A METHOD FOR EXTRACTING MAXIMUM RESOLUTION POWER SPECTRA FROM MICROWAVE SKY MAPS [1]

Max Tegmark

Max-Planck-Institut für Physik, Föhringer Ring 6, D-80805 München
email: max@mppmu.mpg.de

Abstract

A method for extracting maximal resolution power spectra from microwave sky maps is presented and applied to the 2 year COBE data, yielding a power spectrum that is consistent with a standard $n = 1$, $Q_{rms-ps} = 20\,\mu$K model.
By using weight functions that fall off smoothly near the galactic cut, it is found that the spectral resolution $\Delta\ell$ can be more than doubled at $\ell = 15$ and more than tripled at $\ell = 20$ compared to simply using galaxy-cut spherical harmonics. For a future high-resolution experiment with reasonable sky coverage, the resolution around the CDM Doppler peaks would be enhanced by a factor of about 100, down to $\Delta\ell \sim 1$, thus allowing spectral features such as the locations of the peaks to be determined with great accuracy. The reason that the improvement is so large is basically that functions with a sharp edge at the galaxy cut exhibit considerable "ringing" in the Fourier domain, whereas smooth functions do not. The method presented here is applicable to any survey geometry, chopping strategy and exposure pattern whatsoever. The so called signal-to-noise eigenfunction technique is found to be a special case, corresponding to ignoring the width of the window functions.

[1] Published in *MNRAS*, **280**, 299-308 (1996).
Available from *h t t p://www.sns.ias.edu/~max/window.html* (faster from the US)
and from *h t t p://www.mpa-garching.mpg.de/~max/window.html* (faster from Europe).

1 INTRODUCTION

There has been a surge of interest in the cosmic microwave background radiation (CMB) since the first anisotropies of assumed cosmological origin were detected by the COBE DMR experiment (Smoot *et al.* 1992). On the experimental front, scores of new experiments have been carried out and many more are planned or proposed for the near future. On the theoretical front, considerable progress has been made in understanding how the CMB power spectrum C_ℓ depends on various cosmological model parameters – see Hu & Sugiyama (1995), Bond *et al.* 1994 and references therein for recent reviews of analytical and quantitative aspects of this problem. This aim of this paper is to strengthen the link between these two fronts, by presenting a method allowing more accurate estimation of the power spectrum from experimental data.

The usual approach to power-spectrum estimation from CMB experiments has been to assume that some specific model is true, and use a likelihood analysis to find the model parameters that best fit the observed data. Since COBE is mainly sensitive to multipoles $\ell \lesssim 20$, where many models predict a power spectrum C_ℓ that is well fit by a simple power law, most published work on the topic has focused on estimating merely two parameters; the overall normalization and a spectral index n, roughly the logarithmic slope of the power spectrum. This approach has since been generalized to quite a variety of two-parameter models, for instance to a class of curved $\Omega < 1$ cosmologies (Górski *et al.* 1995), where the model parameters are Ω and the normalization, and to flat $\Omega < 1$ cosmologies with a cosmological constant (Bunn & Sugiyama 1995 – hereafter BS95; Stompor *et al.* 1995) where the model parameters are the cosmological constant Λ and the normalization. The most elaborate such scheme to date is that of White & Bunn (1995, hereafter WB95), using a three-parameter model. For the purpose of making such parameter fits efficiently, a number of sophisticated data analysis techniques have been invented, where the data is expanded in some set of basis functions that are orthogonal on the celestial sphere after the galaxy has been cut out. It has recently been shown (Tegmark & Bunn 1995, hereafter referred to as TB95) that the orthogonalized spherical harmonics (Górski 1994, hereafter G94) and the signal-to-noise eigenmodes (Bond 1994 – hereafter B95, BS95) give results that are both virtually identical and virtually optimal for parameter fitting, so that the choice of which method to use is mostly an issue of computational efficiency.

However, as CMB data continues to improve in quantity and quality, it becomes increasingly attractive to go beyond mere parameter fitting and break the shackles of model-dependence. Just as is routinely done with the power spectrum $P(k)$ of galaxies, we wish to simply estimate the angular power spectrum without assuming any model, *i.e.*, estimate the entire sequence of multipole moments C_ℓ from the data one by one, and so to say let the data speak for itself. To estimate say the first 30 multipoles in this fashion thus means fitting a 30-parameter model to the data. Given the numerical difficulties of fitting even three parameters at once, this is clearly unfeasible with the above-mentioned methods at the present time.

Fortunately, this program can readily be carried through by a more direct approach, where the step of maximum-likelihood estimation is omitted, and the multipoles C_ℓ are roughly speaking estimated by simply expanding the data in some modified versions of the spherical harmonics and then squaring the coefficients. Except for technicalities like shot-noise correction, this is analogous to the way that the matter power spectrum $P(k)$ is estimated from galaxy redshift surveys: the observed density field is expanded in some functions that resemble plane waves, and then the expansion coefficients are squared. Pioneering work on this problem (Peebles 1973; Hauser & Peebles 1973) has recently been extended and applied to the 2 year COBE data (Wright *et al.* 1994b, hereafter W94), and the result of such an analysis is shown in Figure 1 (top). When estimating power spectra, it is customary (White, Scott & Silk 1994) to place both vertical and horizontal error bars on the data points, as in Figure 1. The former represent the uncertainty due to noise and cosmic variance, and the latter reflect the fact that an estimate of C_ℓ inadvertently also receives contributions from other multipole moments. As is well-known, this unavoidable effect is caused by incomplete sky coverage, which destroys the orthogonality of the spherical harmonics. Another way of phrasing this is that any estimate of C_ℓ will be the true power spectrum convolved with a *window function* that has some non-zero width $\Delta\ell$. Here and throughout, we will refer to these horizontal bars $\Delta\ell$ as the *resolution* of the power spectrum estimate. Poor resolution (large $\Delta\ell$) clearly destroys information by smearing out features in the true spectrum. For typical ground- and balloon-based experiments probing degree scales, this spectral blurring $\Delta\ell/\ell$ tends to be of order unity, which makes it difficult to resolve details such as the number of Doppler peaks. A much better method is that presented in W94, where the relative resolution $\Delta\ell/\ell$ is brought down to the order of 25% by using the COBE sky map. In this paper, we will see how to reduce the horizontal error bars still further, down to their theoretical minimum, which for the COBE data is seen to be a resolution $\Delta\ell \approx 1$. The method is akin in spirit to that of Tegmark (1995, hereafter T95) for power spectrum estimation from galaxy surveys, in which the best way to weight the galaxies turns out to be given, surprisingly, by the Schödinger equation.

This paper is organized as follows. The problem is formalized in Section 2 and solved in Section 3 for the most general case. The method is applied to the special case of the 2 year COBE DMR data in Section 4. In Section 5, it is seen that, just as for the above-mentioned galaxy survey problem, the gist of the method can be understood from a simple analogy with quantum mechanics. The method is compared to other techniques in Section 6, and its implications for future CMB experiments are discussed.

2 THE PROBLEM

Although the following three sections are rather technical in nature, focusing on the application to real data, we will see in Section 5 that the essence of both the problem and its solution have quite a simple and intuitive interpretation.

Since the computations that follow are all linear algebra, vector and matrix notation will be used as extensively as possible. If the introduction of some notation is found too cursory, the reader is referred to TB95 for more details. Although the true CMB fluctuations are described by some smooth temperature distribution $\frac{\Delta T}{T}$ over the celestial sphere, a real-world CMB sky map is always discretized into some finite number of pixels, say N of them. Let us write the CMB sky map as the N-dimensional vector $\mathbf{x}$, where

$$x_i = \frac{\Delta T}{T}(\widehat{\mathbf{r}}_i), \tag{1}$$

and $\widehat{\mathbf{r}}_i$ is a unit vector in the direction of the i^{th} pixel. $N = 6144$ for the all-sky COBE DMR map.

