OPTIMIZED LARGE-SCALE CMB LIKELIHOOD AND QUADRATIC MAXIMUM LIKELIHOOD POWER SPECTRUM ESTIMATION

E. Gjerløw[1], L. P. L. Colombo[2,3], H. K. Eriksen[1], K. M. Górski[3], A. Gruppuso[4,5],
J. B. Jewell[3], S. Plaszczynski[6], and I. K. Wehus[3]

[1] Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, NO-0315 Oslo, Norway; eirik.gjerlow@astro.uio.no
[2] USC Dana and David Dornsife College of Letters, Arts and Sciences, University of Southern California, University Park Campus, Los Angeles, CA 90089, USA
[3] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA
[4] INAF/IASF Bologna, Via Gobetti 101, Bologna, Italy
[5] INFN, Sezione di Bologna, Via Imero 46, I-40126, Bologna, Italy
[6] Laboratoire de l'Accélérateur Linéaire, Université Paris-Sud 11, CNRS/IN2P3, Orsay, France
Received 2015 June 11; accepted 2015 September 9; published 2015 October 23

ABSTRACT

We revisit the problem of exact cosmic microwave background (CMB) likelihood and power spectrum estimation with the goal of minimizing computational costs through linear compression. This idea was originally proposed for CMB purposes by Tegmark et al., and here we develop it into a fully functioning computational framework for large-scale polarization analysis, adopting *WMAP* as a working example. We compare five different linear bases (pixel space, harmonic space, noise covariance eigenvectors, signal-to-noise covariance eigenvectors, and signal-plus-noise covariance eigenvectors) in terms of compression efficiency, and find that the computationally most efficient basis is the signal-to-noise eigenvector basis, which is closely related to the Karhunen–Loeve and Principal Component transforms, in agreement with previous suggestions. For this basis, the information in 6836 unmasked *WMAP* sky map pixels can be compressed into a smaller set of 3102 modes, with a maximum error increase of any single multipole of 3.8% at $\ell \leqslant 32$ and a maximum shift in the mean values of a joint distribution of an amplitude–tilt model of 0.006σ. This compression reduces the computational cost of a single likelihood evaluation by a factor of 5, from 38 to 7.5 CPU seconds, and it also results in a more robust likelihood by implicitly regularizing nearly degenerate modes. Finally, we use the same compression framework to formulate a numerically stable and computationally efficient variation of the Quadratic Maximum Likelihood implementation, which requires less than 3 GB of memory and 2 CPU minutes per iteration for $\ell \leqslant 32$, rendering low-ℓ QML CMB power spectrum analysis fully tractable on a standard laptop.

Key words: cosmic background radiation – cosmology: observations – methods: statistical

1. INTRODUCTION

Through a series of increasingly sensitive experiments measuring the cosmic microwave background (CMB) led by COBE (Mather et al. 1990), *WMAP* (Bennett et al. 2013), and Planck (Planck Collaboration et al. 2014a), during the last two decades, cosmologists have established a successful cosmological concordance model (Planck Collaboration et al. 2014c). According to this model, the universe is isotropic and homogeneous, and filled with Gaussian random fluctuations drawn from a nearly scale-invariant primordial power spectrum; its energy budget is composed of 68% dark energy, 27% dark matter, and 5% baryonic matter. Remarkably, only six or seven parameters are required to model accurately millions of data points.

The connection between those millions of data points and the handful of cosmological parameters is made through the so-called likelihood function, and cosmological parameter estimation essentially amounts to mapping out this function, for instance, using Markov Chain Monte Carlo (Lewis & Bridle 2002), multi-dimensional gridding (Mikkelsen et al. 2013), or nonlinear optimization (Planck Collaboration et al. 2014d). Since the CMB fluctuations are observed to be (at least close to) Gaussian distributed, the analytic expression for the likelihood is formally given by a multivariate Gaussian. However, this expression is of limited practical use for modern CMB experiments because of the high dimensionality of the associated covariance matrix. For Planck, the number of pixels is $N_{\rm pix} \sim 5 \times 10^7$, and since brute-force likelihood evaluation requires a Cholesky decomposition of this matrix, computationally scaling as $\mathcal{O}(N_{\rm pix}^3)$, a single evaluation would cost $\sim 10^6$ CPU years and require 10^4 TB RAM (see, e.g., Borrill 1999, for a related discussion).

Obviously, the direct brute-force likelihood approach is not feasible for modern full-sky CMB experiments, and a few alternative methods have therefore been proposed and implemented in the literature. These can largely be broken into two groups. First, the most widely adopted approach is that of a hybrid likelihood, which simply splits the full likelihood into two components according to angular scales. Large angular scales are analyzed using some exact method that fully accounts for the non-Gaussian nature of the likelihood, whereas small angular scales are analyzed using faster likelihood approximations motivated by the Central Value theorem. Usually, the two likelihoods are sufficiently uncorrelated that they may be joined into a single all-scale expression either by straight multiplication or by explicitly accounting for overlap correlations (Gjerløw et al. 2013). The second group of methods may be characterized as samplers, for instance, as implemented through Gibbs sampling (Eriksen et al. 2004; Jewell et al. 2004; Wandelt et al. 2004), which draws samples from the CMB posterior. The computational scaling of this approach is $\mathcal{O}(N_{\rm pix}^{3/2})$, and therefore is computationally feasible even for high-resolution data. However, further work is required for this potential to be fully realized because the computational expense is still considerable (Seljebotn

et al. 2014) and, in practice, samplers are still mostly used on large and intermediate angular scales (Planck Collaboration et al. 2014b).

Although one could argue that the ever-advancing progress of computer technology lessens the need for clever likelihood approximations, one could at the same time argue that reducing the time needed to perform likelihood evaluations allows us to expand our field of interest. As an example, there is little to gain in terms of computational time if we restrict our interest to the standard six-parameter ΛCDM model, which can presently be tackled by a standard laptop in a comfortable time frame. Presently, however, considering extensions to this model is gaining more and more interest. Such extensions expand the parameter space of interest, which reintroduces the need to make our likelihood evaluations as fast as possible while still being reasonably accurate.

In this paper, we revisit the problem of exact brute-force likelihood evaluation on large angular scales, exploiting the ideas initially introduced for CMB analysis purposes by Tegmark et al. (1997) to reduce the computational cost through linear compression. Rather than crudely downgrading the data in pixel space until the computational costs are acceptable, we compress the data into a lower-dimensional basis set using a more general linear transformation, thereby reducing computational costs while retaining most of the important information. Furthermore, we also show how this formalism naturally leads to a very efficient implementation of the Quadratic Maximum Likelihood (QML) power spectrum estimator.

2. BASIC DEFINITIONS

In their most basic form, CMB observations may be modeled[7] as a linear sum of a cosmological CMB signal, s, observed by some instrumental beam convolution operator, B, some set of foreground contaminants, f, and random noise, n,

$$d = Bs + f + n. \qquad (1)$$

The signal and noise terms are usually both assumed to be Gaussian distributed with zero-mean and covariances $S = B\langle ss^t\rangle B^t$ and $N = \langle nn^t\rangle$, respectively, and the total data covariance matrix is $C = S + N$, neglecting for the moment the foreground term.

In most cases, the CMB signal is assumed to be isotropic, and it is therefore especially convenient to expand this component into spherical harmonics:

$$s = \sum_{\ell=0}^{\ell_{\max}} \sum_{m=-\ell}^{\ell} s_{\ell m} Y_{\ell m} \equiv Y\tilde{s}. \qquad (2)$$

Here we have defined Y as a matrix listing all of the spherical harmonics (both spin-0 for temperature and spin-2 for polarization; see Zaldarriaga & Seljak 1997 for details) up to some maximum band limit, $\ell_{\max}$ column-wise, and $\tilde{s}$ to be a vector containing the spherical harmonics coefficients of s. We additionally define the symbol Y^{-1} to denote the inverse spherical harmonic transform,

$$\tilde{s} = \int_{4\pi} Y_{\ell m}^* s \, d\Omega \equiv Y^{-1}s, \qquad (3)$$

but emphasize that this is not a true inverse of Y, as neither Y nor Y^{-1} is square, and any spherical pixelization introduce non-orthogonality between modes on small angular scales; we only use Y^{-1} for high-ℓ mode filtering in the combination $P = YY^{-1}$ in this paper, for which exact orthogonality is not required.

Under the assumption of statistical isotropy, the signal covariance takes on a particularly simple form in spherical harmonic space, and is given by the angular power spectrum, C_ℓ. Assuming further that the instrumental beam is circularly symmetric and fully described by a set of Legendre coefficients, b_ℓ, and that any smoothing effects from discrete pixelization may be described in terms of an effective pixel window function, p_ℓ, the harmonic space elements of the signal covariance matrix read

$$\tilde{S}_{\ell m, \ell' m'} \equiv b_\ell p_\ell \left\langle s_{\ell m} s_{\ell' m'}^* \right\rangle p_{\ell'} b_{\ell'} = C_\ell b_\ell^2 p_\ell^2 \, \delta_{\ell\ell'} \delta_{mm'}, \qquad (4)$$

where, for simplicity, we have defined the power spectrum coefficient, C_ℓ, to denote a 3×3 block incorporating all of the temperature and polarization auto- and cross-spectra:

$$C_\ell = \begin{pmatrix} C_\ell^{TT} & C_\ell^{TE} & C_\ell^{TB} \\ C_\ell^{TE} & C_\ell^{EE} & C_\ell^{EB} \\ C_\ell^{TB} & C_\ell^{EB} & C_\ell^{BB} \end{pmatrix}. \qquad (5)$$

From Equation (2) we see that the corresponding signal covariance matrix in the map domain simply reads $S = Y\tilde{S}Y^T$, and it is easy to show that the entries of this matrix are given by the two-point correlation function.

