COSMIC MICROWAVE BACKGROUND LIKELIHOOD APPROXIMATION BY A GAUSSIANIZED BLACKWELL–RAO ESTIMATOR

Ø. RUDIORD1, N. E. GROENEBOOM1, H. K. ERIKSEN1,2, GREG HUEY3, K. M. GÓRSKI3,4, AND J. B. JEWELL3

1 Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029, Blindern, N-0315 Oslo, Norway; oystein.rudjord@astro.uio.no, leuat@rio.co.uk
2 Centre of Mathematics for Applications, University of Oslo, P.O. Box 1053, Blindern, N-0316 Oslo, Norway; h.k.k.eriksen@astro.uio.no
3 Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA
4 Warsaw University Observatory, Aleje Ujazdowskie 4, 00-478 Warszawa, Poland

Received 2008 September 26; accepted 2008 November 13; published 2009 February 24

The American Astronomical Society. All rights reserved. Printed in the U.S.A.

doi:10.1088/0004-637X/692/2/1669

1669

ABSTRACT

We introduce a new cosmic microwave background (CMB) temperature likelihood approximation called the Gaussianized Blackwell–Rao estimator. This estimator is derived by transforming the observed marginal power spectrum distributions obtained by the CMB Gibbs sampler into standard univariate Gaussians, and then approximating their joint transformed distribution by a multivariate Gaussian. The method is exact for full-sky coverage and uniform noise and an excellent approximation for sky cuts and scanning patterns relevant for modern satellite experiments such as the Wilkinson Microwave Anisotropy Probe (WMAP) and Planck. The result is a stable, accurate, and computationally very efficient CMB temperature likelihood representation that allows the user to exploit the unique error propagation capabilities of the Gibbs sampler to high ℓs. A single evaluation of this estimator between ℓ = 2 and 200 takes ~0.2 CPU milliseconds, while for comparison, a single pixel space likelihood evaluation between ℓ = 2 and 30 for a map with ~2500 pixels requires ~20 s. We apply this tool to the five-year WMAP temperature data, and re-estimate the angular temperature power spectrum, C_ℓ, and likelihood, ℒ( C_ℓ ), for ℓ ≤ 200, and derive new cosmological parameters for the standard six-parameter ΛCDM model. Our spectrum is in excellent agreement with the official WMAP spectrum, but we find slight differences in the derived cosmological parameters. Most importantly, the spectral index of scalar perturbations is n_s = 0.973 ± 0.014, 1.9σ away from unity and 0.6σ higher than the official WMAP result, n_s = 0.965 ± 0.014. This suggests that an exact likelihood treatment is required to higher ℓs than previously believed, reinforcing and extending our conclusions from the three-year WMAP analysis. In that case, we found that the suboptimal likelihood approximation adopted between ℓ = 12 and 30 by the WMAP team biased n_s low by 0.4σ, while here we find that the same approximation between ℓ = 30 and 200 introduces a bias of 0.6σ in n_s.

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

1. INTRODUCTION

Detailed measurements of fluctuations in the cosmic microwave background (CMB) have established cosmology as a high-precision science. One striking illustration of this is the fact that it is today possible to predict a vast number of observables based on six numbers only, with only a few (but nevertheless intriguing) “glitches” overall. The key to this success has been making accurate measurements of the CMB power spectrum, perhaps most prominently exemplified by the Wilkinson Microwave Anisotropy Probe (WMAP; Bennett et al. 2003; Hinshaw et al. 2007, 2008).

The primary connection between theoretical models and CMB observations is made through the CMB likelihood, \( \mathcal{L}(C_\ell) = P(d | C_\ell) \). This is a multivariate, non-Gaussian function that quantifies the match between the data and a given power spectrum, \( C_\ell \). Unfortunately, it is impossible to evaluate this function explicitly for modern high-resolution data sets, due to the sheer size of the problem, and one therefore instead typically resolves to various approximations.

However, given the importance of the CMB in modern cosmology, it is of critical importance to characterize this likelihood accurately, and all approximations must be thoroughly verified. One example is the approximation of the large angular scale likelihood, where \( \mathcal{L}(C_\ell) \) is strongly non-Gaussian. This turned out to be a nontrivial issue after the original analysis of the three-year WMAP temperature data by Hinshaw et al. (2007), in which a MASTER-based (Hivon et al. 2002; Verde et al. 2003) approximation was used at \( \ell > 12 \). An exact likelihood analysis (Eriksen et al. 2007a) later demonstrated that this suboptimal approximation, when applied to harmonic modes between \( \ell = 13 \) and 30, biased the spectral index of scalar perturbations, \( n_s \), low by 0.4σ.