For reasons that will become clear later, we define a new vector $\mathbf{z} \equiv A\mathbf{x}$, where A is some completely arbitrary $N' \times N$ matrix. Later on, we will see that considerable savings in computer time can often be made by choosing A wisely. Using equation (2) from TB95, we obtain

$$\mathbf{z} = A(Y\mathbf{a} + \varepsilon) \tag{2}$$

by choosing the matrix A to have all its row vectors orthogonal to the monopole and the dipole, thus eliminating the "nuisance contribution" from these multipoles. The $N \times \infty$-dimensional spherical harmonic matrix Y is defined as $Y_{i\lambda} \equiv Y_{\ell m}(\widehat{\mathbf{r}}_i)$, where we have combined ℓ and m into the single index $\lambda \equiv \ell^2+\ell+m+1 = 1,2,3,....$ (Throughout this paper, we use real-valued spherical harmonics, which are obtained from the standard spherical harmonics by replacing $e^{im\phi}$ by $\sqrt{2}\sin m\phi$, 1, $\sqrt{2}\cos m\phi$ for $m<0$, $m=0$, $m>0$ respectively.) Making the standard assumption that the CMB is Gaussian on the scales probed, $\mathbf{a}$ is an infinite-dimensional Gaussian random vector with zero mean and with the diagonal covariance matrix

$$\langle a_\lambda a_{\lambda'}\rangle = \delta_{\lambda\lambda'} B_\ell^2 C_\ell, \tag{3}$$

where C_ℓ is the angular power spectrum and where B_ℓ^2 is the experimental beam function. The N-dimensional noise vector ε is assumed to be Gaussian with $\langle\varepsilon\rangle = 0$ and $\langle\varepsilon_i\varepsilon_j\rangle = \sigma_i\sigma_j\delta_{ij}$ (for the COBE case, see Lineweaver *et al.* 1994), where σ_i is the rms noise of pixel i.

We wish to estimate the power spectrum C_ℓ from $\mathbf{z}$. As C_ℓ tends to decrease rapidly with ℓ, we will find it convenient to define

$$D_\ell \equiv B_\ell^2 C_\ell/\mu_\ell \tag{4}$$

for some weights μ_ℓ that make the coefficients D_ℓ of similar magnitude, as was done in WB95. We will leave μ_ℓ arbitrary for now, and estimate D_ℓ. Everything below refers to some definite multipole ℓ^*, say $\ell^*=7$, and is to be repeated separately for all other ℓ^*-values of interest. The most general estimate of D_{ℓ^*} that is quadratic in the data can clearly be written as

$$\tilde{D} \equiv \mathbf{z}^t E\mathbf{z}, \tag{5}$$

where E is some arbitrary real symmetric $N'\times N'$ matrix. Let us expand it in an orthogonal set of eigenvectors $\{\mathbf{e}_k\}$ as

$$E = \sum_{k=1}^{N''} \alpha_k \mathbf{e}_k {\mathbf{e}_k}^t, \tag{6}$$

where $\alpha_k \neq 0$ (we simply omit vanishing eigenvalues from the sum, so that N'' is the rank of the matrix). Since D is by definition positive, we clearly want E to be positive semidefinite, *i.e.*, $\alpha_k > 0$. Substituting equations (2), (4) and (6) into (5), we obtain

$$\begin{aligned}\langle\tilde{D}\rangle &= \sum_{k=1}^{N''} \alpha_k \left[\mathrm{tr}\, Y^t A^t \mathbf{e}_k {\mathbf{e}_k}^t A Y \langle \mathbf{a}\mathbf{a}^t\rangle + \mathrm{tr}\, A^t \mathbf{e}_k {\mathbf{e}_k}^t A \langle\varepsilon\varepsilon^t\rangle\right] &(7)\\ &= \sum_{\ell=2}^{\infty} v_\ell^2 D_\ell + B, &(8)\end{aligned}$$

where we have defined the *window function*

$$v_\ell^2 \equiv \mu_\ell \sum_{k=1}^{N''} \alpha_k \sum_{m=-\ell}^{\ell} (V^t \mathbf{e}_k)_\lambda^2 \tag{9}$$

and the *noise bias*

$$B \equiv \sum_{k=1}^{N''} \alpha_k \sum_{i=1}^{N} (A^t \mathbf{e}_k)_i^2 \sigma_i^2, \tag{10}$$

and $V \equiv AY$. This merely reflects well-known fact that the expectation value of any quadratic combination of pixels is just the power spectrum convolved with some window function, plus a positive noise bias. We define the *noise-corrected* estimate by

$$\tilde{D}^{corr} \equiv \tilde{D} - B. \tag{11}$$

We want the window function to represent a weighted average of the D_ℓ-coefficients, which requires it to be normalized so that

$$\sum_{\ell=2}^{\infty} v_\ell^2 = 1. \tag{12}$$

Defining

$$\langle f(\ell) \rangle \equiv \sum_{\ell=2}^{\infty} v_\ell^2 f(\ell) \tag{13}$$

for an arbitrary function f, we can write this normalization condition as simply $\langle 1 \rangle = 1$. Ideally, we would want the window function to be $v_\ell^2 = \delta_{\ell\ell^*}$, but this is of course impossible to achieve if the sky coverage is incomplete. Given this limitation, we simply want the window function to be *as narrow as possible*, centered around ℓ^*. We may for instance choose to minimize its variance (second moment) about ℓ^*, $\langle (\ell - \ell^*)^2 \rangle$, or more generally the quantity $\langle |\ell - \ell^*|^\nu \rangle$ for some positive exponent ν. Let us allow for complete freedom of preferences by minimizing $\langle p_\ell \rangle$ for a completely arbitrary *penalty function* p_ℓ. We thus arrive at the following optimization problem:
Find the matrix E that minimizes $\langle p_\ell \rangle$ subject to the constraint that $\langle 1 \rangle = 1$.

3 THE SOLUTION

We will now solve this constrained minimization problem. It is easy to see that an optimal solution can always be found where the matrix E has rank

1, *i.e.*, where it is of the simple form $E = \mathbf{e}\mathbf{e}^t$ for some vector $\mathbf{e}$. Below we will find a set of N' different rank 1 matrices $\mathbf{e}_k{\mathbf{e}_k}^t$ which have the following properties:

- They are independent in the sense that $\langle(\mathbf{e}_k\cdot\mathbf{z})(\mathbf{e}_{k'}\cdot\mathbf{z})\rangle \propto \delta_{kk'}$.
- $k = 1$ corresponds to an optimal solution, $k = 2$ to the second best solution, *etc.*

If the only objective were to minimize the horizontal error bars in Figure 1, the optimal choice would thus be $E = \mathbf{e}_1{\mathbf{e}_1}^t$. However, we clearly also want to keep the vertical error bars as small as we can. Since cosmic variance drops as we average many different estimates, it is in general better to choose E to be some weighted average of the above-mentioned rank one matrices, as given by equation (6) when normalizing the weights as $\sum_{k=1}^{N''}\alpha_k = 1$. We clearly face a trade-off between vertical and horizontal error bars: as we increase N'', the vertical error bar decreases, but since we are including increasingly wide window functions, the horizontal error bar increases. For reasonable sky coverage except for a galactic cut, the choice of N'' turns out to be quite an easy one: the best $(2\ell^*+1)$ rank one matrices clearly stand out in front of the rest of the pack, and all give quite similar error bars. Once we have fixed N'', changing the weights α_k leaves the horizontal error bars fairly unaffected, and we can choose α_k so as to minimize the vertical error bars. In the limit of complete sky coverage, these $(2\ell^*+1)$ best vectors $\mathbf{e}_k$ approach the $(2\ell^*+1)$ spherical harmonics in question, just as we would expect.

