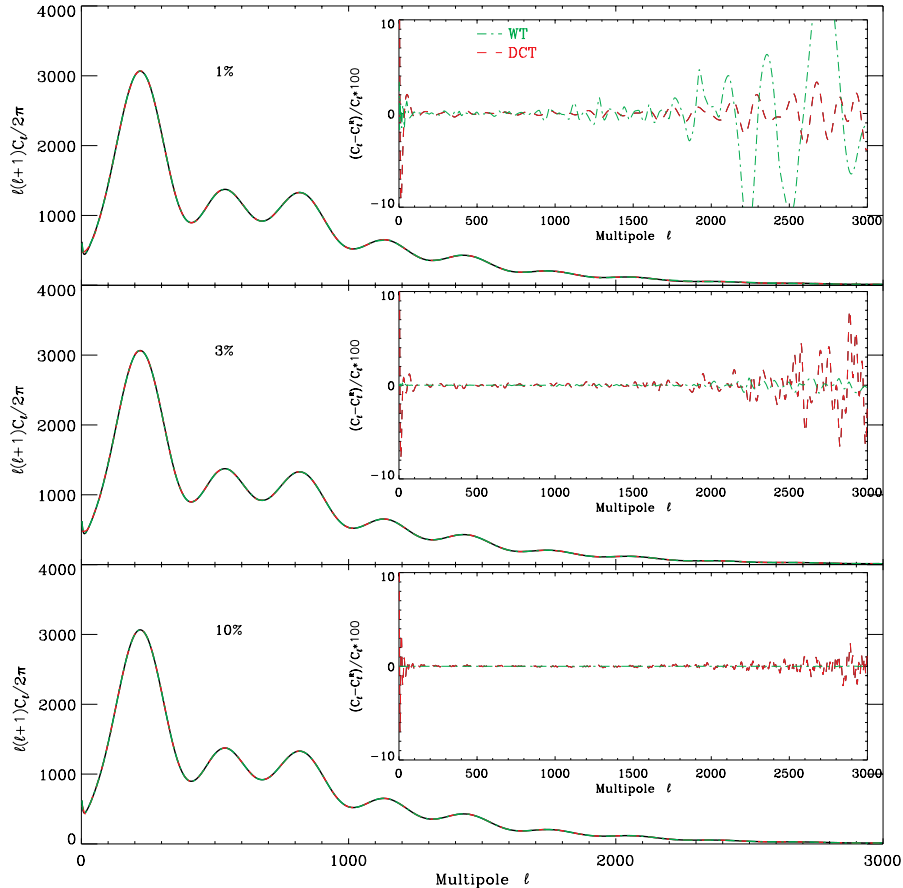


Fig. 1. Theoretical CMB power spectrum along with the reconstructed power spectra, using the DCT and WT dictionaries. The panels show the reconstructions for different fractions of the coefficients. The inner plots show the differences between the actual and the reconstructed power spectra. Both dictionaries suffer from boundary effects, but this is more severe for DCT because the corresponding atoms are not compactly supported. The power spectrum that is decomposed onto the two dictionaries is in the form $\ell(\ell + 1)C[\ell]/2\pi$.

$\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Sparse recovery with an analysis-type sparsity prior amounts to finding the solution of the following problem:

$$\min_X \|\Phi^T X\|_1 \quad s.t. \quad \|Y - X\|_2 \leq \delta, \quad (10)$$

where $\Phi^T X$ represents the transform coefficients of X in the dictionary Φ , and δ controls the fidelity to the data and obviously depends on the noise standard deviation σ .

We now turn to denoising the power spectrum from one empirical realization of it. In this case, however, the noise is highly non-Gaussian and needs to be treated differently. Indeed, as we will see in the next section, the empirical power spectrum will entail a multiplicative χ^2 -distributed noise with a number of degrees of freedom that depends on ℓ . That is, the noise has a variance profile that depends both on the true spectrum and ℓ . We therefore need to *stabilize* the noise on the empirical power spectrum prior to estimation, using a variance stabilization transform (VST). Hopefully, the latter will yield stabilized samples that have (asymptotically) constant variance, say 1, irrespective of the value of the input noise level.

3.1. Variance stabilizing transform

In the statistical literature the problem of removing the noise from an empirical power spectrum is called periodogram denoising (Donoho 1993). Komm et al. (1999) approximated the noise with a correlated Gaussian noise model and derived a threshold at each wavelet scale using the MAD (median of absolute deviation) estimator. A more elegant approach was proposed in Donoho (1993) and Moulin (1994), where the so-called Wahba VST was used. This VST is defined as

$$\mathcal{T}(X) = (\log X + \gamma) \frac{\sqrt{6}}{\pi}, \quad (11)$$

where $\gamma = 0.57721\dots$ is the Euler-Mascheroni constant. After the VST, the stabilized samples can be treated as if the noise contaminating them were white Gaussian noise with unit variance.

We will take a similar path here, generalizing the above approach to the case of the angular power spectrum. Indeed, from Equation 4, one can show that under mild

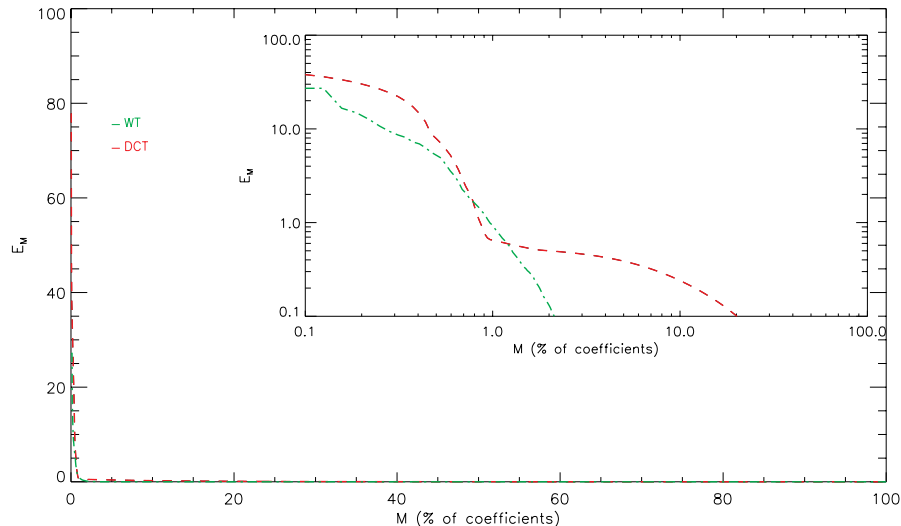


Fig. 2. Non-linear approximation (NLA) error curves for the two dictionaries. Below 1% the DCT curve is dropping faster, which means it is performing better. However, past $\sim 2\%$ the DCT curve flattens off, while WT decreases to ~ 0 very quickly.

