# True cosmic microwave background power spectrum estimation

Paniez Paykari, Jean-Luc Starck, Jalal M. Fadili

# True CMB Power Spectrum Estimation

P. Paykari[1] ⋆, J.-L. Starck[1], and M. J. Fadili[2]

[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

June 29, 2012

**ABSTRACT**

*Context.*
*Aims.* The cosmic microwave background (CMB) power spectrum is a powerful cosmological probe as it entails almost all the statistical information of the 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. In this paper we aim to recover the true underlying CMB power spectrum from the one realization that we have without a need to know 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 only a few percentage of the coefficients in both of these dictionaries and hence is very compressible in these dictionaries.
*Results.* We study the performance of these dictionaries in smoothing a set of simulated power spectra. Based on this, we develop 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 to the true CMB simulations with the correct cosmological parameters. This allows us to make 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 CMB anisotropies are powerful cosmological probes. In the currently favored cosmological model, with the nearly Gaussian-distributed curvature perturbations, almost all the statistical information are 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

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

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

$\mathbb{S}^2 \subset \mathbb{R}^3$ is the unit sphere, $\ell$ is the multipole moment which is related to the angular size on the sky as $\ell \sim 180°/\theta$ and $m$ is the phase ranging from $-\ell$ to $\ell$. 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 all the statistical information of the field. In case where 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 $\ell \geqslant 2$, the triangular array $(a[\ell, m])_{\ell,m}$ represents zero-mean, complex-valued random coefficients, with variance

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

⋆ paniez.paykari@cea.fr

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

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

$$\widehat{C}[\ell] = \frac{1}{2\ell+1} \sum_m |a[\ell,m]|^2 \ . \tag{4}$$

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

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

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

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

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 Planck (Planck Collaboration et al. 2011). 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 due to 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. This estimate will not belong to a set of possible theoretical power spectra (i.e. all $C[\ell]$ that can be obtained by CAMB[2] by varying the cosmological parameters). Instead, such an estimation should be useful for other applications, such as:

- Monte Carlo: we may want to make Monte Carlo simulations in some applications without assuming the cosmological parameters.
- Wiener filtering: Wiener filtering is often used to filter the CMB map and it requires the theoretical power spectrum as an input. We may not want to assume any cosmology at this stage of the processing.
- Some estimators (weak lensing, ISW, etc.) require the theoretical power spectrum to be known. Using a data-based estimation 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 to determine the cosmological parameters.

## 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. In the case where only a few of the entries have large 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, in the sequel, we will 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

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

where $\alpha[i]$ are the synthesis coefficients of $X$, $\Phi = (\phi_1, \dots, \phi_T)$ is the dictionary, and $\phi_i$ are called the atoms (elementary waveforms) of the dictionary $\Phi$. In the language of linear algebra, the dictionary $\Phi$ is a $N \times T$ matrix whose columns are the atoms normalized, supposed here to be normalized to a unit $\ell_2$-norm, i.e. $\forall i \in [1,T], \|\phi_i\|_2^2 = \sum_{n=1}^{N} |\phi_i[n]|^2 = 1$[3]. A function can be decomposed in many dictionaries, but the best dictionary is the one 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 running in linear or almost linear time (unlike matrix-vector multiplications). The Fourier, wavelet and discrete cosine transforms provide certainly the most well known dictionaries. A comprehensive account on sparsity and its applications can be found in the monograph (Starck et al. 2010).

[2] CAMB solves the Boltzmann equations for a cosmological model set out by the given cosmological parameters.

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

### *2.2. Which Dictionary for the Theoretical CMB Power Spectrum?*

We investigate 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).