Properly accounting for the foreground term is a far more complicated problem, and extensive literature has been written on this topic (e.g., Leach et al. 2008; Planck Collaboration et al. 2015, and references therein). In this paper, we limit ourselves to a very basic foreground model in which f may be described by a finite set of spatial templates, each known perfectly up to an overall amplitude:

$$f = \sum_i a_i t_i = Ta, \qquad (6)$$

where T is a matrix listing all of the templates column-wise and a is a vector of template amplitudes. Accounting for such templates is most easily implemented by solving the normal equations for a, and redefining the data vector and data covariance matrix as follows:

$$d \leftarrow d - T\left(T^t(S+N)^{-1}T\right)^{-1}T^t(S+N)^{-1}d \qquad (7)$$

$$N \leftarrow N + \alpha TT^t; \qquad (8)$$

here, α is a parameter that estimates the uncertainty in the template fit and $\alpha \to \infty$ corresponds to full projection. However, from a numerical point of view, it is more convenient to set α to a large numerical value to avoid an otherwise singular covariance matrix. In this paper, we let T consist of the monopole and three dipoles, all normalized to a maximum of unity, and let $\alpha = 10^3$.

With the above data model, the data likelihood depends only on the angular power spectrum, and is given by a multivariate

Gaussian:

$$\mathcal{L}(C_\ell \,|\, \boldsymbol{d}) \equiv P\big(\boldsymbol{d} \,\big|\, C_\ell\big) \propto \frac{e^{-\frac{1}{2}\boldsymbol{d}^T\left(\boldsymbol{S}(C_\ell)+\boldsymbol{N}\right)^{-1}\boldsymbol{d}}}{\sqrt{\left|\boldsymbol{S}(C_\ell)+\boldsymbol{N}\right|}}, \qquad (9)$$

where we have implicitly accounted for template marginalization by the redefinitions in Equations (7) and (8). In principle, this expression can be used directly for CMB power spectrum or cosmological parameter estimation when coupled with some nonlinear optimization or MCMC implementation. However, as already noted, this expression contains both a matrix inverse and a determinant, and therefore scales computationally as $\mathcal{O}(N_{\rm pix}^3)$. Direct likelihood evaluations are therefore computationally very expensive, and the main goal of this paper is to speed up this expression simply by reducing the effective number of pixels.

3. THE 9-YEAR *WMAP* LOW-ℓ LIKELIHOOD

For pedagogical purposes, we specialize the discussion in this paper to the low-ℓ *WMAP* likelihood,[8] as presented by Hinshaw et al. (2013). However, we note that the same approach should be fully applicable to corresponding *Planck* low-ℓ polarization observations once available.

The 9-year *WMAP* low-ℓ likelihood function is implemented as a hybrid between a pure temperature likelihood using a Blackwell–Rao estimator (Chu et al. 2005), and a pure polarization brute-force likelihood, similar to that described in Equation (9). Correlations between the two are handled by explicitly decorrelating the temperature component from the Stokes Q and U maps, given some fixed estimate of the full-sky temperature sky map and C_ℓ^{TE} (Page et al. 2007). For computational speed, the polarization data are degraded onto a very low-resolution grid, defined by the HEALPix[9] pixelization with a resolution parameter of $N_{\rm side} = 8$. This pixelization has a pixel size of $7° \times 7°$ and reliably supports harmonic modes only up to $\ell_{\rm max} = 16$, although the *WMAP* likelihood implementation formally includes modes up to $\ell = 23$. After applying a Galactic mask removing contaminated pixels, a total of 1100 low-resolution polarization (Q and U) pixels are included in the likelihood.

This approach leads to a fast and flexible low-ℓ likelihood. However, several assumptions have been made in the process, most notably that the temperature noise is fully negligible (enabling the temperature–polarization split) and that the full-sky temperature modes are well described by the *WMAP* ILC map. Neither of these assumptions are obvious (see, e.g., Finelli et al. 2013 for a relevant discussion), and in particular the assumption of no temperature noise has significant consequences in terms of the effective prior of the likelihood. An absolute mathematical requirement for any likelihood is that the total covariance matrix, $\boldsymbol{S} + \boldsymbol{N}$, be positive definite, while a softer physical requirement is that the signal covariance $\boldsymbol{S}$ alone be positive definite. Enforcing these requirements consistently is not trivial with a split likelihood, and we will see in Section 7 that the *WMAP* likelihood has nonphysical "holes" as a result of this, as well as a generally complicated behavior near the singularity regions.

A second issue with the *WMAP* likelihood implementation lies in its resource requirements. To accelerate the likelihood evaluations, the *WMAP* code precomputes the Legendre polynomials for each pair of pixels, thereby saving CPU time for building the signal covariance matrix. However, this is costly in terms of memory, and requires 1 GB RAM already at $N_{\rm side} = 8$, which only supports $\ell \lesssim 16$. Doubling the resolution in order to probe scales up to $\ell \lesssim 32$ increases this requirement to 33 GB, which is more than most computers can handle comfortably today.

In this paper, we present a more direct implementation of a low-ℓ *WMAP* likelihood that relies only on the brute-force likelihood expression in Equation (9). Both temperature and polarization sky maps are considered at a common resolution parameter of $N_{\rm side} = 16$. Otherwise, we adopt data combinations that are as close as possible to those used for the official *WMAP* likelihood (Hinshaw et al. 2013). Specifically, for polarization, we include only the foreground-reduced *WMAP* Ka, Q, and V bands in the following, not the K band, which is used for foreground cleaning, or the W band, which is known to have more poorly correlated noise and/or systematics issues than the other frequencies. For the temperature component, we adopt the 9-year *WMAP* ILC map, smoothed to $10°$ FWHM.

The individual foreground-reduced polarization frequency maps are co-added into a single "clean" CMB map by inverse noise variance weighting,

$$\boldsymbol{d} = \left(\sum_i \boldsymbol{N}_i^{-1}\right)^{-1} \sum_i \boldsymbol{N}_i \boldsymbol{d}_i, \qquad (10)$$

where $\boldsymbol{N}_i$ is the full covariance matrix for band i, which also take into account the additional noise contribution from the foreground reduction. The noise covariance of the co-added map reads

$$\boldsymbol{N} = \left(\sum_i \boldsymbol{N}_i^{-1}\right)^{-1}. \qquad (11)$$

Finally, we add $2\,\mu{\rm K}$ regularization noise to the ILC temperature map to make the temperature covariance matrix invertible. The full noise covariance thus consists of a diagonal temperature block and a dense polarization block, with no cross-terms between the two.

We adopt the *WMAP* KQ85 mask for the temperature component and the P06 mask for the polarization components (Bennett et al. 2013), leaving a total of 2326 temperature and 4510 polarization (Q and U) pixels for analysis, or a total of 6836 elements in the data vector. The instrumental beams are taken to be a perfect Gaussian of $10°$ FWHM for the temperature, and a Gaussian of 30.6 arcmin for polarization, corresponding roughly to the Q-band beam, and adopted as a rough average of the three channels; its impact is very small for the multipoles considered in the following, with a minimum amplitude of $b_\ell = 0.993$ at $\ell = 30$.

4. LINEAR COMPRESSION AND BASIS DEFINITIONS

4.1. Basic Formalism

Any linear transformation of a set of Gaussian random variables results in another set of Gaussian random variables.

Therefore, we consider some linear combination of the form

$$\bar{d} = Pd, \quad (12)$$

where P is some $N \times N_{\mathrm{pix}}$ transformation matrix with $N \leqslant N_{\mathrm{pix}}$, and $\bar{d}$ is a transformed data vector. If d is a zero-mean Gaussian field with covariance C, then $\bar{d}$ will be a zero-mean Gaussian field with covariance $\bar{C} = PCP^t$. Using the data model described above, the corresponding likelihood for these compressed data therefore reads

$$\mathcal{L}\left(C_\ell | \bar{d}\right) \propto \frac{e^{-\frac{1}{2}\bar{d}^T\left(\bar{S}(C_\ell)+\bar{N}\right)^{-1}\bar{d}}}{\sqrt{\left|\bar{S}(C_\ell)+\bar{N}\right|}}$$
$$= \frac{e^{-\frac{1}{2}P^T d^T\left(PS(C_\ell)P^T+PNP^T\right)^{-1}Pd}}{\sqrt{\left|PS(C_\ell)P^T+PNP^T\right|}}. \quad (13)$$

The interesting question is now whether or not there exists some transformation P that retains the relevant information in d with a smaller number of data points, $N < N_{\mathrm{pix}}$.

Before explicitly defining a set of candidate bases, it is useful to introduce some additional notation. First, since there are always parts of the sky that are unavailable for cosmological CMB analysis due to foreground contamination from our own Galaxy, we introduce a pixel space masking operator, M, defined in terms of an $N_{\mathrm{mask}} \times N_{\mathrm{pix}}$ matrix that contains one row for each unmasked pixel, with a value of 1 in the column corresponding to the pixel number; all other entries are zero. When applied to full-sky data vectors, this operator simply picks out the unmasked pixels, leaving all of the values numerically unchanged.

Second, we define a harmonic space truncation operator, $P_h \equiv MY_{\ell_t}Y_{\ell_t}^{-1}M^T$, where only spherical harmonics up to some truncation multipole, $\ell_t \leqslant \ell_{\mathrm{max}}$, are included in the spherical harmonics operator. This operator filters out any spherical harmonics above ℓ_t, evaluated only over masked pixels; since only masked pixels are included, the operator is not a sharp operator in multipole space, but rather corresponds to a pseudo-$a_{\ell m}$ projection operator with non-zero coupling to multipoles above ℓ_t (e.g., Hivon et al. 2002).