regularity assumptions on the true power spectrum,

$$\widehat{C}[\ell] \xrightarrow{d} C[\ell]Z[\ell], \text{ where } \forall \ell \geq 2, 2LZ[\ell] \sim \chi_{2L}^2, L = 2\ell + 1, \quad (12)$$

where \xrightarrow{d} means convergence in distribution. From Equation 12, it is appealing then to take the logarithm to transform the multiplicative noise Z into an additive one. The resulting log-stabilized empirical power spectrum reads

$$C^s[\ell] := \mathcal{T}_\ell(\widehat{C}[\ell]) = \log \widehat{C}[\ell] - \mu_L = \log C[\ell] + \eta[\ell], \quad (13)$$

where $\eta[\ell] := \log Z[\ell] - \mu_L$, $L = 2\ell + 1$. Using the asymptotic results from Bartlett & Kendall (1946) on the moments of $\log -\chi^2$ variables, it can be shown that $\mu_L = \psi_0(L) - \log L$, $\mathbb{E}(\eta[\ell]) = 0$ and $\sigma_L^2 = \text{Var}[\eta[\ell]] = \psi_1(L)$, where $\psi_m(t)$ is the standard polygamma function, $\psi_m(t) = \frac{d^{m+1}}{dt^{m+1}} \log \Gamma(t)$.

We can now consider the stabilized $C^s[\ell]$ as noisy versions of the $\log C[\ell]$, where the noise is zero-mean additive and independent. Owing to the central limit theorem, the noise tends to Gaussian with variance σ_L^2 as ℓ increases. At low ℓ , normality is only an approximation. Indeed, the noise $\eta[\ell]$ has a probability density function of the form

$$p_\eta(\ell) = \frac{(2L)^L}{2^L \Gamma(L)} \exp[L(\ell + \mu_L - e^{\ell + \mu_L})], \quad (14)$$

which might be used to estimate the thresholds in the wavelet domain.

To standardize the noise, we slightly modify the VST (13) to the normalized form

$$C^s[\ell] := \mathcal{T}_\ell(\widehat{C}[\ell]) = \frac{\log \widehat{C}[\ell] - \mu_L}{\sigma_L} = X^s[\ell] + \varepsilon[\ell], \quad (15)$$

where now the noise $\varepsilon[\ell]$ is a zero-mean (asymptotically) Gaussian with unit variance, and $X^s[\ell] := \log C[\ell]/\sigma_L$. It can be checked that the Wahba VST (11) is a specialization of Equation 15 to $L = 0$.

In the following, we will use the operator notation $\mathcal{T}(X)$ for the VST that applies Equation 15 entry-wise to each $X[\ell]$, and $\mathcal{R}(X)$ its inverse operator, i.e. $\mathcal{R}(X) := (\mathcal{R}_\ell(X[\ell]))_\ell$ with $\mathcal{R}_\ell(X[\ell]) = \exp(\sigma_L X[\ell])$.

3.2. Signal detection in the wavelet domain

Without loss of generality, we restrict our description here to the wavelet transform. The same approach applies to other sparsifying transforms, e.g., DCT.

To estimate the true CMB power spectrum from the wavelet transform, it is important to detect the wavelet coefficients that are “significant”, i.e. those wavelet coefficients whose absolute value is too high to be due to noise

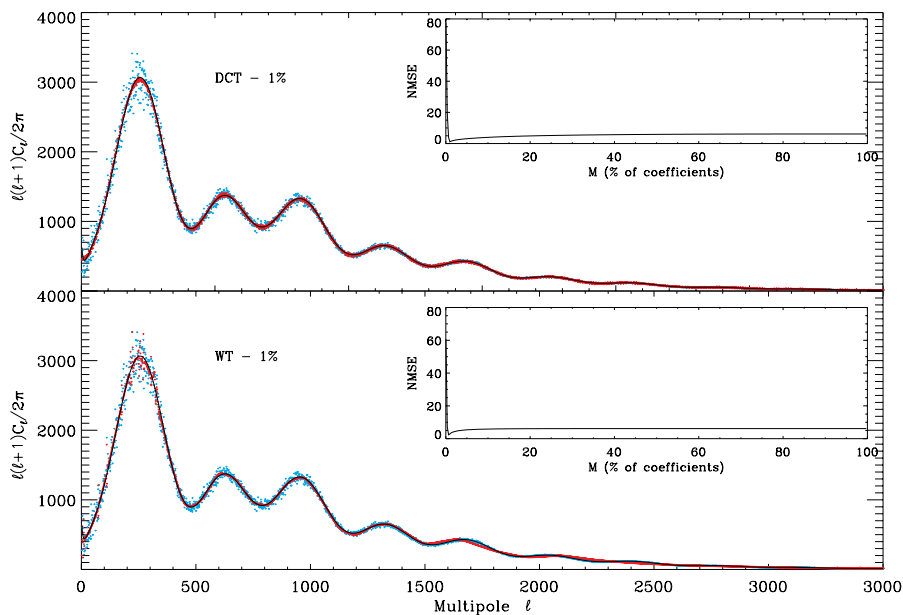


Fig. 3. A simulated CMB power spectrum along with the reconstructed spectra, using the DCT and WT dictionaries. The black solid line is the true underlying power spectrum from which the simulations were made. The blue and red dots show the simulated and the reconstructed power spectra respectively. With only 1% of the coefficients, DCT can recover the input power spectrum (i.e. the black solid line) very well, recovering the peaks and troughs accurately. Unlike DCT, WT seems to have difficulties in recovering the peaks and troughs. The inner plots show the NMSE curves.

(cosmic variance + instrumental noise). Let $w_j[\ell]$ be the wavelet coefficient of a signal Y at scale j and location ℓ . We define the multiresolution support M of Y as

$$M_j[\ell] = \begin{cases} 1 & \text{if } w_j[\ell] \text{ is significant,} \\ 0 & \text{if } w_j[\ell] \text{ otherwise.} \end{cases} \quad (16)$$

For Gaussian noise, it is easy to derive an estimate 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 (Starck & Murtagh 1998). To detect the significant wavelet coefficients, it suffices to compare the wavelet coefficients in magnitude $|w_j[\ell]|$ to a threshold level t_j . This threshold is generally taken to be equal to $\kappa\sigma_j$, where κ ranges from 3 to 5. This means that a small magnitude compared to the threshold implies that the coefficients are very likely to be due to noise and are hence insignificant. This decision corresponds to the hard-thresholding operator

$$\begin{aligned} \text{if } |w_j[\ell]| &\geq t_j && \text{then } w_j[\ell] \text{ is significant,} \\ \text{if } |w_j[\ell]| &< t_j && \text{then } w_j[\ell] \text{ is not significant.} \end{aligned} \quad (17)$$

To summarize, the multiresolution support is obtained from the signal Y by computing the forward transform co-

efficients, applying hard thresholding, and recording the coordinates of the retained coefficients.