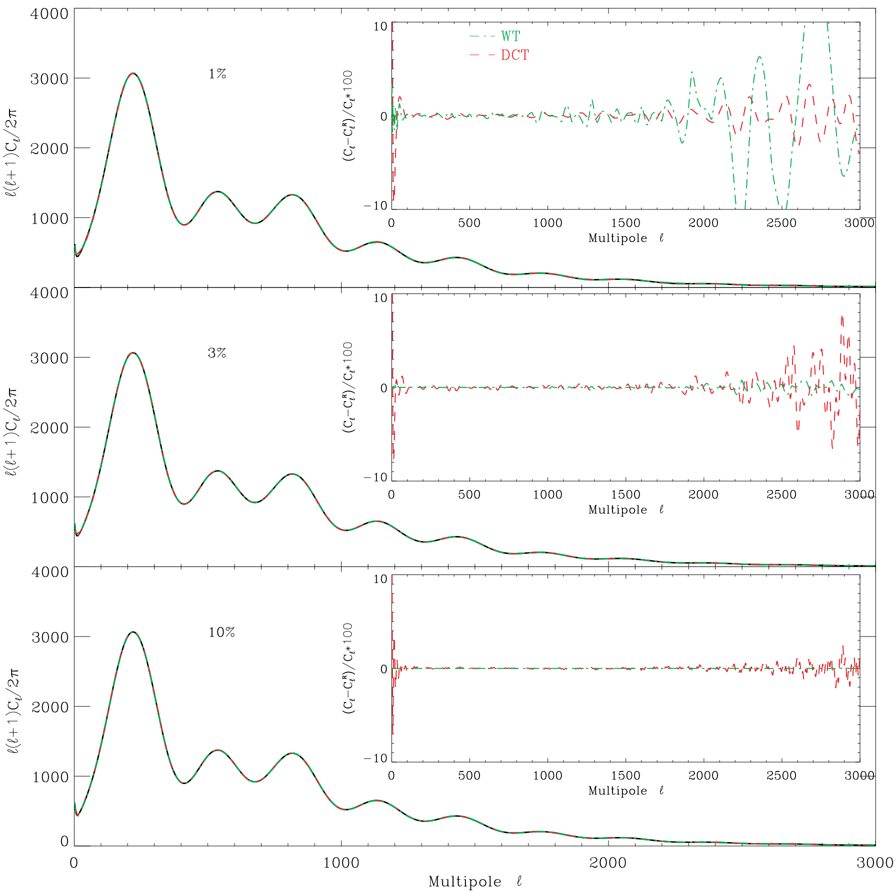


**Fig. 1.** A theoretical CMB power spectrum along with the reconstructed power spectra, using the DCT and WT dictionaries. The panels show the reconstructions for different fractions of the coefficients used. The inner plots show the differences between the actual and the reconstructed power spectra.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

Figure 1 shows an angular power spectrum (calculated by 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. It can be seen that with only a few percentage of the coefficients the shape of the power spectrum is correctly reconstructed in both dictionaries. The height and the position of the peaks and troughs are of great importance here as 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;

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

As $M$ increases we get closer to the complete reconstruction and the error reaches 0 when all the coefficients have been used. Usually the domain with the steepest $E_M$ curve is the sparsest domain. In this case though both dictionaries