Third, we define $[A]_\epsilon$ as the set of eigenvectors of A with a fractional eigenvalue larger than ϵ relative to the maximum eigenvalue. That is, let V be the matrix containing the eigenvectors of A, and W be the diagonal matrix of eigenvalues, such that $A = VWV^t$; then, $[A]_\epsilon$ contains all columns of V with an eigenvalue larger than $\epsilon \cdot \mathrm{max}(W)$. This operator removes those modes that have low eigenvalues, which in turn will be used to eliminate modes with a low signal-to-noise ratio. However, due to the very different signal amplitudes in temperature and polarization, we define two different eigenvalue thresholds, ϵ_T and ϵ_P, for the temperature and polarization modes, and set the temperature–polarization cross-elements in A to zero before performing the eigenvalue decomposition.

4.2. Basis Sets

Using the above notation, we define five candidate bases to be considered for further analysis:

$P_1 = M$	Pixel
$P_2 = \left[P_h\right]_\epsilon M$	Harmonic
$P_3 = \left[P_h N^{-1}P_h^t\right]_\epsilon M$	Inverse-noise
$P_4 = \left[P_h(S+N)P_h^t\right]_\epsilon M$	Signal-plus-noise
$P_5 = \left[P_h\left(S^{1/2}N^{-1}S^{1/2}\right)P_h^t\right]_\epsilon M$	Signal-to-noise,

where $S = S(C_\ell^{\mathrm{fid}})$ is the signal covariance matrix computed from some fiducial model.[10] Each basis is either commonly encountered in the literature (i.e., pixels, harmonics, signal-to-noise eigenmodes) or has a well-defined, specific purpose (e.g., the inverse noise basis is particularly well suited to test systematics by suppressing poorly measured modes, while the signal-plus-noise basis corresponds to numerical regularization of the data covariance matrix). It is also worth noting that the signal-to-noise basis is closely related to the Karhunen–Loeve (or Principal Component) transform originally proposed for cosmological applications by Tegmark et al. (1997). A potential dependence on the assumed fiducial spectrum, C_ℓ^{fid}, is considered in Section 7; we find no significant detrimental effects by adopting a power spectrum far from the best-fit spectrum.

There are two tunable parameters in this framework, ℓ_t and ϵ, both of which have a very intuitive interpretation: lowering ℓ_t removes high-ℓ spherical harmonic modes, while increasing ϵ removes low signal-to-noise modes. However, it is important to note that no choice of either ℓ_t or ϵ can ever *bias* the power spectrum, but only modify the uncertainties. Linear compression simply amounts to removing irrelevant modes, and is mathematically fully equivalent to removing masked pixels. However, lowering ℓ_{max} (as opposed to ℓ_t) will both bias the power spectrum and increase χ^2 because it changes the data model, not simply the data selection. This is an important difference between our approach and that implemented by the official *WMAP* polarization likelihood code, which simply downgrades the actual sky maps from $N_{\mathrm{side}} = 16$ to 8.

Before proceeding with basis optimization, it is useful to make some intuitions about the various basis candidates. Therefore, in Figure 1, we show an example basis vector, and in Figure 2 we show the eigenvalue spectrum for each basis as computed from the 9-year *WMAP* data (Section 3). The example basis vectors all correspond to the vector with the thirtieth largest eigenvalue, ϵ_{30}, for both temperature and polarization. Only the Stokes Q field is shown for polarization, as Stokes U looks qualitatively similar.

Starting with the pixel basis, in Figure 1, we see that, in this case, each pixel corresponds to an independent basis vector. Furthermore, as seen in Figure 2, the eigenspectrum is completely flat, and no truncation limit, ϵ, can remove any degrees of freedom. The pixel basis is therefore always complete and all information stored in the uncompressed data is (by definition) retained in this basis. In the following, we adopt the pixel basis as the reference against which we measure data loss for other bases.

The second row in Figure 1 shows a temperature and polarization mode of the spherical harmonics basis, and the

---

[10] We set $C_\ell^{BB,\mathrm{fid}} = C_\ell^{EE,\mathrm{fid}}$ when constructing the basis signal covariance matrix in the following analyses to ensure good sampling of both spectra.
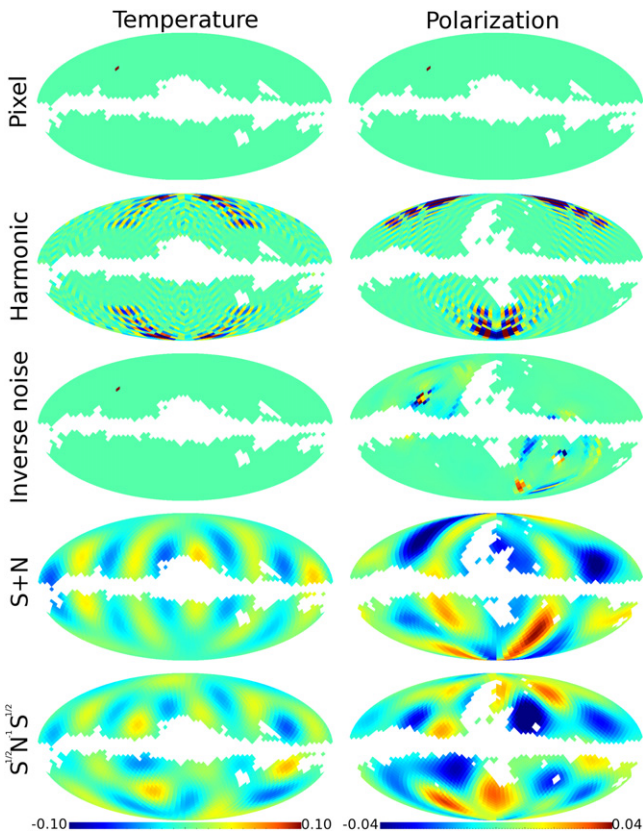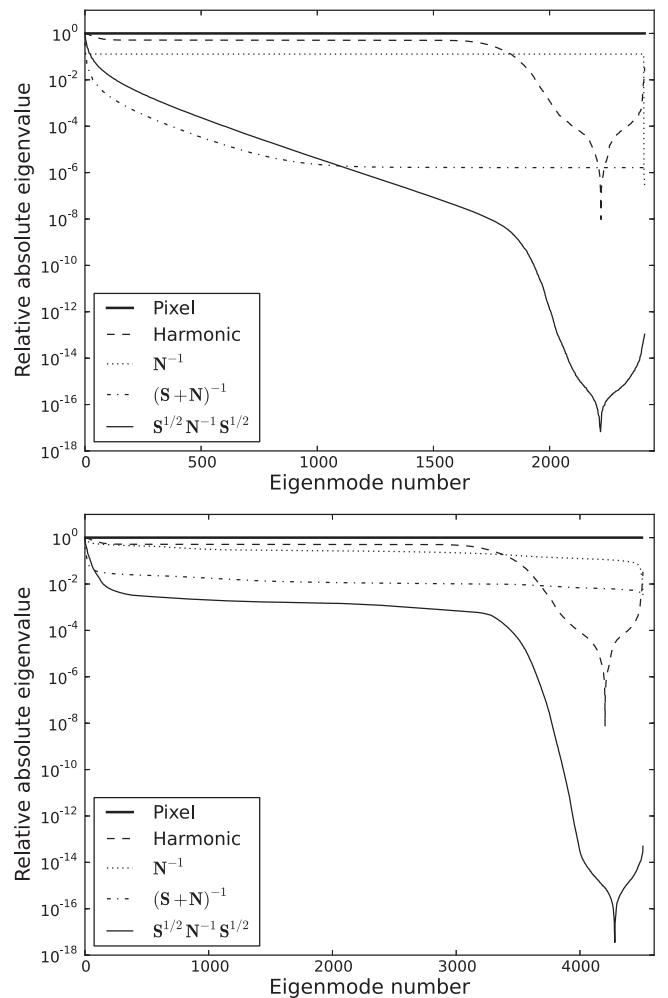
Figure 1. Example basis vectors for temperature (left column) and polarization (right column) for each of the five basis sets considered in this paper, computed from the 9-year *WMAP* data. In each case, the basis vector with the thirtieth highest eigenvalue is shown, and only the Stokes Q component is shown for polarization; the Stokes U components look qualitatively similar.

Figure 2. Eigenvalue spectra for all five bases defined in the text, shown both for temperature (top) and polarization (bottom), as evaluated for the 9-year *WMAP* data, using no high-ℓ truncation. The eigenvalues are normalized to the maximum value and sorted according to decreasing values; increasing values from left to right indicate negative eigenvalues.

dashed line in Figure 2 shows its eigenspectrum. Both of these highlight a problematic feature of this particular basis: it is susceptible to numerical errors at high multipoles. Ideally, $Y_{\ell_t} Y_{\ell_t}^{-1}$ should be identically equal to one for $\ell \leqslant \ell_t$, and zero otherwise. However, because all of the operations are performed on a finite pixelization that supports only a finite number of multipoles, there is always some leakage between multipoles in this operator. Furthermore, there is no guarantee that $\int Y_{\ell m} Y_{\ell m}^* d\Omega$ will be smaller than one. On the contrary, the worst-behaved modes often have a square-integral substantially larger than one. This is observed as three distinct regions in the eigenspectrum of the spherical harmonics basis: the flat plateau from about 50 to 1500 corresponds to well-resolved modes with good support on the masked sky; the rapid decrease above 1500 corresponds to modes that are filtered either by the high-ℓ truncation operator or are degenerate because of the sky mask;[11] and, finally, the modes below 50 are numerically unstable high-ℓ modes with an eigenvalue larger than 1. The harmonic mode shown in Figure 1 is an example of such a mode.