3.3. Power spectrum recovery algorithm

We now turn to the adaptive estimator of the true CMB power spectrum $C[\ell]$ from its empirical estimate $\hat{C}[\ell]$. Because we benefit from the (asymptotic) normality of the noise in the stabilized samples $C^s[\ell]$ in Equation 15, we are in position to easily construct the multiresolution support M of C^s as described in the previous section. Once the support M of significant coefficients has been determined, our goal is to reconstruct an estimate \tilde{X} of the true power spectrum, known to be sparsely represented in some dictionary Φ (regularization), such that the significant transform coefficients of its stabilized version reproduce those of C^s (fidelity to data). Furthermore, because a power spectrum is a positive, a positivity constraint must be imposed. These requirements can be cast as seeking an estimate that solves

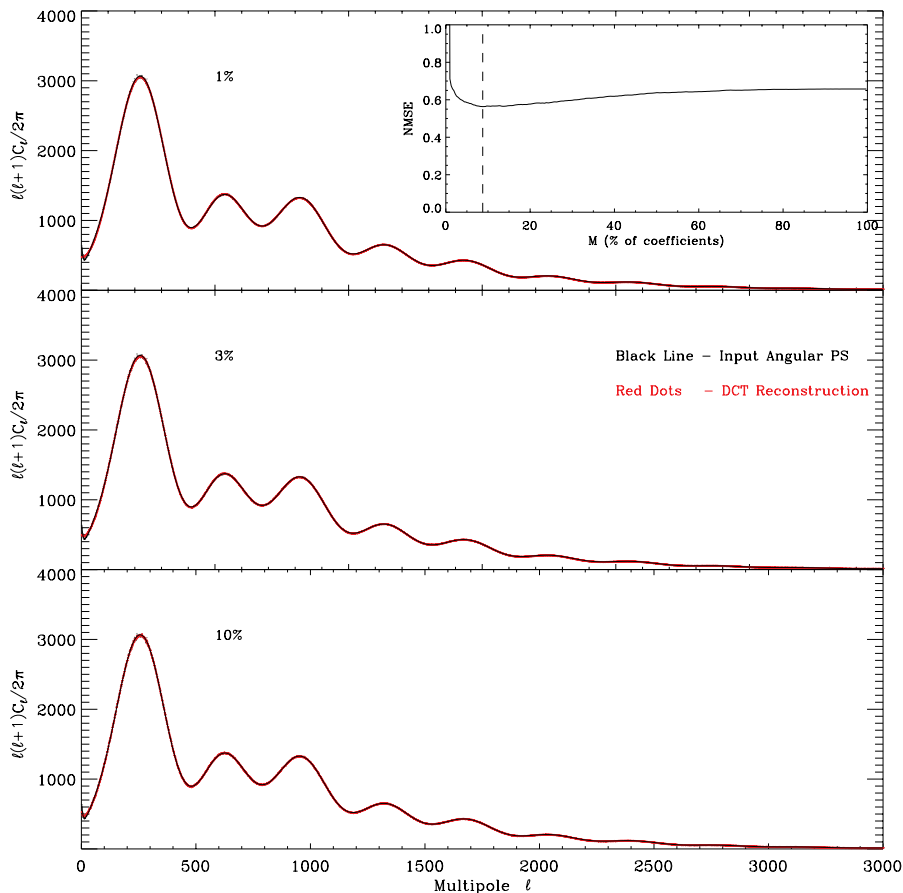


Fig. 4. Average of the DCT-reconstructed power spectra when different fractions of the most significant coefficients used in the reconstruction. The inner plot shows the NMSE curve for this average. The minimum of the curve is at less than 10% of the coefficients, meaning that the true CMB spectrum can be recovered with less than 10% of the coefficients while ensuring a good bias-variance trade-off.

the following constrained optimization problem:

$$\min_X \|\Phi^T X\|_1 \quad \text{s.t.} \quad \begin{cases} X \geq 0 \\ M \odot (\Phi^T \mathcal{T}(X)) = M \odot (\Phi^T C^s) \end{cases}, \quad (18)$$

where \odot stands for the Hadamard product (i.e., entry-wise multiplication) of two vectors. This problem has a global minimizer, which is bounded. However, in addition to non-smoothness of the l_1 -norm and the constraints, the problem is also non-convex because of the VST operator \mathcal{T} . It is therefore far from easy to solve.

We propose the following scheme, which starts with an initial guess of the power spectrum $X^{(0)} = 0$, and then iterates for $n = 0$ to $N_{\max} - 1$,

$$\begin{aligned} \tilde{X} &= \mathcal{R} \left(\mathcal{T} \left(X^{(n)} \right) + \Phi M \odot \left(\Phi^T \left(C^s - \mathcal{T} \left(X^{(n)} \right) \right) \right) \right) \\ X^{(n+1)} &= \mathcal{P}_+ \left(\Phi \text{ST}_{\lambda_n} \left(\Phi^T \tilde{X} \right) \right), \end{aligned} \quad (19)$$

where \mathcal{P}_+ denotes the projection onto the positive orthant and guarantees non-negativity of the spectrum estimator, $\text{ST}_{\lambda_n}(w) = (\text{ST}_{\lambda_n}(w[i]))_i$ is the soft-thresholding

with threshold λ_n that applies term-by-term the shrinkage rule

$$\text{ST}_{\lambda_n}(w[i]) = \begin{cases} \text{sign}(w[i])(|w[i]| - \lambda_n) & \text{if } |w[i]| \geq \lambda_n, \\ 0 & \text{otherwise.} \end{cases} \quad (20)$$

Here, we have chosen a decreasing threshold with the iteration number n , $\lambda_n = (N_{\max} - n)/(N_{\max} - 1)$. More details pertaining to this algorithm can be found in [Starck et al. \(2010\)](#)

Algorithm 1 summarizes the main steps of the sparse denoising algorithm. A similar approach was proposed in [Starck et al. \(2009\)](#) and [Schmitt et al. \(2010\)](#) for Poisson noise removal in 2D and 3D data sets.

3.4. Instrumental noise

In practice, the data are generally contaminated by an instrumental noise, and estimating the true CMB power spectrum $C[\ell]$ from the empirical power spectrum $\hat{C}[\ell]$ requires one to remove this instrumental noise. The instrumental noise is assumed to be stationary and independent of the CMB. We also assumed that we have access to the power spectrum of the noise, or we can compute the empirical