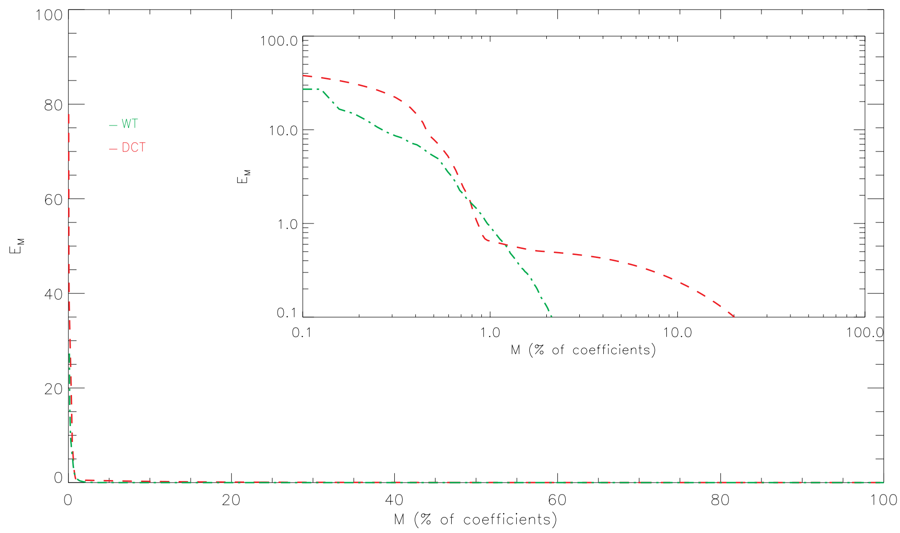


**Fig. 2.** Non-Linear Approximation (NLA) error curves for the two dictionaries. Below 1% the DCT curve is dropping faster, which means it is doing a better job. However, past $\sim 2\%$ the DCT curve flattens off while WT decreases to $\sim 0$ very quickly.

have very similar behaviors. There is only a small window in the coefficients for which DCT does a better job than WT. However, DCT flattens after using $\sim 1\%$ of the coefficients and does not improve the reconstruction until a big proportion of the coefficients have been used.

Both dictionaries seem to suffer from boundary issues at low and high $\ell$s. This can be solved for high $\ell$s as one can always perform the reconstruction beyond the desired $\ell$. For low $\ell$s it can be solved by different means, such as extrapolation of the spectrum. Note that the boundary issues are more severe in the DCT domain than WT; this is due to the fact that DCT atoms are not compactly supported.

Next we investigate the sparsity of a set of *realized* spectra in the two dictionaries. We simulate 100 maps from the theoretical power spectrum used previously and estimate their power spectra using equation 4 . As before, we decompose each realization in the DCT and WT dictionaries and reconstruct keeping increasing fractions of the largest coefficients. At this stage, it is important to note that, as we are dealing with the empirical power spectrum, we are no longer in an approximation setting but rather in an estimation one. Indeed, the empirical power spectrum can be seen as 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 (large bias). The converse is true when a large proportion of coefficients is kept in the reconstruction. Therefore, there will exist a threshold value that will entail a bias-variance tradeoff, 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

$$\mathrm{NMSE}_M = \frac{\left\| C[\ell] - \widehat{C}[\ell]^{(M)} \right\|_2}{\left\| C[\ell] \right\|_2} \times 100 \,, \tag{9}$$

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

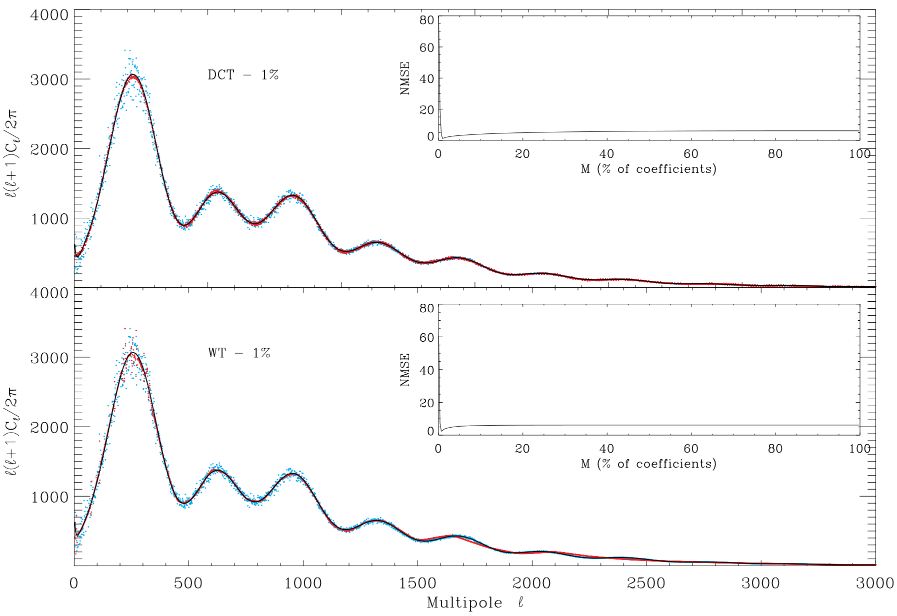


**Fig. 3.** A simulated CMB power spectrum along with the reconstructed spectra, using the DCT and WT dictionaries. The black solid line is the true underlying power spectrum from which the simulations were made. The blue and red dots show the simulated and the reconstructed power spectra respectively. 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 shows the NMSE curves.