Now let us find the vectors $\mathbf{e}_k$. Writing $E = \mathbf{e}_k{\mathbf{e}_k}^t$ (no summation implied), we solve the constrained minimization problem for $\mathbf{e}_k$ using the standard method of Lagrange multipliers. Defining

$$\mathcal{L} \equiv \langle p_\ell\rangle + \gamma B - \lambda(\langle 1\rangle - 1) \tag{14}$$

and requiring the derivatives with respect to the components of $\mathbf{e}_k$ to vanish, we obtain

$$(Q - \lambda R)\mathbf{e}_k = 0, \tag{15}$$

where we have defined the $N'\times N'$ matrices

$$Q_{ij} = \sum_\lambda V_{i\lambda}V_{j\lambda}\mu_\ell p_\ell + \gamma\sum_{k=1}^{N} A_{ik}A_{jk}\sigma_k^2, \tag{16}$$

$$R_{ij} = \sum_\lambda V_{i\lambda}V_{j\lambda}\mu_\ell. \tag{17}$$

Here we have added the term γB to the target function to make the noise bias B (and thus the vertical error bars) small. The choice of γ is discussed below. Since both Q and R are symmetric, this generalized eigenvalue problem has N' real solutions, which we normalize so that $\langle 1\rangle = 1$ and sort by their eigenvalues (the smallest eigenvalue corresponds to $k = 1$, the best solution, *etc.*). Most standard eigenvalue packages (such as the public-domain package EISPACK, or that included in NAG) provide a specialized routine for precisely this problem: the generalized eigenvalue problem where Q and R are real and symmetric, and R is positive definite.

Once we have solved this and chosen N'', the variance of our estimate is

$$V(\tilde{D}) = 2\sum_k\sum_{k'}\alpha_k\alpha_{k'}M_{kk'}^2, \tag{18}$$

where the matrix $M_{kk'} \equiv \langle(\mathbf{e}_k\cdot z)(\mathbf{e}_{k'}\cdot z)\rangle$ is given by

$$M_{kk'} = \sum_\lambda (V^t\mathbf{e}_k)_\lambda(V^t\mathbf{e}_{k'})_\lambda\mu_\ell D_\ell + \sum_{i=1}^N (A^t\mathbf{e}_k)_i(A^t\mathbf{e}_{k'})_i\sigma_i^2. \tag{19}$$

When M is almost diagonal, the variance is approximately minimized if we choose the weights $\alpha_k \propto M_{kk}^{-2}$. It is easy to show that the first term will be strictly diagonal. The second term will typically be close to diagonal, with a slight off-diagonality caused by heteroscedastic noise.

3.1 How to chose the parameters

The above treatment was very general, leaving the choice of the matrix A, the weights μ, the penalty function p_ℓ and the constant γ to be chosen according to the preferences of the user. We now discuss some aspects of the choice of these free parameters.

The matrix A. A is conveniently constructed in two steps:

1. One makes some natural choice such as the spherical harmonic matrix Y^t, the noise-weighted and orthogonalized ditto introduced by G94, the time-saving matrix described in the next section, or the identity matrix if one prefers to simply use the pixel values for a brute-force approach à la TB95, obtaining the best possible result at a high computational cost.

2. One makes all the row vectors orthogonal to the monopole and the dipole, as described in *e.g.* TB95.

The matrix A also allows us to directly incorporate data sets other than sky maps. For example, for a double-chop configuration such as that of the Tenerife experiment, the observed data vector $\mathbf{z}$ is a certain linear combination of the actual sky pixels $\mathbf{x}$ (which are never observed directly), and since the coefficients in these linear combinations (the elements of A) are known, the matrix A incorporates all we need to know about the data acquisition process.

The weights μ_ℓ. This is basically the choice of what to label the vertical axis with in Figure 1. If we probe a linear function with a symmetric and correctly centered window function, the estimates will be unbiased. However, if we are probing the power spectrum C_ℓ, which is generally believed to have convex shape (the second derivative of say $1/\ell(\ell+1)$ is positive), then our estimates tend to be biased high, since the upward bias from coupling to lower multipoles (the so called "red leak") is stronger than the downward bias from coupling to higher multipoles (the "blue leak"). For instance, when attempting to estimate C_{20}, even a very slight sensitivity to the quadrupole C_2 is likely to dwarf the signal one is trying to measure, as pointed out in *e.g.* W94 and illustrated in Figure 3. A simple strategy for reducing such problems is to choose μ_ℓ proportional to the expected power spectrum multiplied by the experimental window function, *i.e.*,

$$\mu_\ell = \frac{B_\ell^2}{\ell(\ell+1)}, \tag{20}$$

so that the coefficients D_ℓ will all be of the same order of magnitude. This is quite a natural choice also because power spectra are conventionally plotted with the quantity $\ell(\ell+1)C_\ell/2\pi \propto D_\ell$ on the vertical axis. Thus the horizontal error bars will refer to exactly the quantities that we plot, *i.e.*, D_ℓ, which is of course the most honest way to present the results.

The penalty function p. The simple choice $p_\ell = (\ell-\ell^*)^2$ is quite adequate if μ_ℓ is a mere constant. This gives $p_\ell\mu_\ell \sim \ell^2$ as $\ell\to\infty$, and corresponds to the quadratic penalty function $p_{\mathbf{k}} = k^2$ as $\ell\to\infty$ in the galaxy survey problem treated in T95. In T95, it was seen that this k^2 penalty in Fourier space corresponded to a Laplace operator ∇^2 in real space, and that this guaranteed that the resulting solutions were continuous and well-behaved. A k^4 term would make the solutions even smoother, guaranteeing that all derivatives would be continuous as well, *etc.* If $p_\ell\mu_\ell$ asymptotically grows slower than ℓ^2, however, high frequencies in the weight function are not sufficiently suppressed, and there is a risk that the solutions will be quite irregular and unnatural. Since we advocated choosing $\mu_l = B_\ell^2/\ell(\ell+1)$

above, we thus need to make $p_\ell \to \infty$ at least as fast as ℓ^4/B_ℓ^2 as $\ell \to \infty$. In addition, if one is particularly worried about say non-cosmic contamination from a particular multipole like the quadrupole, one can increase the value of p_2, thereby reducing the value of the window function at $\ell = 2$. There are of course no free lunches: the price of reducing v_2^2 in this fashion is that other unwanted window function entries increase.

The constant γ. Since there is a trade-off between vertical and horizontal error bars, we minimize a combination of them. The larger we make γ, the greater the emphasis on the vertical ones. A larger γ tends to increase the weights given to the least noisy pixels. As long as the pixel noise levels are of the same order of magnitude all across the map, the results are quite insensitive to the choice of γ, and one might as well set $\gamma = 0$.

It should be stressed that these free parameters reflect a strength rather than a weakness of the method. The goal is to find optimal window functions, and the choices of μ, p and γ simply reflect what we mean by "good". Different choices give slightly different error bars on the resulting plots, but even if one criterion for "good" is used in the optimization process and a different one is used when judging the results, the optimization tends to improve the situation greatly over simply using spherical harmonics. In Section 5, we will see why.