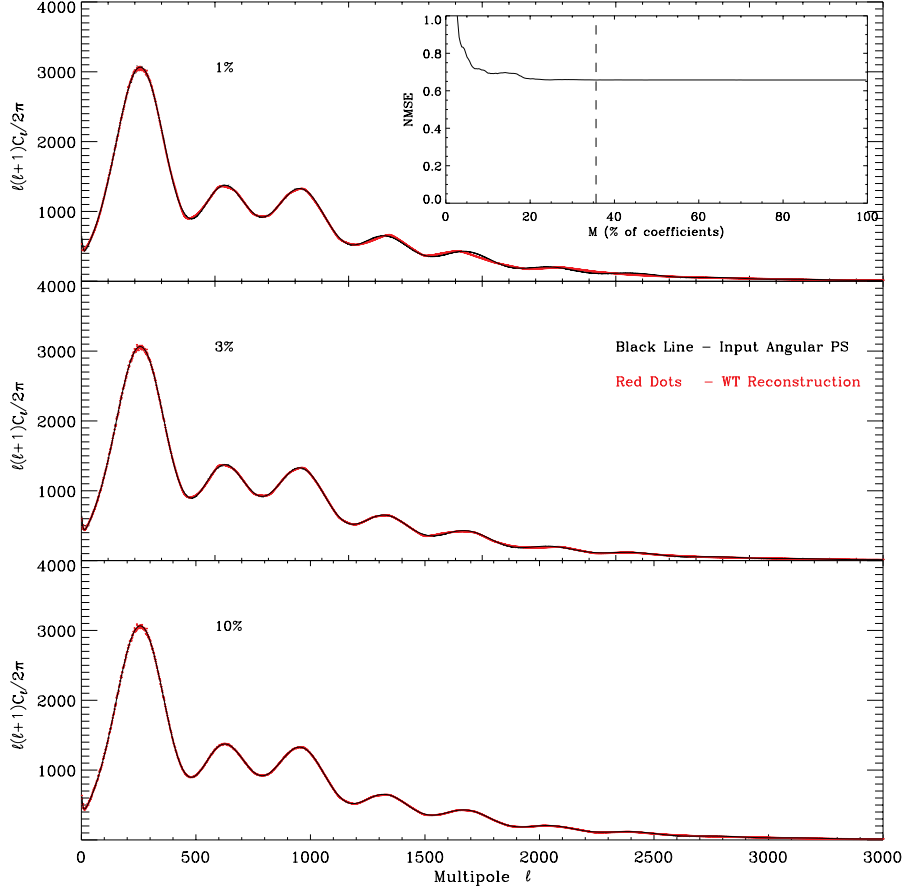


Fig. 5. Same as Figure 4, but for the WT domain. It seems that WT cannot perform well for simulated power spectra, compared to the DCT domain; the minimum of the NMSE curve is at more than 35% — while the curve is quite flat after $\sim 20\%$.

Algorithm 1: TOUSI power spectrum smoothing

Require:

Empirical power spectrum \widehat{C} ,
 number of iterations N_{\max} ,
 threshold κ (default value is 5).

Detection

- 1: Compute C^s using Equation 15.
- 2: Compute the decomposition coefficients W of C^s in Φ , $W = \Phi^T C^s$.
- 3: Compute the support M from W with the threshold κ , assuming standard additive white Gaussian noise.

Estimation

- 4: Initialize $X^{(0)} = 0$.
 - 5: **for** $n = 0$ to $N_{\max} - 1$ **do**
 - 6: $\tilde{X} = \mathcal{R} \left(\mathcal{T} \left(X^{(n)} \right) + \Phi M \odot \left(\Phi^T \left(C^s - \mathcal{T} \left(X^{(n)} \right) \right) \right) \right)$.
 - 7: $X^{(n+1)} = \mathcal{P}_+ \left(\Phi \text{ST}_{\lambda_n} \left(\Phi^T \tilde{X} \right) \right)$.
 - 8: $\lambda_{n+1} = \frac{N_{\max} - (n+1)}{N_{\max} - 1}$.
 - 9: **end for**
 - 10: **Return:** The estimate $\tilde{X} = X^{(N_{\max})}$.
-

power spectrum $\widehat{S}_N[\ell]$ of at least one realization, either from a JackKnife data map or from realistic instrumental noise simulations. The above algorithm can be adapted to handle

this case after rewriting the optimizing problem as follows:

$$\min_X \|\Phi^T X\|_1 \quad \text{s.t.} \quad \begin{cases} X \geq 0 \\ M \odot (\Phi^T \mathcal{T}(X + \widehat{S}_N)) = M \odot (\Phi^T C^s) \end{cases} \quad (21)$$

Thus, Equation 19 becomes

$$\tilde{X} = \mathcal{R} \left(\mathcal{T}(X^{(n)} + \widehat{S}_N) + \Phi M \odot \left(\Phi^T (C^s - \mathcal{T}(X^{(n)} + \widehat{S}_N)) \right) \right) - \widehat{S}_N$$

$$X^{(n+1)} = \mathcal{P}_+ (\Phi \text{ST}_{\lambda_n} (\Phi^T \tilde{X})) \quad (22)$$

Algorithm 1 can be modified accordingly.

3.5. Combining several dictionaries

We have seen in Section 2.2 that the WT and DCT dictionaries had complementary benefits. Indeed, each dictionary is able to capture the features with shapes similar to its atoms well. More generally, assuming that we have D dictionaries Φ_1, \dots, Φ_D . Given a candidate signal Y , we

can derive a support M_d associated to each dictionary Φ_d , for $d \in \{1, \dots, D\}$. The optimization problem to solve now reads

$$\begin{aligned} & \min_X \|\Phi^T X\|_1 \quad \text{s.t.} \\ & \begin{cases} X \geq 0 \\ M_d \odot (\Phi_d^T \mathcal{T}(X + \hat{S}_N)) = M_d \odot (\Phi_d^T C^s), \quad d \in \{1, \dots, D\} \end{cases} \end{aligned} \quad (23)$$

Again, this is a challenging optimization problem. We propose to solve it by applying Equation 22 successively and alternatively on each dictionary Φ_d . Algorithm 2 describes in detail the different steps.

Algorithm 2: TOUSI power spectrum smoothing with D dictionaries

Require:

Empirical power spectrum \hat{C} , D dictionaries Φ_1, \dots, Φ_D ,
 noise power spectrum \hat{S}_N ,
 number of iterations N_{\max} ,
 threshold κ (default value is 5).

Detection

- 1: Compute C^s using Equation 15.
- 2: For all d , compute the decomposition coefficients W_d of C^s in Φ_d , $W_d = \Phi_d^T C^s$.
- 3: For all d , compute the support M_d from W_d with the threshold κ , assuming standard additive white Gaussian noise.

Estimation

- 4: Initialize $X^{(0)} = 0$,
 - 5: **for** $n = 0$ to $N_{\max} - 1$ **do**
 - 6: $Z_d = X^{(n)}$.
 - 7: **for** $d = 1$ to D **do**
 - 8: $T = \mathcal{T}(Z_d + \hat{S}_N)$.
 - 9: $\tilde{Z} = \mathcal{R}\left(T + \Phi_d M \odot \left(\Phi_d^T (C^s - T)\right)\right) - \hat{S}_N$.
 - 10: $Z_{d+1} = \mathcal{P}_+\left(\Phi_d \text{ST}_{\lambda_n}(\Phi_d^T \tilde{Z})\right)$.
 - 11: **end for**
 - 12: $X^{(n+1)} = Z^{D+1}$.
 - 13: $\lambda_{n+1} = \frac{N_{\max} - (n+1)}{N_{\max} - 1}$.
 - 14: **end for**
 - 15: Get the estimate $\tilde{X} = X^{(N_{\max})}$.
-

4. Application: Monte Carlo simulations

We simulated 100 maps from a theoretical CMB power spectrum that was calculated by CAMB. The power spectra of these maps are equivalent to 100 realizations of the true CMB power spectrum — this realized spectrum is what we have access to in reality. Each of these 100 simulated spectra were run through the TOUSI algorithm, with the aim of recovering the theoretical spectrum from which these 100 spectra were simulated (i.e., the one that was calculated by CAMB).