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 are around 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. Conversely, the WT cannot reconstruct the proper shape of the power spectrum, but provides a smoother estimate.

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

- the CMB power spectrum is very sparse in both the 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;
- in the case of realizations, WT recovers more localized (noisy) features than the global ones, 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

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

$$\min_X \left\|\Phi^T X\right\|_1 \quad s.t. \quad \|Y - X\|_2 \leq \delta \ , \tag{10}$$

where $\Phi^T X$ represent the transform coefficients of $X$ in the dictionary $\Phi$, and $\delta$ controls the fidelity to the data and obviously depends on the noise standard deviation $\sigma$.

Let's 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 $\chi^2$-distributed noise with a number of degrees of freedom that depends on $\ell$. That is, the noise has a variance profile that dependents both on the true spectrum and $\ell$. 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 goes by the name of periodogram denoising (Donoho 1993). In (Komm et al. 1999), approximating the noise with a correlated Gaussian noise model, a threshold was derived at each wavelet scale using the MAD (Median of Absolute Deviation) estimator. A more elegant approach was proposed in (Donoho 1993; Moulin 1994), where the so-called Wahba VST was used. This VST is defined as:

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

where $\gamma = 0.57721...$ 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 (4), one can show that under mild regularity assumptions on the true power spectrum,

$$\widehat{C}[\ell] \xrightarrow{d} C[\ell] Z[\ell], \text{ where } \ \forall \ell \geq 2,\ 2LZ[\ell] \sim \chi^2_{2L}, L = 2\ell + 1 \ . \tag{12}$$

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

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

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 $\sigma_L^2 = \mathrm{Var}[\eta[\ell]] = \psi_1(L)$, 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 $\sigma_L^2$ as $\ell$ increases. At low $\ell$, normality is only an approximation. In fact, it can be show that the noise $\eta[\ell]$ has a probability density function of the form

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

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

In order to standardize the noise, the VST (13) will be slightly modified to the normalized form

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

where now the noise $\varepsilon[\ell]$ is 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 (15) to $L = 0$.

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

### 3.2. Signal detection in the wavelet domain

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

In order to estimate the true CMB power spectrum from the wavelet transform, it is important to detect the wavelet coefficients which are "significant", i.e. the wavelet coefficients which have an absolute value too large to be due to noise (cosmic variance + instrumental noise). Let $w_j[\ell]$ the wavelet coefficient of a signal $Y$ at scale $j$ and location $\ell$. We define the multiresolution support $M$ of $Y$ as:

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

For Gaussian noise, it is easy to derive an estimation of the noise standard deviation $\sigma_j$ 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 $\kappa$ ranges from 3 to 5. This means that a small magnitude compared to the threshold implies that the coefficients is very likely to be due to noise and hence insignificant. Such a decision rule corresponds to the hard-thresholding operator

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

To summarize, The multiresolution support is obtained from the signal $Y$ by computing the forward transform coefficients, applying hard thresholding, and recording the coordinates of the retained coefficients.

### 3.3. Power Spectrum Recovery Algorithm

Let's now turn to the adaptive estimator of the true CMB power spectrum $C[\ell]$ from its empirical estimate $\widehat{C}[\ell]$. As we benefit from the (asymptotic) normality of the noise in the stabilized samples $C^s[\ell]$ in (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 reconstruct an estimate $\widetilde{X}$ of the true power spectrum, known to be sparsely represented in some dictionary $\Phi$(regularization), such that the significant transform coefficients of its stabilized version reproduce those of $C^s$ (fidelity to data). Furthermore, as a power spectrum is a positive, a positivity constraint must be imposed. These requirements can be cast as seeking an estimate that solves the following constrained optimization problem:

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

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, beside 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 obvious to solve.