3.2 Analyzing very large data sets

When applying this method to COBE, where the matrices are never larger than 4038×4038, the solution of the resulting eigenvalue problem is numerically straightforward. Future high resolution CMB experiments, however, are likely to produce data sets where the pixels are counted in millions rather than thousands. Although it is obviously not feasible to diagonalize matrices of this size, the method presented above can nonetheless be readily applied to such data sets, as follows:

- For small ℓ, almost no information is lost if the number of pixels is reduced by smoothing prior to the analysis.
- For large ℓ, it is convenient to split the sky into many smaller patches, analyze them separately, and then combine the results. This of course lowers the spectral resolution somewhat, but as is discussed in Section 5, $\Delta\ell/\ell$ can still be kept extremely small, of the order of a few percent.

4 APPLICATION TO COBE

We now apply the method to the two year COBE DMR data, combining the 53 and 90 GHz channels and removing all pixels less than $20°$ away from the galactic plane. This leaves $N = 4038$ pixels.

4.1 Parameter choices

Just as in WB95, we choose $\mu_\ell = B_\ell^2/\ell(\ell+1)$, where B_ℓ^2 is the COBE beam window function published by Wright *et al.* (1994a). In line with our discussion above, we estimate D_{ℓ^*} by chosing the penalty function

$$p_\ell = \begin{cases} (\ell-\ell^*)^2/\mu_{\ell^*+2} & \text{if } \ell \leq \ell^*+2, \\ (\ell-\ell^*)^2/\mu_\ell & \text{if } \ell \geq \ell^*+2. \end{cases} \tag{21}$$

This behaves as a simple quadratic penalty function $p_\ell \propto (\ell-\ell^*)^2$ near ℓ^*, whose symmetry ensures that the "red leak" will be similar to the "blue leak" and thus that $\langle\ell\rangle \approx \ell^*$, but also has the above-mentioned desired property that $p_\ell\mu_\ell \propto \ell^2$ as $\ell \to \infty$. As the COBE signal-to-noise is quite good and the pixel noise does not vary radically across the sky, we make the simple choice $\gamma = 0$. As the noise bias B is of order $B_* \equiv (70\mu\mathrm{K})^2\ell(\ell+1)/NB_\ell^2$ and the optimal horizontal error bars $\Delta\ell$ are of order unity, a natural alternative choice would be say $\gamma = 1/5B_*$. This choice would mean roughly that we value a 1% reduction in the horizontal error bar as much as a 5% reduction in the vertical error bar. We choose A to be the first 961 rows of Y^t, which corresponds to the spherical harmonic components up to $\ell = 30$. We place the cutoff at $\ell = 30$ simply because the COBE-beam effectively washes out the signal from $\ell \gg 20$, leaving mostly noise at these high frequencies. Indeed, by comparing G94 and TB95, one concludes that $\mathbf{z}$ contains almost all the cosmological information in $\mathbf{x}$ with this high frequency cutoff. We then throw out the first four rows of A, corresponding to the monopole and dipole, and make the remaining rows orthogonal to them.

4.2 A trick for saving time

Because the galactic cut preserves reflection symmetry about the galactic plane, even and odd multipoles remain orthogonal to one another (to very good accuracy, the only slight correction arising from pixelization effects). When estimating the D_ℓ corresponding to an even ℓ, we thus throw out all rows corresponding to odd ℓ, and vice versa.

Although diagonalizing 500×500 matrices (N' being of order $961/2$) is numerically straightforward, the solution of the eigenvalue problem given by equation (15) can be simplified further. Since the galactic cut preserves azimuthal symmetry, spherical harmonics with different m-values are also orthogonal to each other, so by simply regrouping the rows of A by m-values, the matrices Q and R become block-diagonal. In other words, we can do the following:

1. Pick some m-value and choose A to be the rows of Y^t that are not orthogonal to the row corresponding to $Y_{\ell,m}$. For say $m = 0$ and ℓ^* even, A will consist of the $N' = 14$ rows corresponding to $Y_{2,m}$, $Y_{4,m}$, ... , $Y_{30,m}$.
2. Compute Q and R, truncating the sums (16) and (17) at say $\ell = 60$, since the COBE beam B_ℓ exponentially suppresses multipoles $\ell \gg 15$.
3. Find the smallest eigenvalue of equation (15), normalize the corresponding eigenvector so that $\langle 1 \rangle = 1$ and denote it $\mathbf{e}_m$.
4. Repeat the corresponding procedure for the other m-values, finally obtaining the $N'' = 2\ell^* + 1$ vectors $\{\mathbf{e}_m\}$.
5. Compute the matrix M (as with all other methods, error bar estimation of course requires assuming some reasonable power spectrum — here the spectrum $n = 1$, $Q_{rms-ps} = 20\mu\mathrm{K}$ is used) and the optimal weights α_m.
6. Compute $\tilde{D}^{corr}$ and the vertical and horizontal error bars.
7. Repeat everything for all other ℓ^*-values of interest.

This is what has been done to produce Figure 1 (bottom).

4.3 Results

As expected, Figure 1 (bottom) shows the vertical error bars being dominated by cosmic variance for low ℓ and by noise for large ℓ, especially for $\ell \gtrsim 15$, where beam smearing takes its heavy toll. As to the horizontal error bars, the bars that specify the spectral resolution, we see that they exhibit hardly any ℓ-dependence — we will return to this fact in the next section, and see that it has a simple intuitive explanation. Having said this about the error bars, we now turn to the location of the data points. If the true power spectrum C_ℓ were a simple $n = 1$ Harrison-Zel'dovich spectrum with normalization $Q_{rms-ps} = 20\mu$K, then we would expect about 68% of the data points in Figure 1 to lie within the shaded region, which is the $1 - \sigma$ error region for this model, and this is clearly the case, in agreement with previous analyses such as W94 and de Oliveira-Costa & Smoot (1995). Comparing our results to those obtained with the truncated spherical harmonic method (top), we see that the individual multipole estimates agree quite well for low ℓ, where the spectral resolution $\Delta\ell$ of the two methods is comparable, but start to differ as ℓ increases, reflecting the fact that the data points in the upper figure are probing the average power in a much wider band of ℓ-values.

For verification, 1000 Monte-Carlo simulations were performed, where fake COBE-skies with noise and galaxy cut were generated with $n = 1$, $Q_{rms-ps} = 20\mu$K and then piped though the analysis software. The result was in perfect agreement with the theoretically predicted error bars, with zero bias, about 68% of the estimated multipoles falling within the shaded region in Figure 1, *etc.* The COBE analysis was also repeated for a number of different choices of the parameters γ, μ and p, and the results were found to remain virtually unaffected. In other words, the results obtained with the method we have presented are quite robust.

5 THE QUANTUM ANALOGY

Whereas sections 2, 3 and 4 of this paper were rather technical, this section shows that all qualitative features of the method can be understood from a simple analogy with quantum mechanics. Although all the elements of realism included above are important when applying the method to real data, one can in fact understand the qualitative features by ignoring most of these complications.

First of all, let us ignore the fact that a real CMB sky map is pixelized. We let the function $x(\widehat{\mathbf{r}})$ denote the temperature fluctuation $\Delta T/T$ in the direction of the unit vector $\widehat{\mathbf{r}}$. In the discrete real-world case, each pixel has some r.m.s. noise variance σ^2 and effectively covers some solid angle Ω of the sky. For our continuous analogue, let us define the noise variance function $V(\widehat{\mathbf{r}})$ by $V \equiv \sigma^2\Omega$, where σ and Ω refer to a pixel in the direction $\widehat{\mathbf{r}}$. If we neglect contamination problems, it is easy to see that $V(\widehat{\mathbf{r}})$ is simply proportional to the inverse of the time spent observing the direction $\widehat{\mathbf{r}}$ (the small-scale details of the pointing of the antenna are irrelevant, as the beam smearing will ensure that the function V is smooth on angular scales below the beam width). Secondly, let us ignore the nuisance terms from monopole, dipole, *etc.*, so that we can omit A from equation (2). Finally, we choose $\mu_\ell = 1$.