Figure 6 shows an example of this; the empirical power spectrum of one realization (blue dots) that was fed into the TOUSI algorithm, the average of the 100 estimated power spectra using TOUSI (red line) and the input theoretical spectra (black line). The black line is the input theoretical spectrum that was calculated by CAMB, which is what we are trying to recover. The reconstruction of the peaks

and troughs of the power spectrum by this algorithm is very impressive. This is very important because these features define the cosmological parameters. To additionally check the accuracy of these reconstructed spectra, we estimated a set of cosmological parameters from these spectra, using CosmoMC (Lewis & Bridle 2002). First, a set of cosmological parameters were estimated from the 100 simulated spectra. The results are shown in Figure 4 as black solid lines. Then a ‘mean’ reconstructed spectrum was calculated by averaging the 100 reconstructed spectra. This, in principle (i.e., if the algorithm has worked), should be an estimation of the true input spectrum with the same characteristics and the same cosmological parameters. To test this, we used this ‘mean’ reconstructed spectrum to simulate another 100 maps and then 100 spectra. These simulated spectra were run through CosmoMC to estimate the same set of cosmological parameters. These are shown as red lines in Figure 4. Clearly, the TOUSI algorithm can reconstruct the true underlying power spectrum with great accuracy in the cosmological parameters.

4.1. Data with instrumental noise

Here we present the performance of the TOUSI algorithm in the presence of instrumental noise. The noise maps were simulated using a theoretical (Planck level) noise power spectrum. They were added to the CMB maps simulated above and the power spectra of the combined maps were estimated using Equation 4.

Figure 8 shows the reconstruction of the theoretical CMB spectrum in the presence of noise. The blue dots show the empirical power spectrum of one realization with instrumental noise. Yellow dots show the estimated power spectrum of one of the simulated noise maps. Green dots show the the spectrum with the noise power spectrum removed. The black and red solid lines are the input and reconstructed power spectra, respectively. The theoretical power spectrum can be reconstructed up to the point where the structure of the power spectrum has not been destroyed by the instrumental noise. In our case, having Planck level noise, this goes to ℓ up to 2500. Obviously, the TOUSI performs very well in reconstructing the input power spectrum even in the presence of instrumental noise.

4.2. Test on WMAP7 power spectrum

We tested our algorithm on real data. Figure 9 shows the application of our technique to the WMAP7 power spectrum. The method indeed works well up to ℓ of ~ 800 . After this point the instrumental noise becomes so dominant that the features of the spectrum are washed out and cannot be recovered.

5. Sparsity versus averaging

A very common approach to reduce the noise on the power spectrum is the moving average filter, i.e., average values in a given window,

$$\tilde{C}^A[b] = \frac{1}{b(b+1)\omega_b} \sum_{\ell=b-\frac{\omega_b}{2}}^{b+\frac{\omega_b}{2}} \ell(\ell+1)\hat{C}[\ell], \quad (24)$$

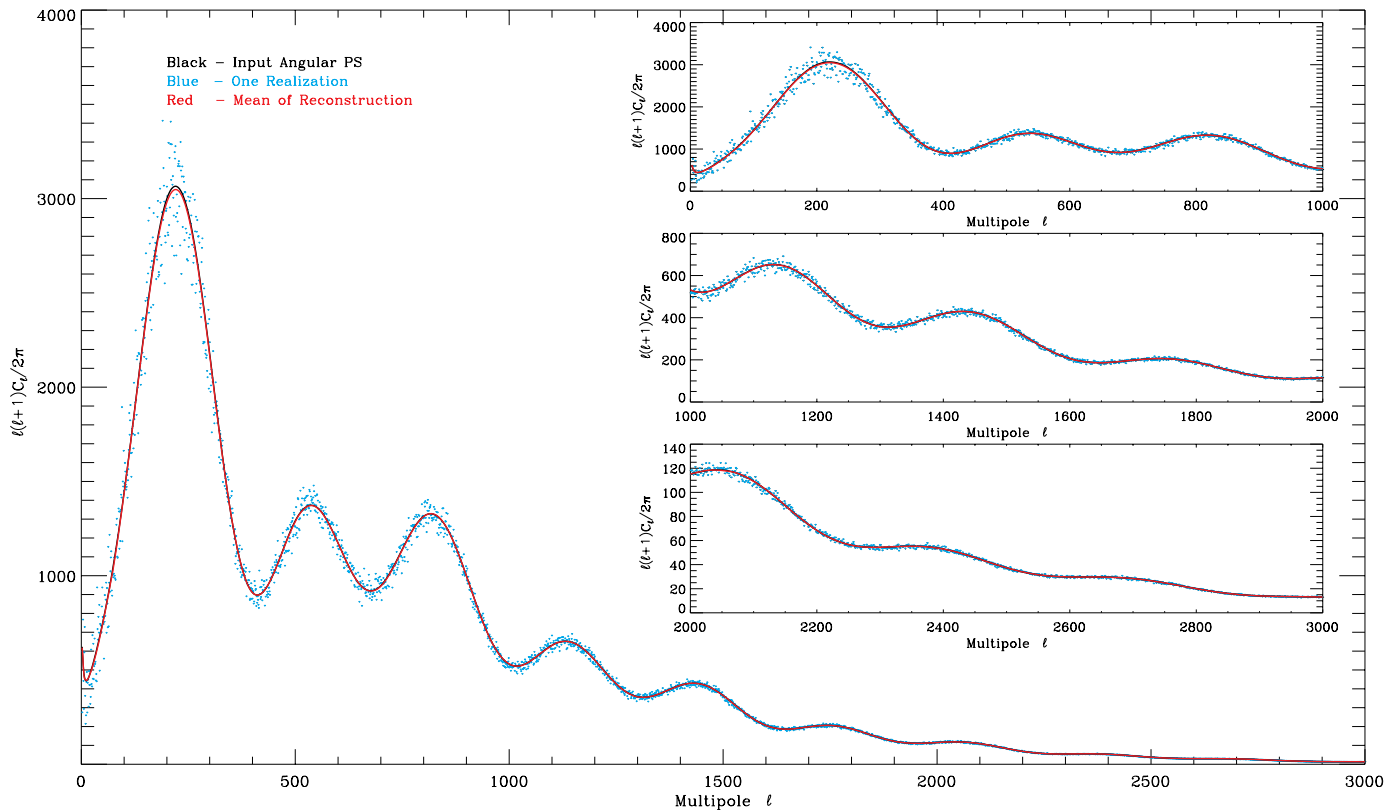


Fig. 6. Theoretical CMB power spectrum (black line), the empirical power spectrum of one realization (blue dots), and the averaged of estimated power spectra (red line) using the TOUSI algorithm. The inner plots show a zoomed-in version.

and the window size ω_b is increasing with ℓ . Here, we used window sizes of $\{1, 2, 5, 10, 20, 50, 100\}$, respectively, for ℓ ranging from $\{2, 11, 31, 151, 421, 1201, 2501\}$ to $\{10, 30, 150, 420, 1200, 2500, 3200\}$, which have also been used in the framework of the Planck project in Leach et al. (2008).