The third basis corresponds to the eigenvectors of the inverse noise covariance matrix. For the low-resolution *WMAP* data, this matrix is given by a spatially constant regularization noise RMS amplitude of $2\,\mu$K for temperature and the actually measured instrumental noise covariance for polarization,

including both scanning strategy and correlated noise effects. For temperature, the inverse noise basis functions are therefore identical to the pixel basis, with one pixel per value, with one exception: this basis explicitly highlights the effect of foreground template projection in the form of a sharp drop in the eigenspectrum, corresponding to the monopole and dipole modes which are manually assigned a numerically large uncertainty. For polarization, the dominant feature is the scanning strategy, which can clearly be seen in the example basis mode in Figure 1. This basis may be useful for systematics studies, since instrumental systematics are often strongly associated with poorly measured modes.

The fourth basis is defined as the eigenvectors of the total data covariance matrix, $S + N$. This could be a relevant basis for cases that have an ill-conditioned covariance matrix, as often happens for strongly signal-dominated temperature data, $S \gg N$, for instance. Since S, by construction, is spanned by $(\ell_{max} + 1)^2 < N_{pix}$ modes, this situation leads to a poorly conditioned total covariance matrix that needs to be regularized before further analysis. The two most common approaches are either to add a small amount of white noise to the data (known as "regularization noise") or to increase ℓ_{max} beyond the
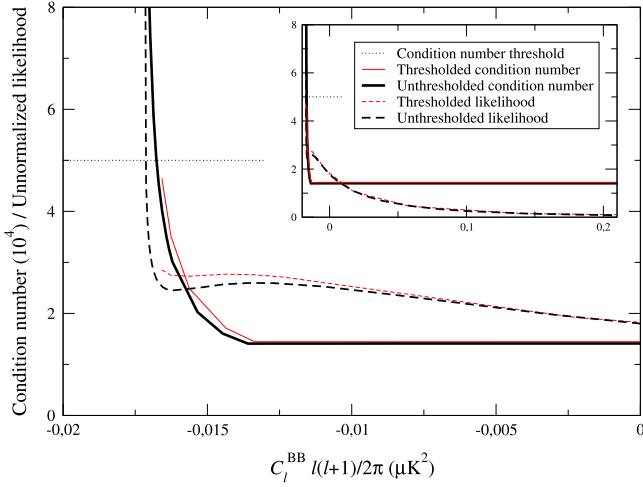
---

[11] Note that the absolute value of the eigenvalue is plotted here; the sharp feature indicates the mode for which the eigenvalue becomes negative.

Figure 3. Regularizing the likelihood with a condition number prior. Dashed curves show slices through $\mathcal{L}(C_2^{BB})$ while fixing all the other multipoles at their maximum-likelihood points. Solid lines show the condition number of the data covariance matrix, $S + N$, as a function of the power spectrum. Thick curves show results with no prior on the condition number, and thin curves show results when requiring the condition number to be smaller than 50,000. Condition number regularization eliminates likelihood artifacts near the singularity boundary.

Nyquist limit formally supported by the pixelization. A third option would be to use the $S + N$ basis proposed here, which simply removes by hand poorly conditioned modes from the data set.

Finally, the fifth basis is given by the eigenvectors of the signal-to-noise covariance matrix, $S^{1/2}N^{-1}S^{1/2}$, written in an explicitly symmetric form to minimize numerical errors. In this case, a prior spectrum is introduced that allows one to select modes based on individual signal-to-noise ratios, only retaining those which actually contribute useful information. Several variations of this have already been discussed extensively in the literature, resulting in various implementations of the same underlying ideas, two of which are the Karhunen–Loève and Principal Component transforms. This was also the basis set originally proposed by Tegmark (1997). For a related application to non-Gaussianity, see Rocha et al. (2001).

4.3. A Condition Number-based Prior

Before proceeding to further analysis, we comment briefly on a technical issue already mentioned in Section 3, namely, that the total data covariance matrix, C, must be positive definite in order for the likelihood to be well defined. Intuitively, and not mathematically rigorously, this condition breaks down whenever $S_{\ell m,\ell m} < -N_{\ell m,\ell m}$. However, the likelihood surface actually becomes unstable well before this limit due to numerical errors, as illustrated in Figure 3. The only difference between the main frame and the inset is the x-axis range. First, the thick dashed line shows a slice (arbitrarily normalized) through our pixel basis likelihood for $\mathcal{L}(C_2^{BB})$, keeping all of the other multipoles fixed at their maximum-likelihood values. This slice exhibits perfectly normal behavior for large values of C_2^{BB}, following roughly the behavior of an inverse Gamma distribution. However, near the value of $C_2^{BB} = -0.0175\,\mu K^2$, the likelihood rapidly increases and essentially diverges to infinity.

This behavior is a generic feature of any likelihood near the boundary at which it becomes singular: even if the matrix may

be positive definite and invertible, the numerical value cannot be trusted sufficiently near the singularity boundary. Fortunately, this problem can be resolved in several ways, and our preferred solution is to monitor the covariance matrix condition number, i.e., the ratio of the largest to smallest eigenvalue. This quantity is shown as a solid thick line in Figure 3 for the above case. For any power spectrum value not close to the singularity boundary, we see that the condition number is highly stable, with a numerical value around 15,000 for this particular case. However, once the covariance matrix approaches singularity, it starts to increase rapidly and does so sooner than the actual likelihood. The combination of a high degree of stability within the main parameter volume and a rapid increase toward the edges makes the condition number an effective monitor of the likelihood robustness. Therefore, rather than requiring that the covariance matrix simply be positive definite (for instance, as is enforced by the official *WMAP* likelihood), we demand that the condition number must be smaller than some pre-defined threshold. The specific value of this threshold must be determined by some initial likelihood scans, but in practice this is very straightforward. For the above basis, we adopt a numerical threshold of 50,000, and the resulting regularized likelihood is shown as a dashed thin line. The slight difference with respect to the unregularized likelihood at higher values is due to the slightly different maximum-likelihood power spectrum coefficients at other multipoles caused by the same prior.

5. EFFICIENT AND STABLE QML IMPLEMENTATION

The formalism described in Section 4 can be used to derive a computationally efficient variation of the QML estimator, initially introduced by Tegmark (1997) and Bond et al. (1998) as an efficient route to the maximum-likelihood CMB power spectrum. For example applications, see, e.g., Gruppuso et al. (2009, 2011, 2013) and references therein. Let $C_{,b} = \partial C / \partial C_b$ denote the derivative of the data covariance matrix with respect to some power spectrum parameter, C_b, with $b = \{\ell_1, ..., \ell_n\}$, where n denotes the number of multipoles included in the spectral bin. The first derivative and the Fisher matrix of the log-likelihood may then be written as

$$\frac{\partial \ln \mathcal{L}}{\partial C_b} = \frac{1}{2}\,\mathrm{tr}\Big[\big(\bar{d}\bar{d}^t - \bar{C}\big)\big(\bar{C}^{-1}\bar{C}_{,b}\bar{C}^{-1}\big)\Big] \qquad (14)$$

$$F_{bb'} = \frac{1}{2}\,\mathrm{tr}\Big[\bar{C}^{-1}\bar{C}_{,b}\bar{C}^{-1}\bar{C}_{,b'}\Big]; \qquad (15)$$

see Section IIC of Bond et al. (1998) for full details.

The QML estimator is now defined as follows.

1. Make some initial guess at the power spectrum, $C_b^{(0)}$.
2. Update the spectrum according to the following rule:

$$C_b^{(i)} = C_b^{(i-1)} + \sum_{b'}\big(F^{-1}\big)_{bb'}\frac{\partial \ln \mathcal{L}}{\partial C_{b'}}. \qquad (16)$$

3. Iterate until convergence.

This algorithm is closely related to the Newton-Raphson optimization method, with the one difference being that it employs the (computationally cheaper) Fisher matrix instead of the curvature matrix. The two algorithms converge to the same (maximum-likelihood) solution (Bond et al. 1998).

In this paper, we note that Equations (14) and (15) can be slightly rewritten to facilitate fast numerical evaluation.

Specifically, the signal matrix may be written as $\bar{C} = PY\tilde{S}Y^{\dagger}P^{t} + \bar{N}$, where $\tilde{S}$ is the *full-sky* signal covariance matrix in harmonic space, and all geometry and data selection effects are encoded in the constant projection operators P and Y. The derivative of this matrix with respect to C_b reads

$$\frac{\partial C}{\partial C_b} = PYI_bY^{\dagger}P^{t}, \tag{17}$$

where I_b is a harmonic space matrix containing the value 1 for entries containing C_{ℓ} in $\tilde{S}$ for $\ell \in b$, and otherwise 0; it is very sparse, and multiplication with this matrix is fast.

Inserting this expression into Equations (14) and (15), and noting that the trace operator is invariant under cyclic permutations, we see that

$$\frac{\partial \ln \mathcal{L}}{\partial C_b} = \frac{1}{2}\,\mathrm{tr}\left[\left(Y^{\dagger}P^{t}\bar{C}^{-1}\right)\left(\bar{d}\bar{d}^{t} - \bar{C}\right)\left(\bar{C}^{-1}PY\right)I_b\right] \tag{18}$$

$$F_{bb'} = \frac{1}{2}\,\mathrm{tr}\left[\left(Y^{\dagger}P^{t}\bar{C}^{-1}PY\right)I_b\left(Y^{\dagger}P^{t}\bar{C}^{-1}PY\right)I_{b'}\right]. \tag{19}$$

While these expressions look somewhat formidable at first glance, they are in fact computationally very efficient. Starting with the first derivative, the important point is that all multipole dependencies have been factorized away from expensive dense matrix products. After precomputing PY (which only has to be done once for every basis set) and grouping the matrix products as indicated with parentheses in the above equations, the computational cost of the first derivative is given by only two matrix products plus one Cholesky factorization/solve, and the total memory consumption is equivalent to four dense matrices. The memory consumption is independent of the number of power spectrum bins, and the CPU time is only weakly dependent on the number of bins, involving only a single sparse trace evaluation.