Since the power $D_{\ell*}$ has units of $\mu\mathrm{K}^2$, we clearly want to estimate it by some quantity $\tilde{D}$ that is quadratic in the data. The simplest estimator of this type is

$$\tilde{D} \equiv \left| \int \psi(\widehat{\mathbf{r}}) x(\widehat{\mathbf{r}}) d\Omega \right|^2, \tag{22}$$

where ψ is some function on the sphere, and above we showed that the most general estimate is just a weighted average of such estimates. A straightforward calculation shows that

$$\langle \tilde{D} \rangle = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} |\widehat{\psi}_{lm}|^2 D_\ell + \int |\psi(\widehat{\mathbf{r}})|^2 V(\widehat{\mathbf{r}}) d\Omega, \tag{23}$$

where $\widehat{\psi}_{lm}$ denotes the spherical Fourier transform of ψ, *i.e.*, the coefficients in an expansion of ψ in spherical harmonics. In other words, we see that the expectation value of our estimator is the sum of two terms of quite different character. The first, the contribution from cosmology, is the power spectrum convolved with a window function $v_\ell^2 \equiv \sum_m |\widehat{\psi}_{lm}|^2$. The second, the contribution from noise, is just an average value of V, the weights being $|\psi(\widehat{\mathbf{r}})|^2$. This is very similar to the result when estimating the power spectrum from a galaxy survey (T95). Just as in that paper, we will find it convenient to use the standard Dirac quantum mechanics notation with kets, bras and linear operators. This allows us to write $\widehat{\psi}_{\ell m} = \langle \ell m | \psi \rangle$. A window function should always integrate to unity, so the correct normalization for ψ is just $\langle \psi | \psi \rangle = 1$. Defining the operator

$$L \equiv \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \ell |\ell m\rangle\langle \ell m|, \tag{24}$$

equation (23) becomes simply

$$\langle \tilde{D} \rangle = \langle \psi | D_L + V(\widehat{\mathbf{r}}) | \psi \rangle. \tag{25}$$

Note that L is a scalar operator satisfying $L|\ell m\rangle = \ell|\ell m\rangle$, and is related to the (vector) angular momentum operator $\mathbf{L} = -i\mathbf{r} \times \nabla$ through $\mathbf{L}^2 = L(L+1)$.

Now what is the best choice of ψ? Equation (22) tells us that $\tilde{D}$ is the square modulus of a random variable whose real and imaginary parts are both Gaussian. Thus if ψ is real, the standard deviation of $\tilde{D}$ (the vertical error bar) is simply $\sqrt{2}$ times its expectation value. (With random phases, the real and imaginary parts contribute equally, and this decreases the variance by a factor of $\sqrt{2}$.) This means that we minimize the vertical error bars by minimizing $\langle\tilde{D}\rangle$. But assuming that our window function is narrow enough that we are measuring mostly what we want to measure, D_{ℓ^*}, the first term in equation (25) satisfies $\langle\psi|D_L|\psi\rangle \approx D_{\ell^*}$, independent of ψ, so we minimize the vertical error bars by simply minimizing the second term, $\langle\psi|V(\widehat{\mathbf{r}})|\psi\rangle$. As a measure of the horizontal error bars, we will use $\Delta\ell$, the r.m.s. deviation of the window function from ℓ^*. With our quantum notation, we have simply $\Delta\ell^2 = \langle\psi|(L-\ell^*)^2|\psi\rangle$. As was discussed above, it is impossible to minimize both error bars at the same time, since there is a trade-off between them. It would be like asking for the best and cheapest car. Instead, the best we can do is minimize some linear combination $E \equiv \langle H\rangle$, where we have defined

$$H \equiv (L-\ell^*)^2 + \gamma V(\widehat{\mathbf{r}}), \tag{26}$$

and the parameter γ specifies how concerned we are about the vertical error bar relative to the horizontal one. Continuing our quantum analogy, we see that we want to find the ψ that minimizes the total "energy", where the "kinetic energy" $(L-\ell^*)^2$ corresponds to the horizontal error bar and the "potential energy" $\gamma V(\widehat{\mathbf{r}})$ corresponds to the vertical error bar. If we for the sake of illustration set $\ell^* = -1/2$, we simply want to minimize $\langle\psi|\mathbf{L}^2 + \gamma V(\widehat{\mathbf{r}})|\psi\rangle$, given the constraint $\langle\psi|\psi\rangle = 1$. Introducing a Lagrange multiplier E, we arrive at the Schrödinger equation

$$[\mathbf{L}^2 + \gamma V(\widehat{\mathbf{r}})]|\psi\rangle = E|\psi\rangle. \tag{27}$$

In other words, we want to find the ground state wavefunction for a particle confined to a sphere with some potential. Numerically solving for various integer values of ℓ^* gives functions with similar behavior, modulated by a wiggling similar to that of the corresponding spherical harmonics.

From our knowledge of quantum mechanics, we can immediately draw a number of conclusions about the solutions, all which turn out to agree well with the exact numerical results.

- $|\psi|^2$ will be small in regions where the noise variance is large, so regions that received little observation time will receive low weights in the analysis.

- Except for the case of complete sky coverage, we will have $\Delta\ell > 0$.

- If incomplete sky coverage confines ψ to a region of the sky whose angular diameter in the narrowest direction is of order $\Delta\theta$, then the uncertainty principle tells us that the minimum $\Delta\ell$ must be at least of order $1/\Delta\theta$.

- This limit on the spectral resolution is independent of ℓ (that this remains approximately true in the real-world case as well is illustrated in Figure 4).

  For a sky map of COBE type, where $\Delta\theta$ is of order a radian given a $20°$ galactic cut, the uncertainty principle thus gives $\Delta\ell \gtrsim 1$. This agrees well with the horizontal error bars actually attained in Figure 1 (bottom).

- If the sky-coverage is incomplete, V is infinite outside of the region covered, and we recover the quantum-mechanical particle-in-a-box problem. From this we know that ψ will always go to zero smoothly as it approaches the survey boundary. This is illustrated in Figure 2 (bottom) by a numerical example.

- This smoothness of ψ is really the gist of the method, as it radically reduces "ringing" in Fourier space, "kinetic energy", without increasing the "potential energy" $\langle\psi|V(\widehat{\mathbf{r}})|\psi\rangle$ much at all. This is seen by comparing figures 2 and 3: whereas the upper and lower weight functions look quite similar in real space (Figure 2), they differ dramatically in the Fourier (multipole) domain (Figure 3).

- If the noise level V is constant wherever we have data, then the potential energy term will reduce to $\gamma\langle\psi|V|\psi\rangle = \gamma V\langle\psi|\psi\rangle = \gamma V$, *i.e.*, become independent of ψ. This means that the solution will be independent of γ. For any evenly sampled data set, the choice of γ is thus irrelevant, so we might as well chose $\gamma = 0$ for simplicity, as we did in our COBE analysis.

- If the survey volume consists of several disconnected parts, then $\Delta\ell$ is limited by the $\Delta\theta$ of the largest part. For the galaxy-cut COBE case, for instance, using only the northern half of the sky gives almost the same $\Delta\ell$ as using both the northern and southern skies combined. (However, including both of course helps reduce the *vertical* error bars.)

6 DISCUSSION

A method for extracting maximal resolution power spectra from CMB sky maps has been presented and applied to the 2 year COBE DMR data.