Figure 10 shows a reconstruction of the power spectrum for TOUSI versus averaging. From the NMSE curve it is clear that our algorithm is much more efficient. Figure 11 shows the average error the 100 realizations as a function of ℓ

$$E[\ell] = \frac{1}{100} \sum_{i=1}^{100} \|C[\ell] - \tilde{C}_i[\ell]\|_2, \quad (25)$$

where \tilde{C}_i is the estimated power spectrum from the i -th realization. We display the errors for the spectra estimated by the empirical estimator (the realization, black dotted line), the averaging estimator (red dashed line) and TOUSI (solid blue line). The cosmic variance is overplotted as a solid black line. Evidently, the expected error is highly reduced when using the sparsity-based estimator.

6. Conclusion

Measurements of the CMB anisotropies are powerful cosmological probes. In the currently favored cosmological

model, with the nearly Gaussian-distributed curvature perturbations, almost the entire statistical information is contained in the CMB angular power spectrum. We have investigated the sparsity of the CMB power spectrum in two dictionaries, the DCT and WT. In both dictionaries the CMB power spectrum can be recovered with only a few percentages of the coefficients, meaning the spectrum is very sparse. The two dictionaries have different characteristics and can accommodate reconstructing different features of the spectra. The DCT can help recover the global features of the spectrum, while WT helps recover small localized features. The sparsity of the CMB spectrum in these two domains has helped us develop an algorithm, TOUSI, which estimates the true underlying power spectrum from a given realized spectrum. This algorithm uses the sparsity of the CMB power spectrum in both WT and DCT domains and takes the best from both worlds to obtain a highly accurate estimate from a single realization of the CMB power spectrum. This could be a replacement for CAMB if knowing the cosmological parameters is not necessary. The developed IDL code will be released with the next version of ISAP (Interactive Sparse astronomical data Analysis Packages) via the web site

<http://jstarck.free.fr/isap.html>.

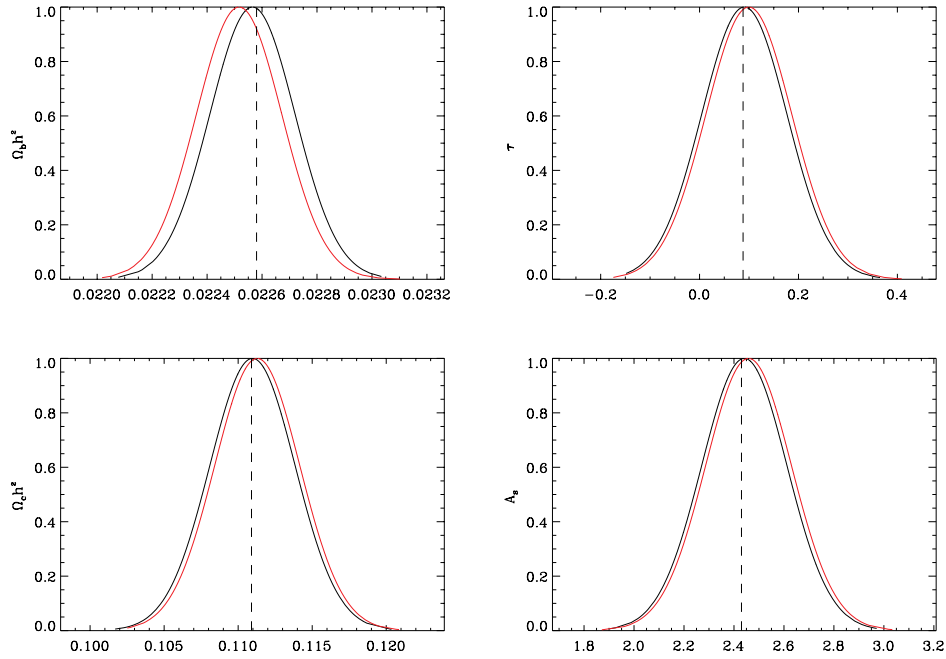


Fig. 7. Cosmological parameters estimated from the true CMB power spectrum (black line) and the mean of the reconstructed power spectra (red line). The dashed line is the true input parameters, i.e., those used to calculate the theoretical power spectrum using CAMB.

Acknowledgments

The authors thank Marian Douspis, Olivier Doré, and Amir Hajian for useful discussions. This work is supported by the European Research Council grant SparseAstro (ERC-228261)

References

- Aghamousa, A., Arjunwadkar, M., & Souradeep, T. 2012, *ApJ*, 745, 114
- Bartlett, M. S. & Kendall, D. G. 1946, *Journal of the Royal Statistical Society, Series B*, 8, 128
- Bennett, C. L., Hill, R. S., Hinshaw, G., et al. 2003, *APJS*, 148, 97
- Donoho, D. 1993, in *Proceedings of Symposia in Applied Mathematics*, ed. A. M. Society, Vol. 47, 173–205
- Faÿ, G., Guilloux, F., Betoule, M., et al. 2008, *Phys. Rev. D*, 78, 083013
- Hivon, E., Górski, K. M., Netterfield, C. B., et al. 2002, *APJ*, 567, 2–17
- Komm, R. W., Gu, Y., Hill, F., Stark, P. B., & Fodor, I. K. 1999, *ApJ*, 519, 407
- Larson, D., Dunkley, J., Hinshaw, G., et al. 2010, *ArXiv e-prints*
- Leach, S. M., Cardoso, J.-F., Baccigalupi, C., et al. 2008, *A&A*, 491, 597–615
- Lewis, A. & Bridle, S. 2002, *PRD*, 66, 103511
- Lewis, A., Challinor, A., & Lasenby, A. 2000, *Astrophys. J.*, 538, 473
- Moulin, P. 1994, *IEEE Transactions on Signal Processing*, 42, 3126–3136
- Mukherjee, P. & Wang, Y. 2003, *Astrophysical Journal*, 599, 1–6
- Mukherjee, P. & Wang, Y. 2004, *Astrophysical Journal*, 613, 51–60
- Oh, S. P., Spergel, D. N., & Hinshaw, G. 1999, *APJ*, 510, 551
- Readhead, A. C. S., Mason, B. S., Contaldi, C. R., et al. 2004, *APJ*, 609, 498
- Reichardt, C. L., Ade, P. A. R., Bock, J. J., et al. 2009, *APJ*, 694, 1200
- Schmitt, J., Starck, J. L., Casandjian, J. M., Fadili, J., & Grenier, I. 2010, *A&A*, 517, A26+
- Starck, J.-L., Fadili, J. M., Digel, S., Zhang, B., & Chiang, J. 2009, *A&A*, 504, 641
- Starck, J.-L. & Murtagh, F. 1998, *Publications of the Astronomical Society of the Pacific*, 110, 193–199
- Starck, J.-L., Murtagh, F., & Fadili, M. 2010, *Sparse Image and Signal Processing* (Cambridge University Press)