A similar consideration holds for the Fisher matrix. In this case, the main computational cost lies in evaluating $Y^{\dagger}P^{t}\bar{C}^{-1}PY$ once, at the cost of one Cholesky factorization/solve and one matrix multiplication. Computing the remaining product and traces is computationally fast because of the high sparsity of the I_b operator.

The iterative QML algorithm as described above has one major weakness: the power spectrum proposed in iteration i does not necessarily yield a positive definite total data covariance matrix, $\bar{C}$. This typically happens whenever one or more likelihood conditionals have a sharp edge beyond which (symbolically) $S_{\ell m} < -N_{\ell m}$, which is not uncommon in the noise-dominated regime (see Section 7 for explicit examples).

As a safeguard against this problem, we modify the QML algorithm as follows.

1. Make some initial guess at the power spectrum, $C_b^{(0)}$.
2. Update the spectrum according to the following rule:

$$C_b^{(i)} = C_b^{(i-1)} + \alpha \sum_{b'}\left(F^{-1}\right)_{bb'}\frac{\partial \ln \mathcal{L}}{\partial C_{b'}}, \tag{20}$$

where the step length, α, maximizes $\mathcal{L}(C_b^{(i)})$. We implement the latter optimization with a standard line optimizer (linmin; Press et al. 2007).

3. Convergence is defined when the log-likelihood has changed by less than 0.1 over the last three iterations.

The underlying intuition is simply to tune the step size along the proposed QML direction such that the likelihood is maximized. Each step will necessarily lead to a higher likelihood value, and the algorithm cannot diverge.

Unfortunately, this stability comes at a non-negligible computational cost, as one now has to perform a nonlinear optimization within each main QML iteration, and this operation requires repeated likelihood evaluations. However, since each likelihood evaluation is quite fast due to the compression step described above (after all, the likelihood function is designed to be an active component in an MCMC cosmological parameter estimation framework), this is not a showstopper; the benefit of additional stability more than compensates for this expense.

Before turning to applications, we make one note regarding error estimation. Often, $\sqrt{(F^{-1})_{bb}}$ is adopted as an uncertainty on the QML estimate, a choice that is primarily driven by computational efficiency. In this paper, we quote asymmetric 68% confidence limits, computed by mapping out the likelihood conditionally around the maximum-likelihood point for each parameter, and finding the smallest range that encompass 68% of the conditional likelihood volume.

6. BASIS OPTIMIZATION

We now turn our attention to basis set optimization, considering each of the five candidates defined in Section 4 as applied to the 9-year *WMAP* data described in Section 3. In our framework, basis optimization corresponds simply to determining the harmonic space truncation multipole, ℓ_{t}, and eigenvalue thresholds, ϵ_{T} and ϵ_{P}, which result in the smallest number of accepted modes under the constraint that the information content over some range of multipoles be conserved. For the main analysis, we consider $2 \leqslant \ell \leqslant 32$ to be the multipole range of interest, matching that of the official *WMAP* temperature likelihood, and as a secondary test, we consider a case in which the polarization range is constrained to $\ell \leqslant 10$, directly targeting the low-ℓ *EE* reionization peak.

The goal of our first test is to compare the efficiency of the five candidate bases. For this, we base our statistic on the Fisher information: for each combination of truncation multipole and eigenvalue thresholds, we compare the Fisher uncertainty (i.e., $F_{ii}^{-1/2}$) for the proposed basis with the corresponding value computed from the full-pixel basis. We then require that the uncertainty not increase by more than 10% for any multipole within the range of interest. Thus, this test is designed only to compare the relative compression efficiency of the various bases, not measure the absolute information content, as correlations between multipoles are not properly quantified.

The results from these calculations are summarized in Figure 4, plotting the lowest number of accepted modes for each basis as a function of truncation multipole. First, we see that higher truncation multipoles generally require more modes in the basis in order to produce stable low-ℓ results. This makes sense because many of the newly added high-ℓ modes have a higher eigenvalue than some of the previous low-ℓ modes, and therefore more modes have to be included to retain the same low-ℓ information. Selecting modes based on harmonic content rather than eigenvalue would circumvent this issue. Second, and this is the main point of the plot, we see that the four candidate bases behave quantitatively differently: while the inverse noise basis requires more than 4500 modes to produce
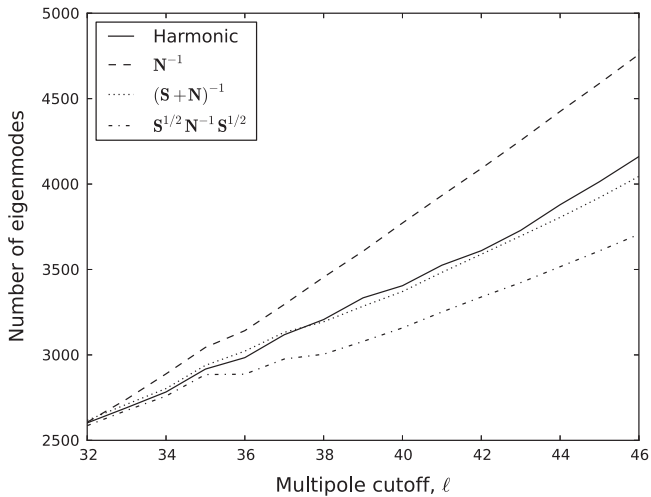
Figure 4. Number of modes required for the Fisher uncertainty to increase by no more than 10% relative to the pixel basis, including modes between $2 \leqslant \ell \leqslant 32$ and plotted as a function of truncation multipole.

Figure 5. Relative Fisher uncertainty increase as a function of truncation multipole for each of the four cosmologically interesting power spectra (C_ℓ^{TT}, C_ℓ^{TE}, C_ℓ^{EE} and C_ℓ^{BB}), evaluated for the signal-to-noise basis.

robust results for high truncation multipoles, only 3500 modes are needed in the signal-to-noise basis, corresponding to a reduction of 22% in the number of modes or a theoretical speed-up of 2. The two other bases lie in between, and are fairly close to each other. All four candidate bases achieve substantial compression compared to the original pixel basis, including a total of 6836 modes. In the following, we adopt the signal-to-noise basis as our default compression basis.

In Figure 5, we plot the relative Fisher uncertainty increase as a function of the truncation multipole for each of the four cosmologically interesting power spectra (C_ℓ^{TT}, C_ℓ^{TE}, C_ℓ^{EE}, and C_ℓ^{BB}) for this basis. Decreasing ℓ_t gradually from 32 to 26, we observe two main effects. First, the most striking feature is that the uncertainties increase by almost an order of magnitude for any multipoles at $\ell > \ell_t$. However, they do not become infinite because of the non-orthogonality introduced by the mask. In other words, there is information about high multipoles in cut-sky harmonics ("pseudo-$a_{\ell m}$s"). Conversely, the second effect is that the uncertainty on multipoles below ℓ_t also increases when removing high-ℓ modes, gradually increasing the low-ℓ noise floor.

From Figure 4, we know that no Fisher uncertainties between $2 \leqslant \ell \leqslant 32$ increase by more than 10% when including 2500 modes or more in the signal-to-noise basis. However, this is a quite crude criterion, and is not sufficient for establishing a proper production likelihood; for this, we have to make sure that correlations are also properly accounted for. We therefore define a more directly applicable statistic through a simple two-parameter amplitude–tilt model of the form $C_\ell(q, n) = q(\ell/\ell_{\mathrm{pivot}})^n C_\ell^{\mathrm{fid}}$, and map out the two-dimensional (2D) (q, n) likelihood for each effective basis. The search is performed in terms of the number of modes, and only the signal-to-noise basis is subjected to this analysis.