6.1 COBE results

The power spectrum extracted from the 2 year COBE data in Figure 1 (bottom) is seen to be consistent with the standard $n = 1$, $Q_{rms-ps} = 20$ model. This model is close to the best-fit models found in the various two-parameter Bayesian likelihood analyses published (*e.g.*, G94, B95, BS95, TB95, WB95), which we can interpret as the best-fit straight line through the data points in Figure 1 being close to the horizontal heavy line. As has frequently been pointed out (see *e.g.* WB95), Bayesean methods by their very nature can only make *relative* statements of merit about different models, and never address the question of whether the best-fit model itself is in fact inconsistent with the data. As an absurd example, the best fit straight line to a parabola on the interval $[-1, 1]$ is horizontal, even though this is a terrible fit to the data. It is thus quite reassuring that the power spectrum in Figure 1 (bottom) not only has the right average normalization and slope, but that each and every one of the multipoles appear to be consistent with this standard best fit model.

6.2 How the method differs from other techniques

A number of other linear techniques for CMB analysis have recently been published and applied to the COBE data. Both the Karhunen-Loève (KL) signal-to-noise eigenmode method (B95, BS95) and the orthogonalized spherical harmonics method (G94) are devised to solve a different problem than that addressed here. If one is willing to parametrize the power spectrum by a small number of parameters, for instance a spectral index and an amplitude, then these methods provide an efficient way of estimating these parameters via a likelihood analysis. Why cannot the basis functions of these methods be used to estimate the angular power spectrum C_ℓ directly, as they are after all orthogonal over the galaxy-cut sky? The answer is that these basis functions are orthogonal to *each other*, whereas in our context, we want them to be as orthogonal as possible not to each other but to the *spherical harmonics*. This is illustrated in Figure 3, which contrasts $\ell^* = 20$ window functions of the optimal method and the generalized Hauser-Peebles method

(W94, de Oliveira-Costa & Smoot 1995). We want the window function to be centered on $\ell^* = 20$, and be as narrow as possible, so the lower one is clearly preferable. Equation (9) shows that the window function will vanish for a given ℓ if the weight function is orthogonal to all spherical harmonics with that ℓ-value, so we can interpret Figure 3 as the optimal weight function being essentially orthogonal to all spherical harmonics except $\ell = 18$, $\ell = 20$ and $\ell = 22$ (the reason that $\ell = 19$ and $\ell = 21$ do not cause a problem is that even and odd multipoles remain orthogonal after the galactic cut, since parity symmetry is preserved). The other weight function is seen to couple strongly to many of the lower multipoles, and picks up a contribution from the quadrupole that is even greater than that from $\ell = 20$. This of course renders it quite inappropriate for estimating the power at $\ell = 20$. Analogous window functions can of course be computed for the orthogonalized spherical harmonics of G94 or the signal-to-noise eigenmodes of B95 and BS95, and they also exhibit window functions that are broader than the optimal one in Figure 3 — our derivation of the optimal weight functions of course guarantees that *any* other basis functions will give broader window functions. However, it should be emphasized that generating such window functions for the basis functions of G94, B95 and BS95 would be quite an unfair criticism of these methods, since this would be grading them with respect to a property that they were not designed to have. These authors have never claimed that their basis functions were optimal for multipole estimation, merely (and rightly so) that they were virtually optimal for parameter fitting with a likelihood analysis.

The data points in Figure 1 are placed at the mean ℓ-values obtained from histograms like that in Figure 3, and as mentioned, the horizontal error bars are simply the r.m.s. widths of the histograms about this ℓ-value. The horizontal error bars resulting from the optimal method are contrasted with those obtained by the generalized Hauser-Peebles method in Figure 4. It is seen that the former gives error bars $\Delta\ell \approx 1$, whereas the latter gives a resolution $\Delta\ell$ that increase strongly with ℓ. The new method is seen to more than double the spectral resolution at $\ell = 15$ and more than triple it at $\ell = 20$. At $\ell = 300$, the gain would be about a factor of 100 for a high-resolution all-sky map with a $20°$ galactic cut.

As the low multipoles C_ℓ are intrinsically much larger, histograms like that of Figure 3 exhibit mainly leakage from the left, from lower multipoles. Thus the generalized Hauser-Peebles method probes an effective ℓ-value $\langle\ell\rangle < \ell$ for large ℓ, as pointed out in W94. It is seen that the new method eliminates this "read leak" problem.

There is quite an interesting connection between this method and the KL signal-to-noise eigenmode technique of B95 and BS95. The KL-method is that which optimizes the signal-to-noise ratio, *i.e.*, loosely speaking that which minimizes the vertical error bars without regard for the horizontal ones. Indeed, by taking $\gamma \to \infty$ (which corresponds to dropping the first term in equation (16)) and choosing μ_ℓ to be the expected power spectrum, it is readily seen that the present method reduces to the KL method. In other words, the KL method is a special case of the method presented here, corresponding to ignoring the spectral resolution.

6.3 Qualitative features of the method

Figure 4 can be easily understood from the analogy with quantum mechanics discussed in the previous section. Consider a free particle constrained to the surface of a sphere. If its wavefunction is required to vanish in some region, then it cannot be in an angular momentum eigenstate, and the Heisenberg uncertainty principle tells us that $\Delta\ell$ must be bounded from below by some positive constant $\Delta\ell_*$. Which wavefunction minimizes $\Delta\ell$ given this constraint? Basically, $\mathbf{e}_k$ does. So when the forbidden region is a 20° strip above and below the equator, $\Delta\ell_*$ corresponds to half the height of the double-shaded strip. If the wave-function is abruptly truncated at the edge of this strip (as is the case if we simply use spherical harmonics or orthogonalized versions thereof), this sharp edge causes considerable ringing in Fourier space, which is why the single-shaded region is so much wider. Like the Gaussian minimum-uncertainty states, the optimal set of orthogonal functions $\mathbf{e}_k$ fall off smoothly rather than abruptly as they approach the galactic cut. This difference is readily seen in Figure 2. Although the difference between the naive and optimal weight functions may appear minor in real space (Figure 2), they are seen to be quite radical in the Fourier (multipole) domain (Figure 3).

6.4 Implications for future experiments

The fact that our method produces a spectral resolution $\Delta\ell$ of order $1/\Delta\theta$, independent of ℓ, where $\Delta\theta$ is the smallest survey dimension, has quite encouraging implications for future CMB experiments. It means that a high-resolution satellite experiment covering a square radian on the sky can produce $\Delta\ell \approx 1$ all the way out to the Doppler peaks, thus allowing very precise determination of cosmological parameters. If a long-duration balloon flight (or a satellite map with dusty regions removed) produces a map of a $10° \times 10°$ patch of sky, then $\Delta\ell \approx 6$ and the location of a Doppler peak at say $\ell = 200$ could be pinpointed to an accuracy $\Delta\ell/\ell$ of a few percent. When designing such future CMB experiments, it should be borne in mind that since the *smallest* dimension is what limits the resolution, a square map is always superior to a rectangular one of the same area.

The author wishes to thank Ted Bunn, George Efstathiou, Carlos Frenk, Joseph Silk, George Smoot and an anonymous referee for useful comments on the manuscript. This work was partially supported by European Union contract CHRX-CT93-0120 and Deutsche Forschungsgemeinschaft grant SFB-375. The COBE data sets were developed by the NASA Goddard Space Flight Center under the guidance of the COBE Science Working Group and were provided by the NSSDC.