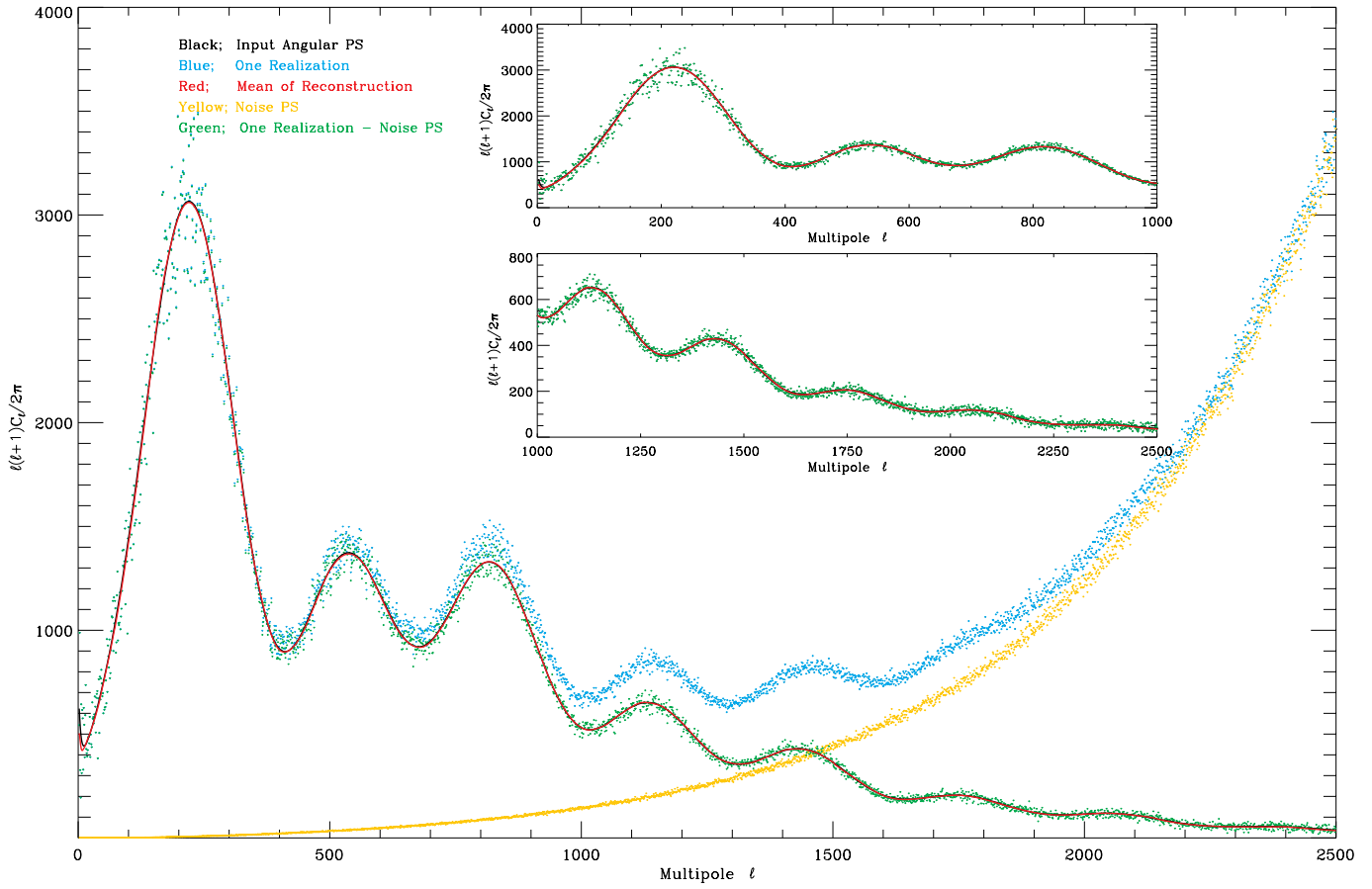


Fig. 8. Power spectrum estimate in the presence of instrumental noise. The blue dots show the empirical power spectrum of one realization with instrumental noise. Yellow dots show the estimated power spectrum of one of the simulated noise maps. Green dots show the the spectrum with the noise power spectrum removed. The black and red solid lines are the input and reconstructed power spectra, respectively. The inner plots show a zoomed-in version.

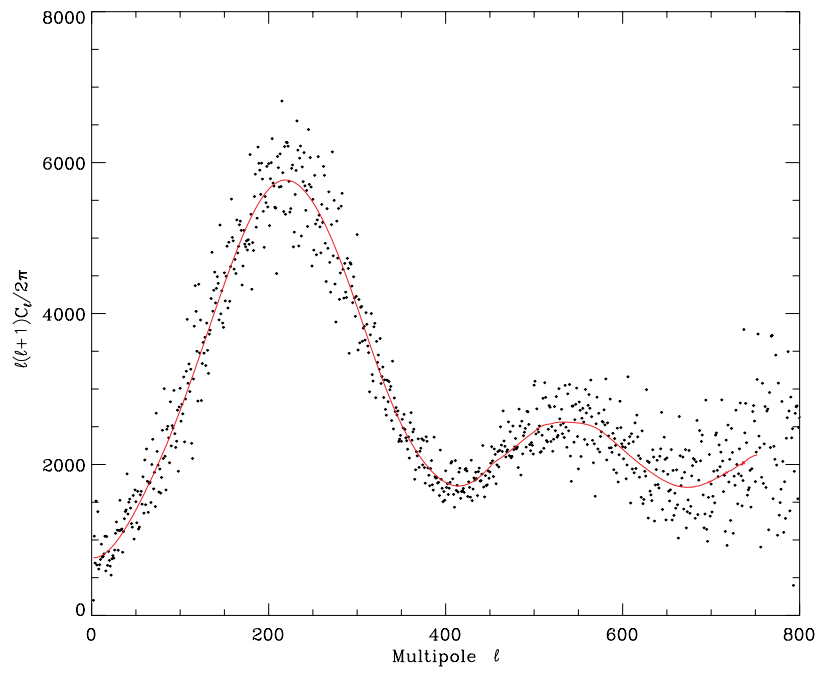


Fig. 9. TOUSI algorithm applied to the WMAP7 power spectrum. The technique works well up to ℓ of ~ 800 , i.e., before the instrumental noise becomes dominant washes the features of the spectrum out.