In this paper 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 $N_{\max} - 1$,

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

where $\mathcal{P}_+$ denotes the projection on the positive orthant and guarantees non-negativity of the spectrum estimator, $\mathrm{ST}_{\lambda_n}(w) = \left(\mathrm{ST}_{\lambda_n}(w[i])\right)_i$ is the soft-thresholding with threshold $\lambda_n$ that applies term-by-term the shrinkage rule

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

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)

### 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 $\widehat{C}[\ell]$ requires to remove this instrumental noise. The instrumental noise is assumed stationary and independent from the CMB. We will also suppose that we have access to the power spectrum of the noise, or we can compute the empirical 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:

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

Thus, (19) becomes

$$\begin{aligned} \widetilde{X} &= \mathcal{R}\left(\mathcal{T}\left(X^{(n)} + \widehat{S}_N\right) + \Phi M \odot \left(\Phi^T\left(C^s - \mathcal{T}\left(X^{(n)} + \widehat{S}_N\right)\right)\right)\right) - \widehat{S}_N \\ X^{(n+1)} &= \mathcal{P}_+\left(\Phi\ \mathrm{ST}_{\lambda_n}(\Phi^T \widetilde{X})\right) . \end{aligned} \tag{22}$$

*3.5. Combining Several Dictionnaries*

We have seen in Section 2.2 that the WT and DCT dictionaries had complementary benefits. Indeed each dictionary is able to capture well features with shapes similar to its atoms. More generally, assume that we have $D$ dictionaries $\Phi_1, \cdots, \Phi_D$. Given a candidate signal $Y$, we can derive a support $M_d$ associated to each dictionary $\Phi_d$, for $d \in \{1, \cdots, D\}$. The optimization problem to solve now reads

$$\min_X \|\Phi^T X\|_1 \quad \text{s.t.} \quad \begin{cases} X \geqslant 0 \\ M_d \odot \left(\Phi_d^T \mathcal{T}(X + \widehat{S}_N)\right) = M_d \odot \left(\Phi_d^T C^s\right),\ d \in \{1, \cdots, D\} \end{cases} . \tag{23}$$

Again, this is a challenging optimization problem. We propose to attack it by applying successively and alternatively (22) on each dictionary $\Phi_d$. Algorithm 1 describes in detail the different steps.

---

**Algorithm 1:** TOUSI Power Spectrum Smoothing with $D$ dictionaries

**Require:**
Empirical power spectrum $\widehat{C}$, $D$ dictionaries $\Phi_1, ..., \Phi_D$, noise power spectrum $\widehat{S}_N$,
Number of iterations $N_{\max}$,
Threshold $\kappa$ (default value is 5).

***Detection***

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

***Estimation***

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

---

## 4. 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 previously and the power spectra of the combined maps were estimated using equation 4.

Figure 4 shows the reconstruction of the theoretical CMB spectrum in the presence of noise. The blue dots show the empirical power spectrum of one realization having 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 $\ell$ up to 2500. It can be seen that TOUSI can do a great job in reconstructing the input power spectrum even in the presence of instrumental noise.

## 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,

$$\widetilde{C}^A[b] = \frac{1}{b(b+1)\omega_b} \sum_{\ell = b - \frac{\omega_b}{2}}^{b + \frac{\omega_b}{2}} \ell(\ell+1)\widehat{C}[\ell] \ , \tag{24}$$

and the window size $\omega_b$ is increasing with $\ell$. Here, we use window sizes of $\{1, 2, 5, 10, 20, 50, 100\}$ respectively for $\ell$ 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).