7 References

Bond, J. R. *et al.* 1994, *Phys. Rev. Lett.*, **72**, 13.
Bond, J. R. 1995, *Phys. Rev. Lett.*, **74**, 4369 ("B95").
Bunn, E.F. & Sugiyama, N. 1995, *ApJ*, **446**, 49 ("BS95").
de Oliveira Costa, A. & Smoot, G.F. 1995, *ApJ*, **448**, 477.
Górski, K. M. 1994, *ApJ*, **430**, L85 ("G94").
Górski, K. M. *et al.* 1995, *ApJ*, **444**, L65.
Hauser, M. G. & Peebles, P. J. E. 1973, *ApJ*, **185**, 757.
Hu, W. & Sugiyama, N. 1995, *Phys. Rev. D*, **51**, 2599.
Lineweaver, C. H. *et al.* 1994, *ApJ*, **436**, 452.
Peebles, P. J. E. 1973, *ApJ*, **185**, 413.
Smoot, G.F. *et al.* 1992, *ApJ*, **396**, L1.
Stompor, R., Górski, K. M. & Banday, A. J. 1995, *MNRAS*, **277**, 1225.
Tegmark, M. & Bunn, E. F. 1995, *ApJ*, **455**, 1 ("TB95").
Tegmark, M. 1995, *ApJ*, **455**, 429 ("T95").
White, M. & Bunn, E. F. 1995, *ApJ*, **450**, 477 ("WB95").
White, M., Scott, D. and Silk, J. 1994, *ARA&A*, **32**, 319.
Wright, E.L. *et al.* 1994a, *ApJ*, **420**, 1.
Wright, E.L. *et al.* 1994b, *ApJ*, **436**, 443 ("W94").

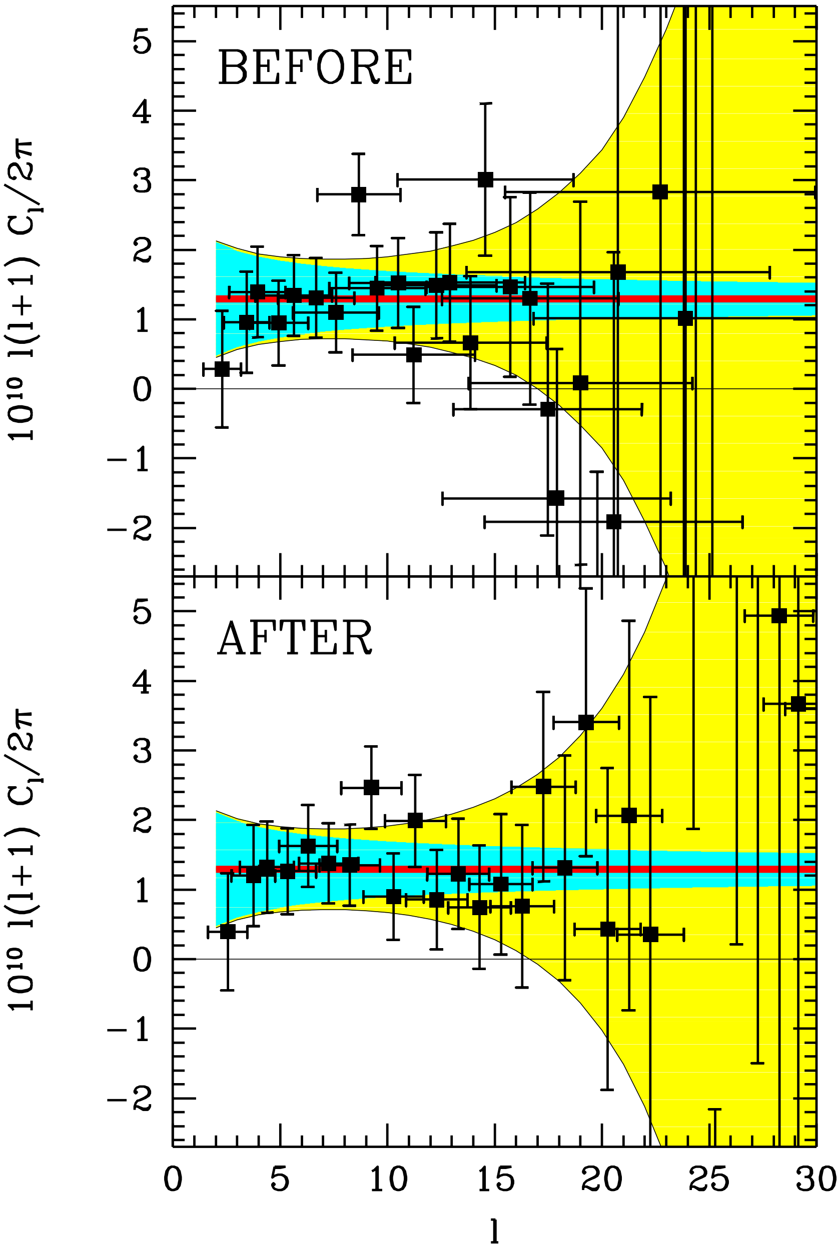

Figure 1: The COBE power spectrum before and after optimization. The observed multipoles $D_\ell = \ell(\ell+1)C_\ell$ are plotted with $1-\sigma$ error bars. The vertical error bars include both pixel noise and cosmic variance, and the horizontal error bars show the width of the window functions used. If the true power spectrum is given by $n = 1$ and $Q_{rms-ps} = 20\mu$K (the heavy horizontal line), then the shaded region gives the $1-\sigma$ error bars and the dark-shaded region shows the contribution from cosmic variance. The multipoles have been estimated by expansion in galaxy-cut spherical harmonics (top) and with the optimal method (bottom).