A subset of the results derived in this calculation is shown in Figure 6, in the form of 2D likelihood contours. Here, we see that when including only 2473 modes, which resulted in a less than 10% increase in any single multipole error bar, the integrated uncertainties over the entire range lead to significant changes. However, the agreement rapidly improves when adding more modes, and with 3102 modes the agreement with
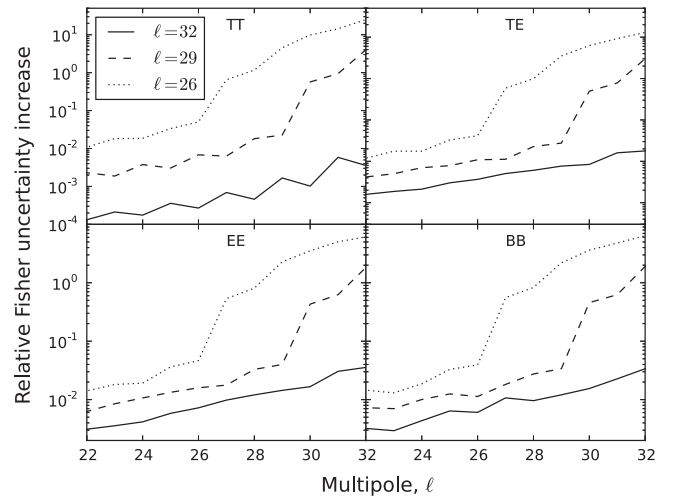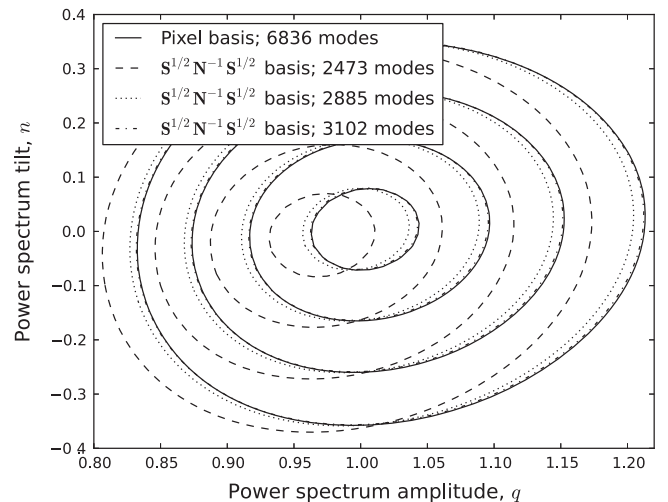
Figure 6. Two-parameter amplitude–tilt likelihoods for various eigenmode cutoffs for the signal-to-noise basis (broken contours), compared to the corresponding likelihood evaluated from the pixel basis (solid contours). The contours indicate the peak and 1, 2, and 3σ confidence regions, respectively.
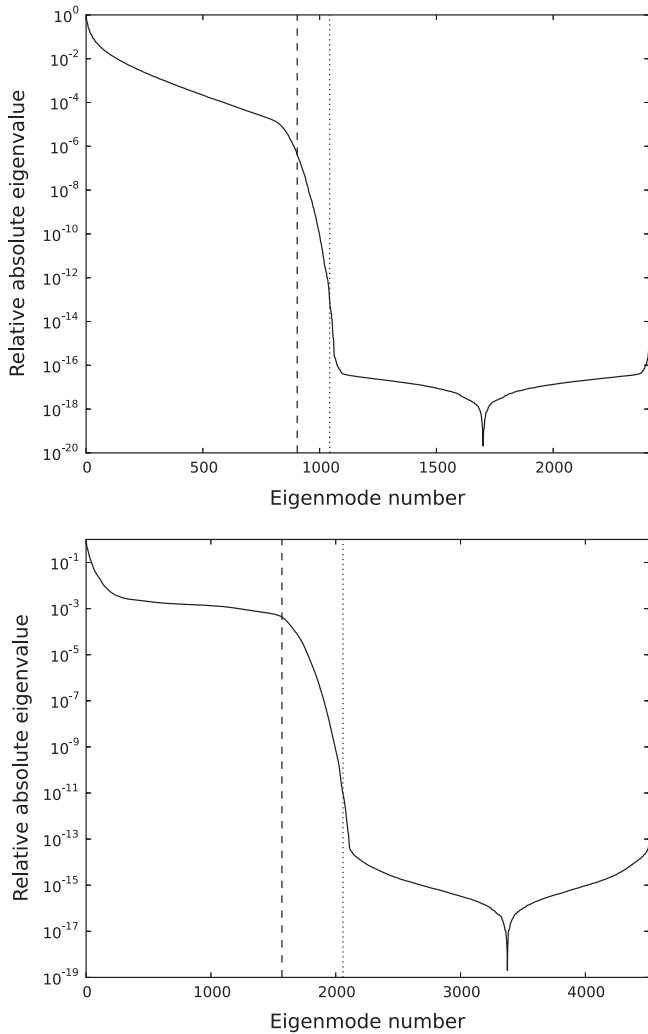
the pixel basis is very good. To quantify this statement, we calculate the integrated absolute difference between the two distributions,

$$\Delta = \int |\mathcal{L}_1 - \mathcal{L}_2| \, dq \, dn, \qquad (21)$$

and compare the resulting parameter with that computed from two bi-variate Gaussians with identical covariances but different means. Numerically, we find a value of $\Delta = 0.002$ for the basis including 3102 modes, which corresponds to a shift of 0.006σ for two bi-variate Gaussians. Recomputing the Fisher uncertainties with this new basis, we find that the relative error increase is now smaller than 3.8% for all multipoles.

We emphasize that this particular number of modes is not to be taken as a universal prescription, and will generally depend on the signal-to-noise ratio of the data in question. However, we should note that the required number of modes needed for convergence lies close to the number of modes that are left after truncating the multipole expansion for temperature and
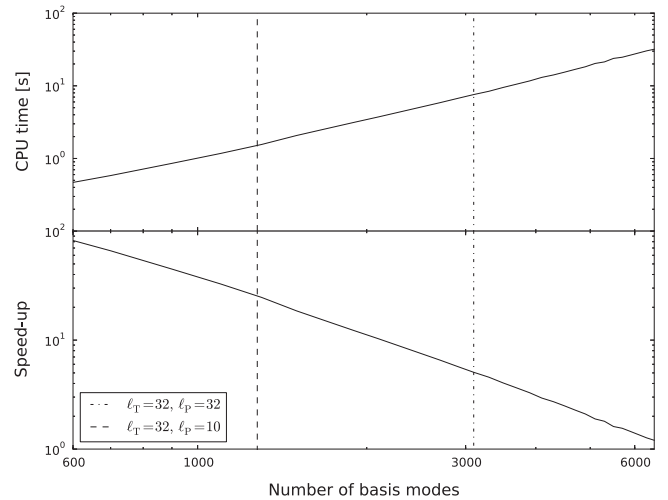
Figure 8. CPU time per likelihood evaluation as a function of the number of basis modes (top), and corresponding speed-up relative to pixel basis evaluation (bottom). The average was calculated from 50 evaluations running on a single CPU. Vertical lines indicate the timing estimates for the two bases found in the text.
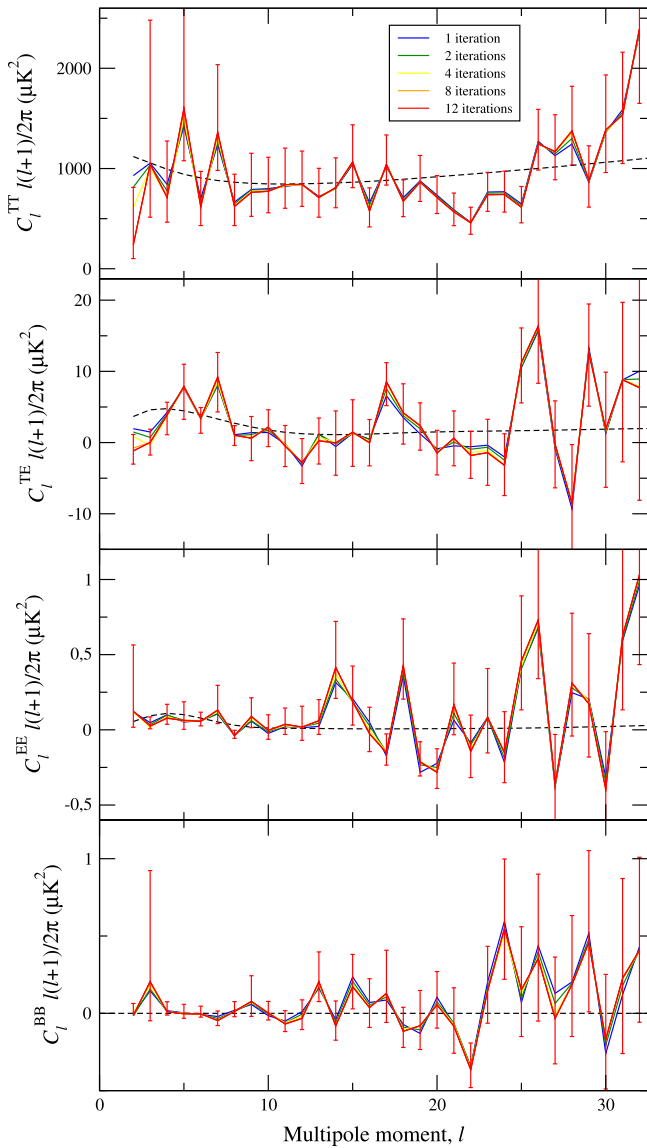
the eigenspectrum of the signal-to-noise basis, and determining the cutoff at which the numerically singular region begins.

Before concluding this section, in Figure 8, we show the CPU time per likelihood evaluation as a function of the number of basis modes (top panel), as well as the corresponding speed-up (bottom panel). Each point in this plot is computed as the average of 50 consecutive single-CPU evaluations. For the pixel basis that includes 6836 modes, each likelihood evaluation requires 35 CPU seconds, while each signal-to-noise basis evaluation (including 3102 modes) requires 7.5 CPU seconds. The realized speed-up is thus a factor of 5, which, although significant, is lower than the theoretical limit of $(35/7.5)^3 = 11$ by a factor of two. On the other hand, all of the expensive operations are implemented using standard Lapack routines, which are already highly optimized, and fully gaining this factor is not trivial.

7. POWER SPECTRUM, LIKELIHOOD, AND PARAMETERS

In this section, we assess the performance of the compressed likelihood formalism in terms of the CMB power spectrum, the likelihood, and cosmological parameters. Figure 9 shows the low-ℓ *WMAP* temperature and polarization power spectrum as derived with the QML estimator described in Section 5, using the signal-to-noise basis containing 3102 basis modes from the last section. Different colors show the results after different numbers of QML iterations going from few (blue) to many (red). The error bars indicate the asymmetric 68% errors for the last iteration. Figure 10 shows the corresponding log-likelihood as a function of the iteration.

Several interesting features can be seen in these plots. First, the initial guess adopted for this calculation was the best-fit *Planck* 2013 model, indicated as dashed lines in Figure 9, while formal convergence was achieved after 12 iterations. However, we see that already a single QML iteration results in a solution which for most multipoles is quite close to the actual maximum-likelihood solution. For exploratory work, for instance, when trying to understand the effect of systematics on low-ℓ polarization studies, a single-iteration QML power

Figure 7. Normalized temperature (top) and polarization (bottom) eigenvalue spectrum for the signal-to-noise basis with $\ell_t = 32$. The vertical lines indicate the cutoffs determined by the Fisher uncertainty (dashed) and amplitude–tilt (dotted) analyses. For robust results, all of the modes below the sharp decreases should be included.

polarization at $\ell = 32$, namely, $3 \cdot (32 + 1)^2 = 3267$, with the difference between this number and the required number of modes (3102) perhaps explained by the low signal-to-noise ratio of the polarization data.