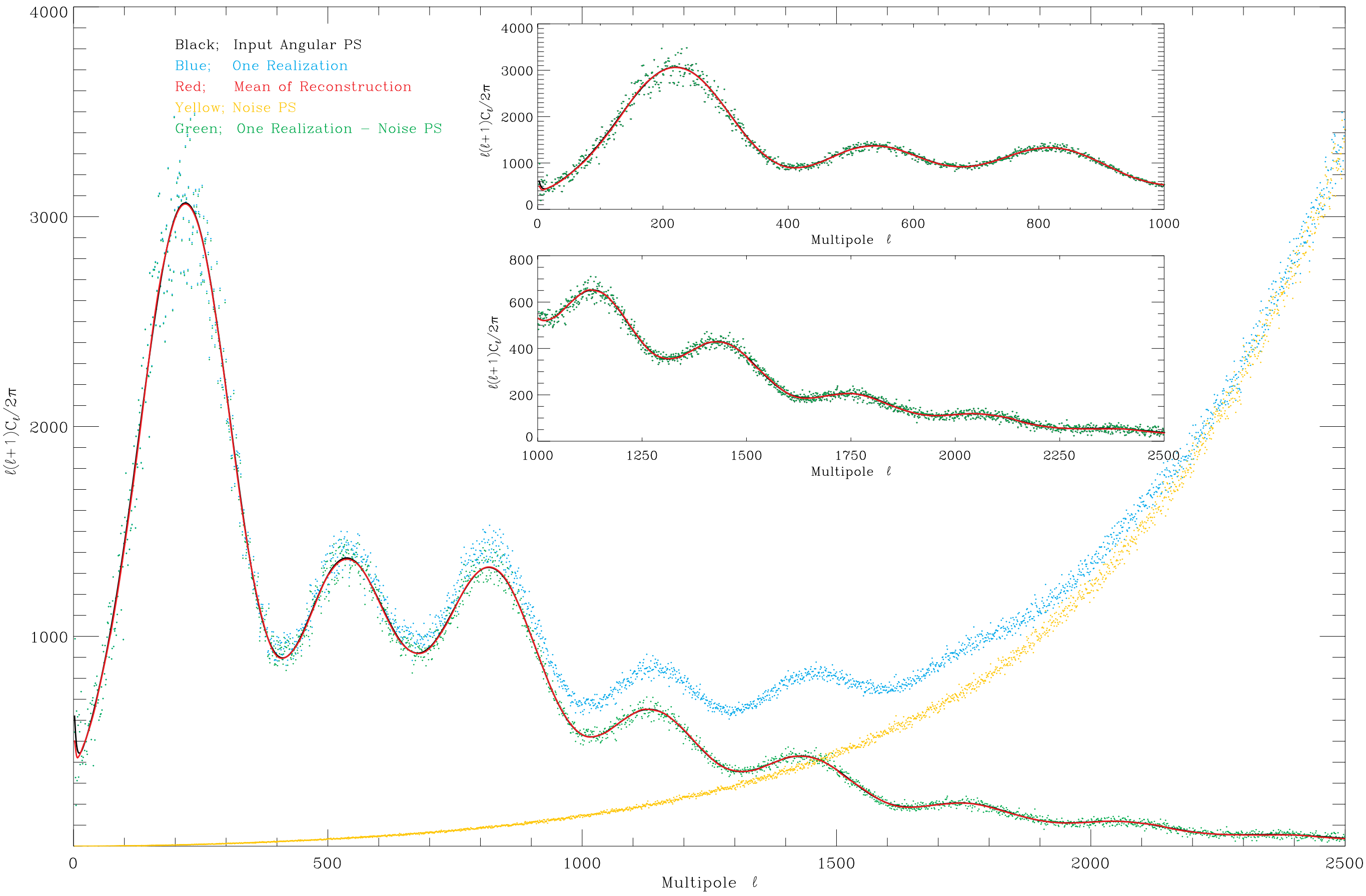


**Fig. 4.** Power spectrum estimation in the presence of instrumental noise. The blue dots show the empirical power spectrum of one realization having instrumental noise. Yellow dots show the estimated power spectrum of one of the simulated noise maps. Green dots show the the spectrum with the noise 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.