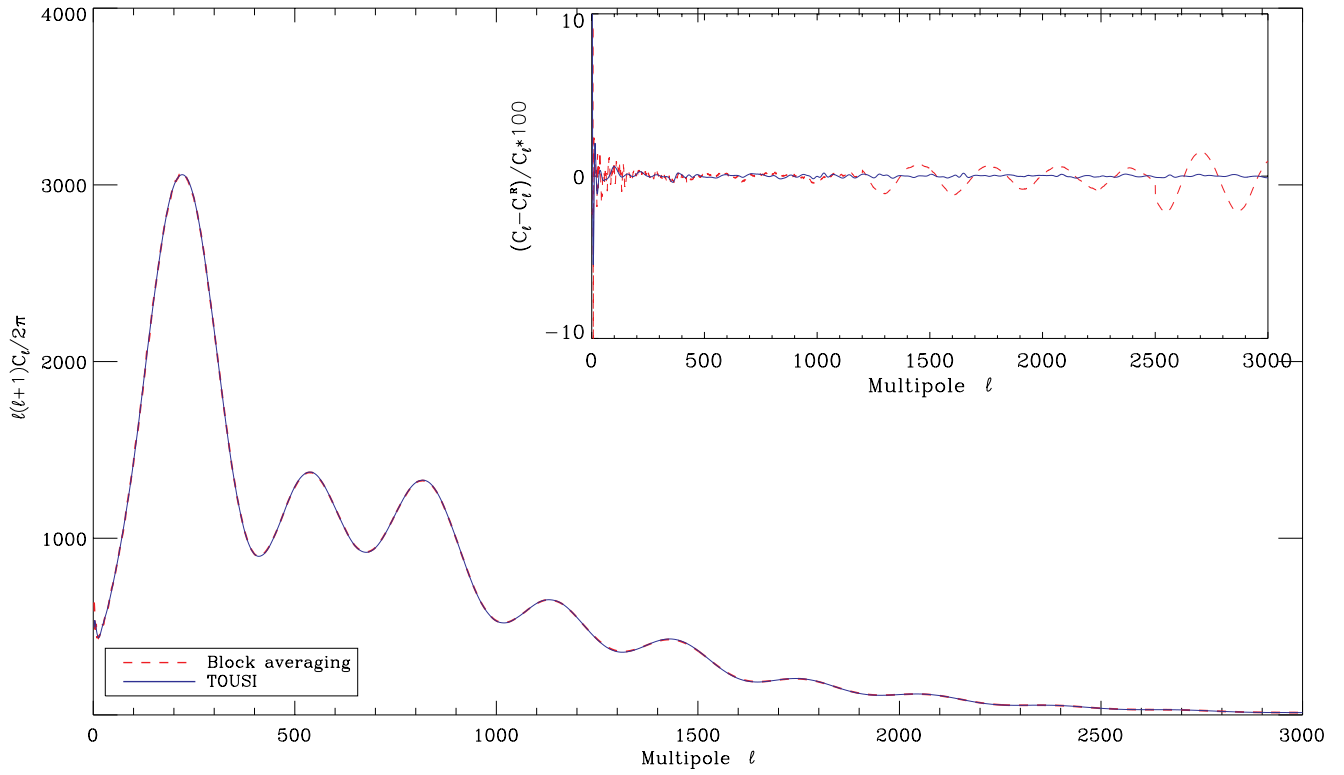


Fig. 10. Mean denoised spectra with wavelets (blue solid line) and averaging (red dashed line) from 100 realizations. The inner plot shows the normalized error for both dictionaries.

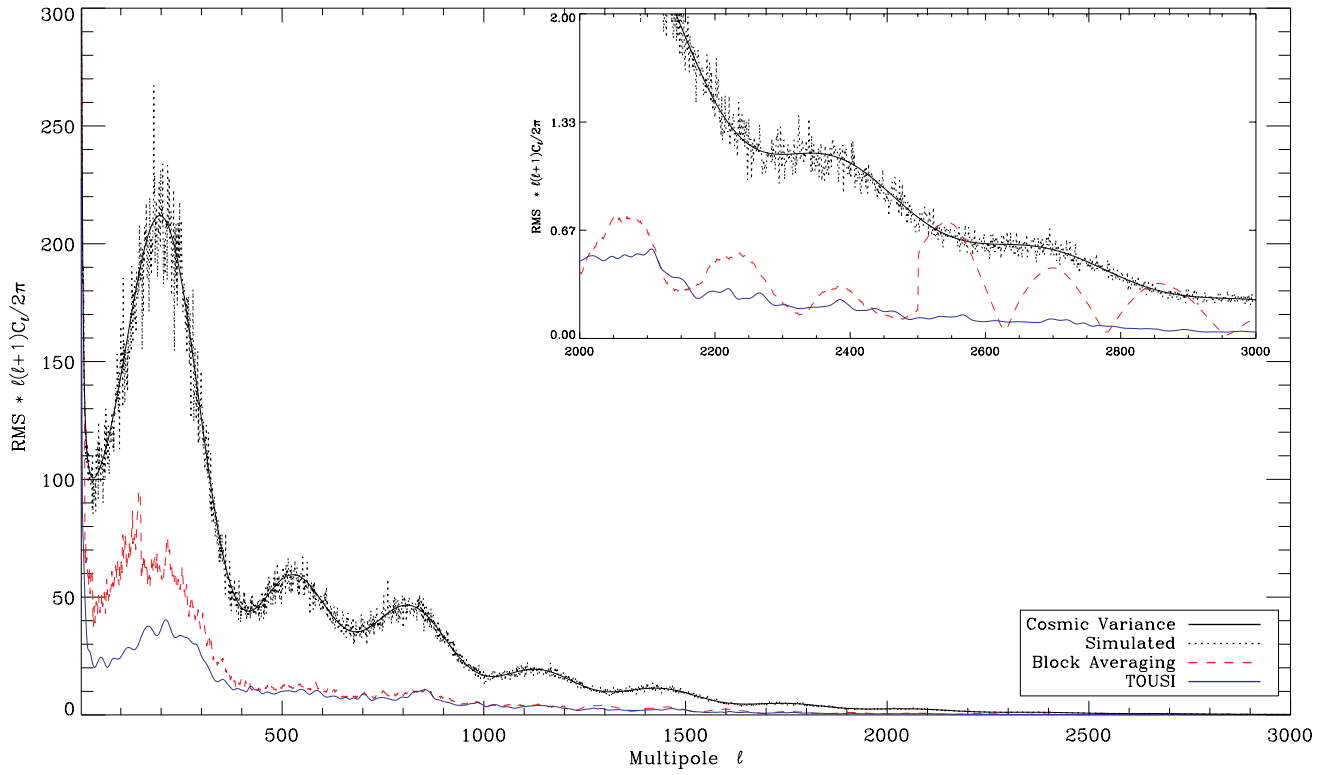


Fig. 11. Mean error for the 100 realizations, for the realizations (black dotted line), the averaging denoising (red dashed line) and the sparse wavelet filtering (blue solid line). The inner plot shows a zoom between $l = 2000$ and $l = 3000$.