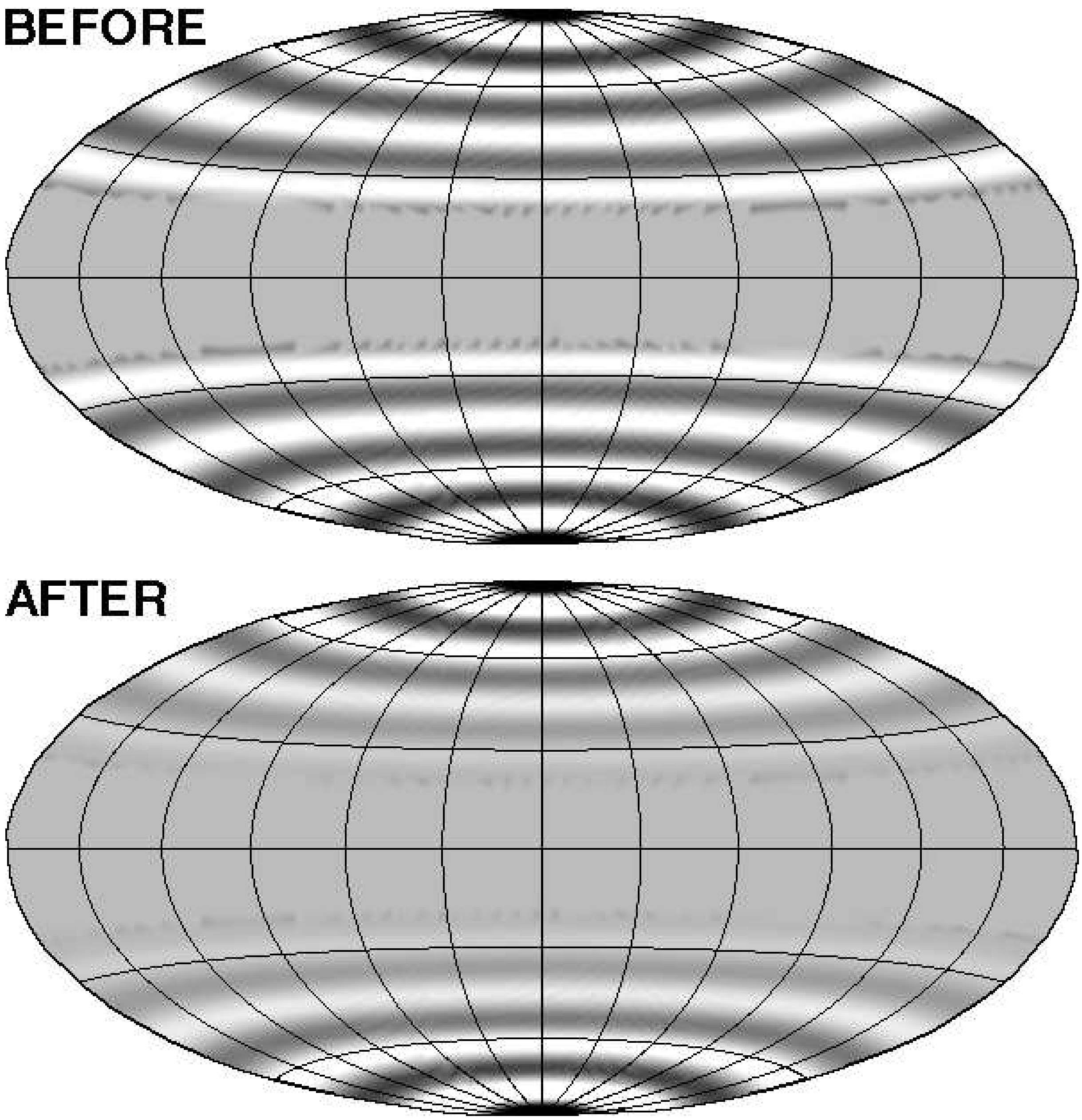

Figure 2: Weight functions before and after optimization.
The weight function $\mathbf{e}$ used to estimate the multipole $\ell = 20$, $m = 0$ is plotted in Hammer-Aitoff projection in galactic coordinates. Black pixels receive large positive weight, white pixels receive large negative weight, and the grey shade of the galactic cut corresponds to zero weight. The functions shown are the relevant galaxy-cut spherical harmonic (top) and the eigenfunction $\mathbf{e}$ with the smallest eigenvalue (bottom).

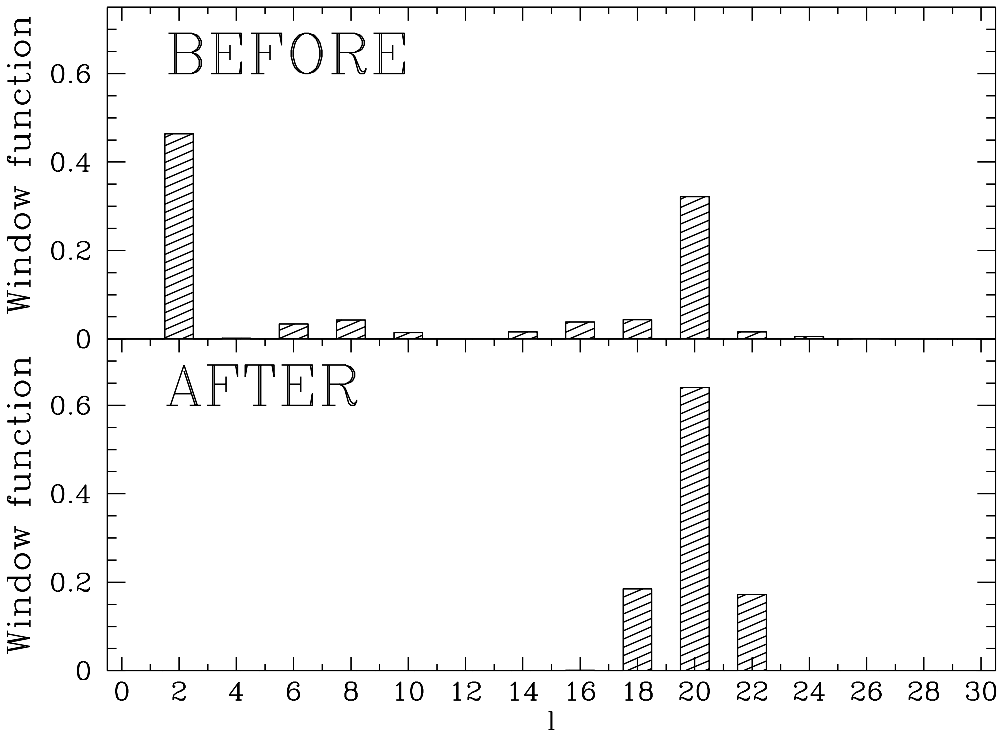

Figure 3: Window functions before and after optimization.
Two window functions for estimation of the multipole $\ell = 20$, $m = 0$ are shown. The upper one is that of the spherical harmonic method (W96), which exhibits a strong leakage from lower multipoles such as the quadrupole. The lower one is the one resulting from the optimal method.

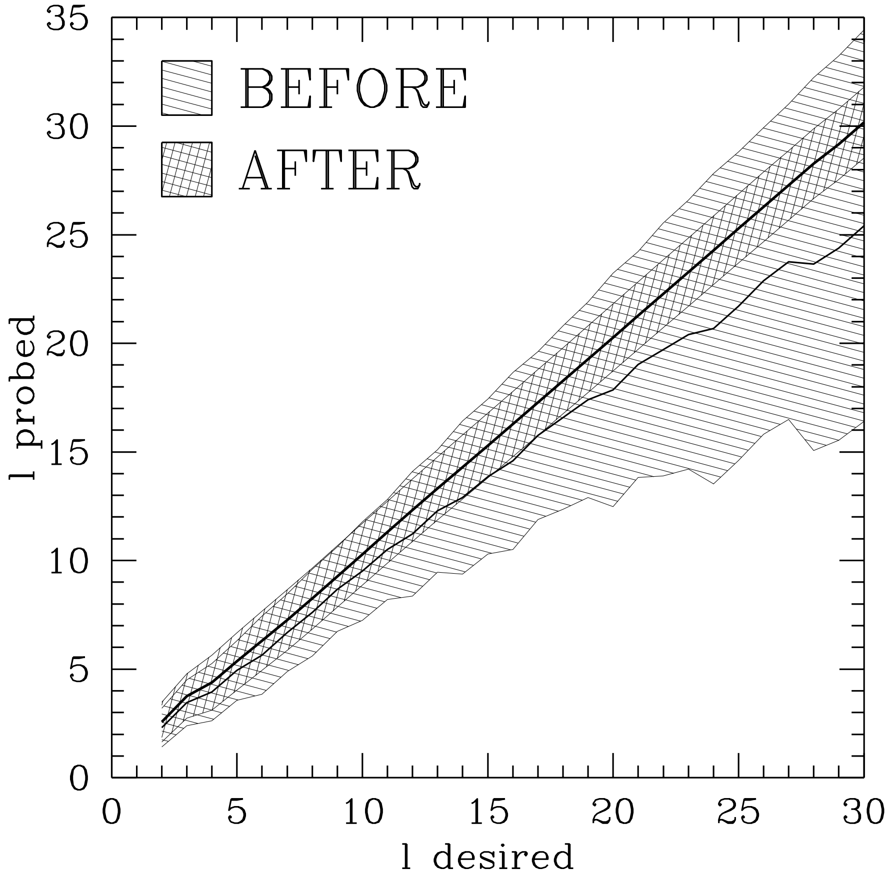

Figure 4: Horizontal error bars before and after optimization.
The shaded region shows the range of ℓ-values $\langle\ell\rangle\pm\Delta\ell$ probed by the window function devised to estimate D_ℓ, the heavy line showing the mean $\langle\ell\rangle$. The wide, jagged region is the result of using spherical harmonics, whereas the narrow strip results from the optimal method described.