Figure 5 shows the average error the 100 realizations as a function of $\ell$

$$E[\ell] = \frac{1}{100} \sum_{i=1}^{100} \parallel C[\ell] - \widetilde{C}_i[\ell] \parallel_2 \,, \tag{25}$$

where $\widetilde{C}_i$ 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 over-plotted as a solid black line. We can see that 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 all the statistical information are contained in the CMB angular power spectrum. In this paper we have investigated the sparsity of the CMB power spectrum in two dictionaries; 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, that 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 get a highly accurate estimate from a single realization of the CMB power spectrum. This could be a replacement for CAMB in cases where 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`

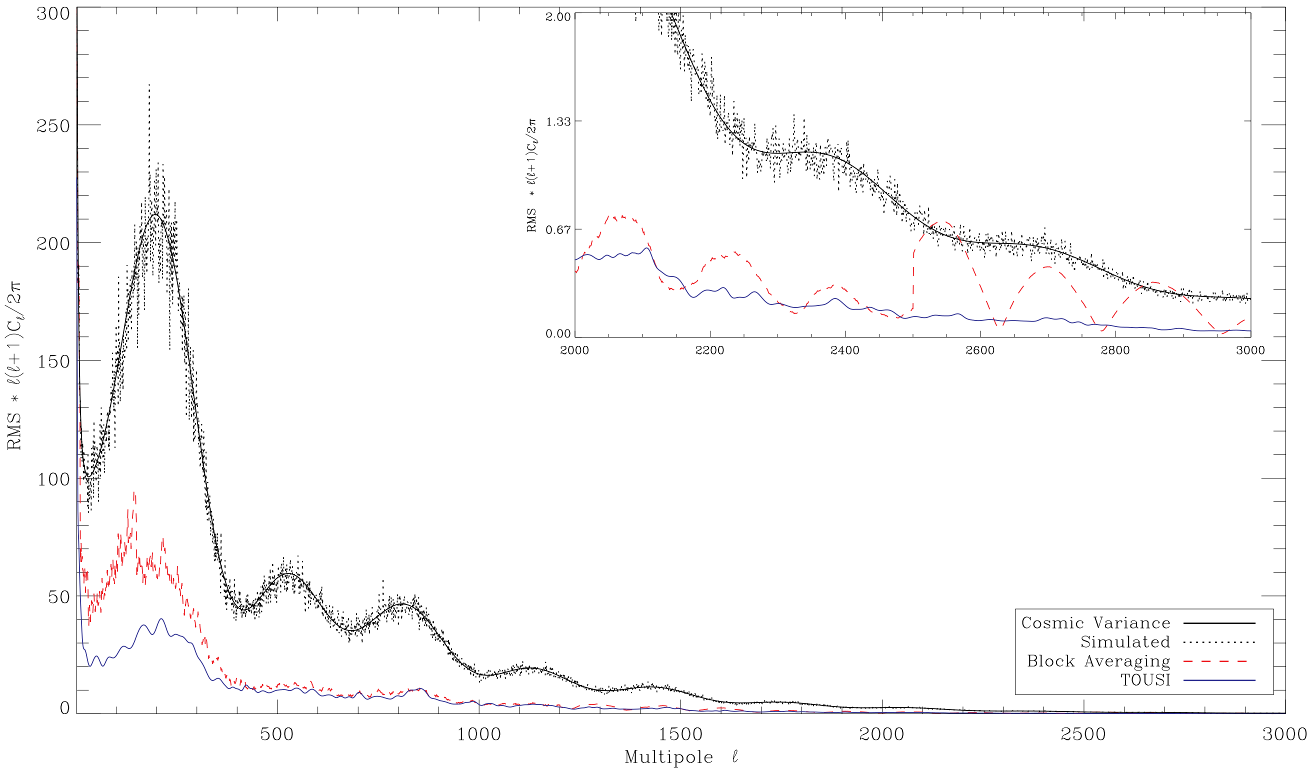


**Fig. 5.** Mean error for the 100 realizations, for the realizations (black dotted line), the averaging denoising (red dashed line) and the sparse wavelet filtering (blue solid line). The inner plot shows a zoom between $l = 2000$ and $l = 3000$.

## Acknowledgments

The authors would like to 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

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
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., Challinor, A., & Lasenby, A. 2000, Astrophys. J., 538, 473
Moulin, P. 1994, IEEE Transactions on Signal Processing, 42, 3126–3136
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2011, ArXiv e-prints
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
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)