Figure 7 shows the eigenspectrum of the signal-to-noise eigenbasis once again, this time with the two proposed eigenmode cutoffs marked as vertical lines. To achieve reasonable accuracy on individual multipoles, it is sufficient to include only the high signal-to-noise modes in the flat high eigenvalue plateaus. However, to properly account for correlations, it is important to also include the modes that lie in the rapidly dropping regime; these are partially degenerate modes that still carry some information. On the other hand, beyond this rapid decrease, the remaining eigenvalues for all practical purposes are zero, and can be excluded safely. Note that this region starts around $(32 + 1)^2$ and $2 \cdot (32 + 1)^2$ for temperature and polarization, respectively, which makes intuitive sense, given that these are the number of modes left after truncating the multipole expansion at $\ell = 32$. Thus, future basis optimization can be performed quite simply by computing the

Figure 9. 9-year *WMAP* QML power spectrum estimates as a function of QML iteration (from blue to red). Error bars are asymmetric 68% confidence limits computed from the conditional likelihood evaluated around the maximum-likelihood point for the last iteration.
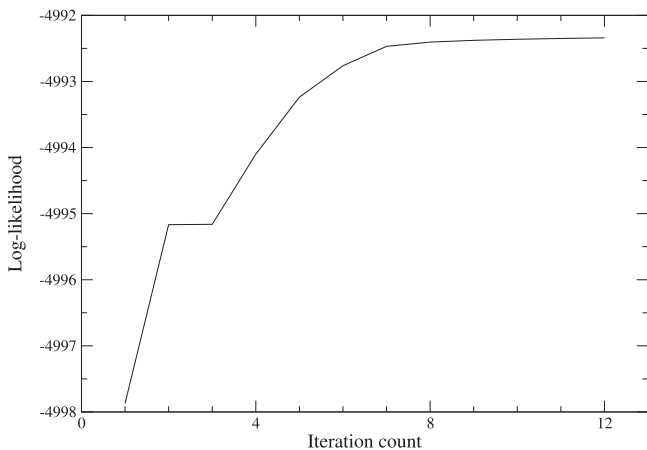

Figure 10. Log-likelihood of the QML power spectrum estimate as a function of QML iteration.
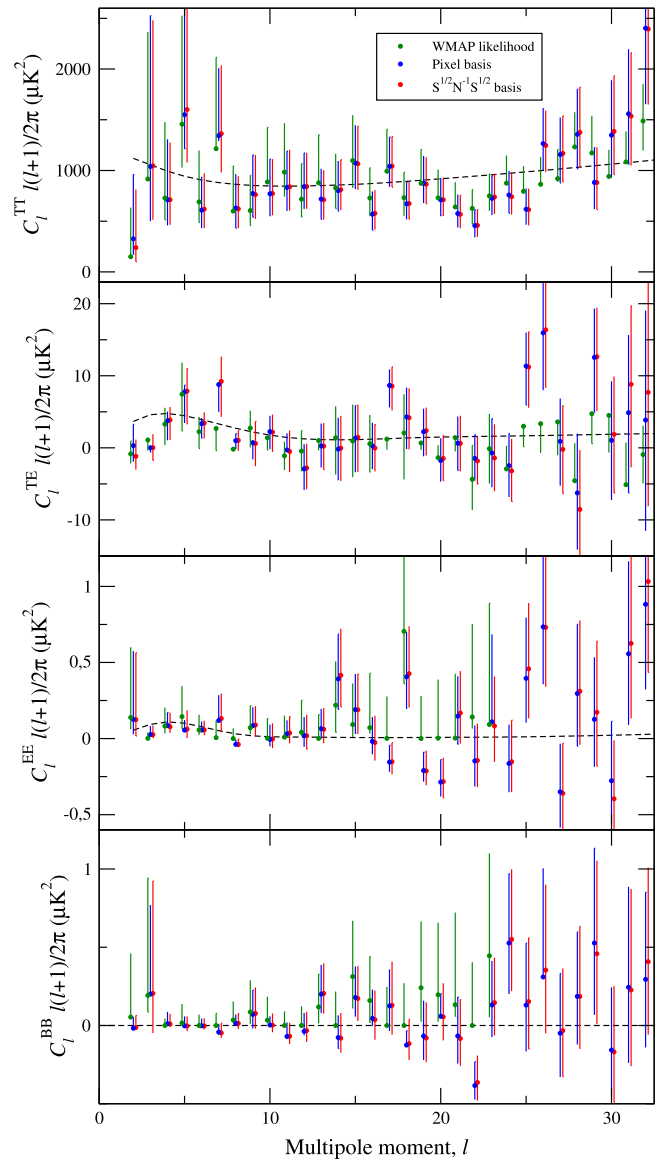
Figure 11. Comparison of maximum-likelihood power spectra derived from three different *WMAP* low-ℓ likelihood implementations. Error bars indicate asymmetric 68% confidence regions computed from conditional likelihood slices around the joint maximum-likelihood point.

spectrum approximation may be quite useful, providing a near optimal power spectrum estimate in less than two minutes of CPU time. However, for final analysis, it is clear that several iterations are indeed highly desirable, as the log-likelihood increases by more than $\Delta \ln \mathcal{L} = 5.5$ and $\Delta \chi^2 = -2\Delta \ln \mathcal{L}$ by more than 11. As a concrete example, the temperature quadrupole converges slowly because of its intrinsically non-Gaussian shape, and requires at least eight iterations before stabilizing.

In Figure 11, we compare three different power spectrum estimates, all computed by maximum-likelihood techniques using the 9-year *WMAP* data, but with different underlying likelihoods. Green points are derived directly from the official *WMAP* low-ℓ likelihood through a nonlinear multivariate Powell search (Press et al. 2007); blue points are derived from the pixel basis likelihood described in this paper using the iterative QML estimator; and red points are derived from the corresponding signal-to-noise basis that includes 3102 modes.
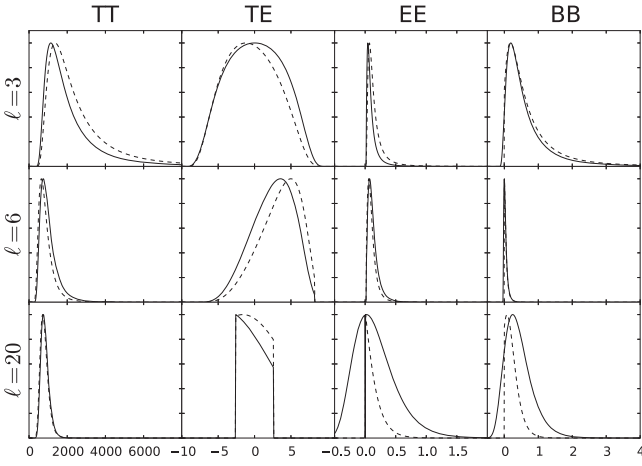
Figure 12. Comparison of conditional likelihood slices computed from the official *WMAP* likelihood (solid lines) and the signal-to-noise basis defined in the current paper (dashed lines). Other multipoles are fixed at the best-fit *Planck* 2013 ΛCDM power spectrum.

Figure 12 compares individual conditional likelihood slices computed from the *WMAP* likelihood and the signal-to-noise basis.

Overall, the agreement between the three spectra is very good, and for most multipoles the relative shifts are much less than 1σ. However, there are notable differences as well, and perhaps the most striking is the different behavior of the error estimates associated with the C_ℓ^{TE} spectrum. Considering first the *WMAP* spectrum in Figure 11, a total of 16 out of 31 power spectrum coefficients have a vanishing error bar either toward low of high values, indicating a maximum-likelihood point that lies on a sharp likelihood boundary. For comparison, for the pixel and signal-to-noise basis likelihoods, only 4 and 0 out of 31 coefficients, respectively, show similar behavior. This difference is primarily due to the different effective priors imposed by the two approaches; while the *WMAP* likelihood only requires the data covariance to be positive definite, we impose the stronger criterion that the condition number must also be well behaved (see Section 4.3). The latter prior prevents the nonlinear search algorithm from finding nonphysical power spectrum solutions near the singularity boundary with artificially high likelihood values, which in turn forces the correlated power spectrum coefficients away from their best-fit values.

We also note that the pixel-based likelihood error bars are, in a sense, less symmetric than those of the signal-to-noise basis. This is most likely a cause of the fact that the pixel-based likelihood is more ill conditioned due to the higher number of redundant modes. This degeneracy makes the likelihood slices behave worse in the pixel-based case than for the signal-to-noise basis.

A different but related issue is seen in the plot of $\mathcal{L}(C_{20}^{EE})$ in Figure 12. Here, one can see that the *WMAP* likelihood allows for significantly negative values of C_{20}^{EE}, but not values very close to zero; there is a "hole" in the likelihood surface. This is an artifact of the temperature–polarization split implemented by the *WMAP* likelihood, in that positive definiteness is assessed separately for the temperature and polarization components, effectively resulting in three independent criteria (i.e., $C_\ell^{TT} > 0$ for the Blackwell–Rao temperature component, $|\boldsymbol{S} + \boldsymbol{N}| > 0$ for the polarization component, and $C_\ell^{TE} < \sqrt{C_\ell^{TT} C_\ell^{EE}}$ for the
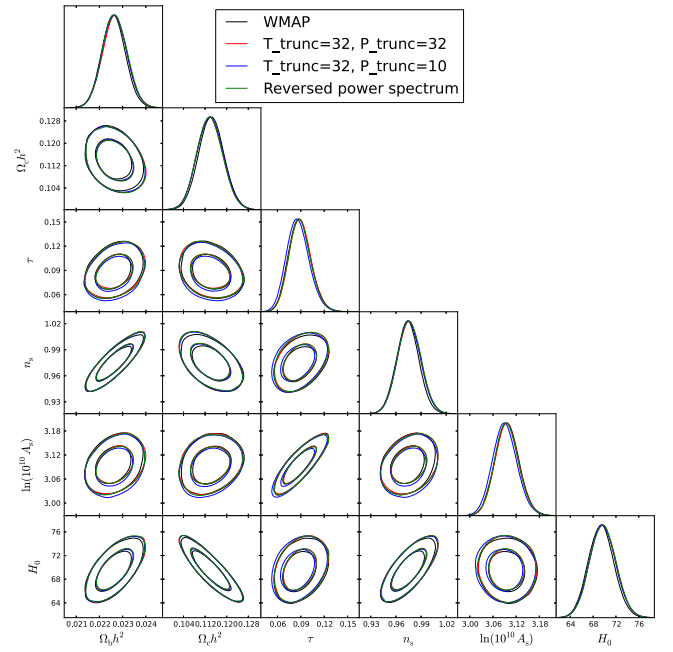
Figure 13. Cosmological parameter distributions evaluated with CosmoMC for three different likelihoods; the official *WMAP* likelihood (black), the signal-to-noise basis derived in the current paper, truncated at $\ell = 32$ (red), the same but truncated $\ell = 10$ for polarization (blue), and the signal-to-noise basis using a reversed power spectrum (green).

hybrid likelihood). With single joint likelihood implemented in this paper, this problem becomes much simpler, in that there is only a single (condition number based) numerical prior, and an optional physical prior, $C_\ell^{TE} < \sqrt{C_\ell^{TT} C_\ell^{EE}}$, whose valid parameter volume lies fully within the numerical prior region.

Our final test relates to cosmological parameters, as estimated using CosmoMC (Lewis & Bridle 2002) coupled to different versions of the 9-year *WMAP* likelihood. All of the cases used the same high-ℓ likelihood, including C_ℓ^{TT} for $33 \leqslant \ell \leqslant 1200$ and C_ℓ^{TE} for $33 \leqslant \ell \leqslant 800$, while at low ℓ's four different variations were considered:

1. the standard *WMAP* low-ℓ hybrid temperature–polarization likelihood;
2. the signal-to-noise basis likelihood derived in Section 6 including 3102 modes;
3. a similar signal-to-noise basis likelihood as (2), but with polarization truncated at $\ell = 10$;
4. the same as (2), but with the fiducial power spectrum used for basis definition extracted from an incorrect part of the original spectrum, specifically $D_\ell \leftarrow D_{1000-\ell}$ with $D_\ell \equiv C_\ell \ell(\ell + 1)/2\pi$.

The latter is a simple test of potential sensitivity to the assumed fiducial spectrum.

The results from these calculations are summarized in Figure 13 and Table 1. First and foremost, we see that all of the results are highly robust against these variations, with a maximum change of any marginal mean of at most 0.2σ. Of course, most of these parameters are dominated by small-scale information, but the optical depth of reionization, τ, which depends critically on the low-ℓ EE spectrum, also shows very small variations. Even filtering away all of the polarization multipoles above $\ell > 10$ only affects the results by 0.18σ. Finally, reversing the power spectrum does not make any

Table 1
Summary of Cosmological Parameters

| | Default *WMAP* | P trunc = 32 | | P trunc = 10 | | Reversed power spectrum | |
	Constraint	Constraint	Deviation (σ)	Constraint	Deviation (σ)	Constraint	Deviation (σ)
$\Omega_b h^2$	0.0226 ± 0.0005	0.0227 ± 0.0005	0.12	0.0227 ± 0.0005	0.09	0.0227 ± 0.0005	0.12
$\Omega_c h^2$	0.114 ± 0.005	0.114 ± 0.005	0.11	0.114 ± 0.005	0.09	0.114 ± 0.005	0.13
θ	1.040 ± 0.002	1.040 ± 0.002	0.06	1.040 ± 0.002	0.04	1.040 ± 0.002	0.05
τ	0.088 ± 0.014	0.089 ± 0.014	0.04	0.086 ± 0.014	0.18	0.089 ± 0.014	0.02
n_s	0.974 ± 0.013	0.976 ± 0.014	0.14	0.976 ± 0.013	0.11	0.976 ± 0.013	0.15
$\log[10^{10} A_s]$	3.10 ± 0.03	3.10 ± 0.03	0.001	3.09 ± 0.03	0.20	3.10 ± 0.03	0.03
H_0	69.4 ± 2.17	69.7 ± 2.22	0.13	69.7 ± 2.21	0.11	69.7 ± 2.21	0.14

Note. Parameters derived with four different low-ℓ *WMAP* likelihood implementations; the official *WMAP* low-ℓ likelihood, and the signal-to-noise basis likelihoods derived in this paper, truncated at $\ell = 32$ and 10, respectively, for polarization, and one truncated at $\ell = 32$ with a reversed power spectrum as the basis signal. The fourth, sixth, and eighth columns show the relative shifts with respect to the *WMAP* approach measured in units of σ. The confidence intervals are 1σ, and the best-fit points are marginal posterior means.

difference whatsoever, with results that are identical to the default case up to the second digit in the uncertainties.

8. SUMMARY

Building on an idea proposed by Tegmark et al. (1997), we have developed a framework for efficient low-ℓ CMB polarization likelihood analysis using linear compression, and we have applied this framework to the 9-year *WMAP* data. Five different basis definitions were compared in terms of compression efficiency and, in agreement with earlier suggestions, we find that an optimal basis may be defined in terms of the eigenvectors of $S^{1/2}N^{-1}S^{1/2}$, picking out those modes with high signal to noise. Within this basis, the original low-ℓ *WMAP* data set comprising 6834 pixels may be compressed to a smaller set of 3102 basis vectors with negligible loss of accuracy, reducing the computational cost of a single likelihood evaluation by a factor of five.

Next, we used the same framework to implement an efficient and stable version of the QML power spectrum estimator, slightly rewriting the expressions for the covariance matrix derivatives to use an explicit projection operator. The corresponding code requires about 3 GB of memory and 2 CPU minutes per QML iteration for the *WMAP* data at a HEALPix resolution of $N_{\rm side} = 16$, which is well within the capabilities of a standard laptop. Additionally, we have shown how to stabilize the QML estimator and avoid regions of parameter space in which the data covariance matrix become non-positive definite. This increases the overall computational cost by a small factor, as it relies on a nonlinear optimization within each main QML iteration, but for most cases this is not a major problem.

On a related topic, we have introduced a new and more effective prior for removing nonphysical artifacts on the likelihood surface. Previously, this was done by only requiring the data covariance matrix to be positive definite, but this leaves significant anomalies near the singularity boundary. A better option is to place a constraint on the condition number of the covariance matrix.

Finally, we note that while only the *WMAP* likelihood was considered in this paper, we expect that the methods presented here should be directly applicable to the upcoming *Planck* polarization data.

REFERENCES

Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, ApJS, 208, 20
Bond, J. R., Jaffe, A. H., & Knox, L. 1998, PhRvD, 57, 2117
Borrill, J. 1999, 3K Cosmology, 476, 277
Chu, M., Eriksen, H. K., Knox, L., et al. 2005, PhRvD, 71, 103002
Eriksen, H. K., O'Dwyer, I. J., Jewell, J. B., et al. 2004, ApJS, 155, 227
Finelli, F., De Rosa, A., Gruppuso, A., & Paoletti, D. 2013, MNRAS, 431, 2961
Gjerløw, E., Mikkelsen, K., Eriksen, H. K., et al. 2013, ApJ, 777, 150
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Gruppuso, A., De Rosa, A., Cabella, P., et al. 2009, MNRAS, 400, 463
Gruppuso, A., Finelli, F., Natoli, P., et al. 2011, MNRAS, 411, 1445
Gruppuso, A., Natoli, P., Paci, F., et al. 2013, JCAP, 7, 047
Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19
Hivon, E., Górski, K. M., Netterfield, C. B., et al. 2002, ApJ, 567, 2
Jewell, J., Levin, S., & Anderson, C. H. 2004, ApJ, 609, 1
Leach, S. M., Cardoso, J.-F., Baccigalupi, C., et al. 2008, A&A, 491, 597
Lewis, A., & Bridle, S. 2002, PhRvD, 66, 103511
Mather, J. C., Cheng, E. S., Eplee, R. E., Jr., et al. 1990, ApJL, 354, L37
Mikkelsen, K., Næss, S. K., & Eriksen, H. K. 2013, ApJ, 777, 172
Page, L., Hinshaw, G., Komatsu, E., et al. 2007, ApJS, 170, 335
Planck Collaboration et al. 2014a, A&A, 571, A1
Planck Collaboration et al. 2014b, A&A, 571, A12
Planck Collaboration et al. 2014c, A&A, 571, A16
Planck Collaboration et al. 2014d, A&A, 566, A54
Planck Collaboration et al. 2015, A&A, submitted (arXiv:1502.01588)
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 2007, Numerical Recipes: The Art of Scientific Computing (3rd ed.; New York: Cambridge Univ. Press)
Rocha, G., Magueijo, J., Hobson, M., & Lasenby, A. 2001, PhRvD, 64, 063512
Seljebotn, D. S., Mardal, K.-A., Jewell, J. B., Eriksen, H. K., & Bull, P. 2014, ApJS, 210, 24
Tegmark, M. 1997, PhRvD, 55, 5895
Tegmark, M., Taylor, A. N., & Heavens, A. F. 1997, ApJ, 480, 22
Wandelt, B. D., Larson, D. L., & Lakshminarayanan, A. 2004, PhRvD, 70, 083511
Zaldarriaga, M., & Seljak, U. 1997, PhRvD, 55, 1830