A second example is that of noncosmological foregrounds. Unless properly accounted for, such foregrounds bias the observed power spectrum to high values and can seriously compromise any cosmological conclusions. While important for temperature observations, this is an absolutely crucial issue for polarization observations, as the desired CMB in amplitude is comparable to or weaker than the interfering foregrounds over most of the sky.

In recent years, a new set of statistical methods has been developed that allows the user to address these issues within a single well-defined framework (Jewell et al. 2004; Wandelt et al. 2004; Eriksen et al. 2004). The heart of this method is the Gibbs sampling algorithm (see e.g., Gelfand & Smith 1990), in which samples from a (typically complicated) joint distribution are drawn by alternately sampling from (simpler) conditional distributions. In the CMB setting, this is realized by drawing joint samples from \( P(s, C_\ell | d) \), by alternately sampling from \( P(C_\ell | s, d) \), where \( C_\ell \) is the CMB power spectrum, \( s \) is the CMB sky signal, and \( d \) are the observed data. In addition to allow for exact likelihood analysis at reasonable computational cost, an equally important feature of this framework is its unique capability of including additional degrees of freedom, such as noncosmological foregrounds, into the analysis (Eriksen...
Further, very recently an additional Metropolis–Hastings Markov Chain Monte Carlo (MCMC) sampling step was introduced by Jewell et al. (2008), which effectively resolves the previously described inefficiency of the Gibbs sampler at low signal to noise (Eriksen et al. 2004).

The framework has also been extended to handle polarization (Larson et al. 2007; Eriksen et al. 2007a) and anisotropic universe models (Groeneboom & Eriksen 2009). By now, the CMB Gibbs sampler is well established and demonstrated to sample efficiently from the exact CMB posterior. However, a long-standing issue has been the characterization of the joint likelihood, given a set of such samples. Originally, Wandelt et al. (2004) proposed to use the so-called Blackwell–Rao (BR) estimator for this purpose, and this approach was later implemented and studied in detail by Chu et al. (2005). While highly accurate for the large angular scale and high signal-to-noise temperature likelihood, it suffers from one major drawback: because it attempts to describe the full \( \ell \)-max-dimensional likelihood without any constraints on allowed correlations, the number of samples required for convergence scales exponentially with \( \ell_{\text{max}} \). In practice, this limits the BR estimator to \( \ell \ll 30 \) for temperature data and just \( \ell \ll 3\)–4 for low signal-to-noise polarization data.

In this paper, we introduce a new temperature likelihood approximation based on samples drawn from the CMB posterior, by modifying the original BR estimator in a way that restricts the allowed N-point functions of \( \mathcal{L}(C_i) \), but still captures most of the relevant information. Explicitly, this is done through a specific change of variables, such that the observed marginal posterior for each multipole, \( P(C_i|\mathbf{d}) \), is transformed into a Gaussian. Then, in these new variables the joint distribution is approximated by a multivariate Gaussian. As long as the correlation between any two multipoles is reasonably small, as is the case for nearly full-sky experiments such as WMAP and Planck, we shall see that this provides an excellent approximation to the exact joint likelihood. As a result, the new approach greatly reduces the overall number of samples required for convergence and allows us to obtain a highly accurate likelihood approximation to arbitrary \( \ell_{\text{max}} \). Generalization to a full polarized likelihood will be discussed in a future paper (H. K. Eriksen et al. 2009, in preparation).

This paper is organized as follows: in Section 2, we first briefly review the Gibbs sampling algorithm together with the original BR estimator, and in Section 3 we introduce the new Gaussianized Blackwell–Rao (GBR) estimator. Next, in Section 4, we apply the new estimator to simulated data and compare results with brute-force likelihood evaluations in pixel space. In Section 5, we analyze the five-year WMAP temperature data and provide an updated power spectrum and set of cosmological parameters. We summarize and conclude in Section 6.

### 2. REVIEW OF THE CMB GIBBS SAMPLER

We start by reviewing the current state of the CMB Gibbs sampling framework, as previously developed through a series of papers (Jewell et al. 2004; Wandelt et al. 2004; Eriksen et al. 2004, 2008b; Larson et al. 2007), and highlight the problem of likelihood modeling as currently presented in the literature.

#### 2.1. Elementary CMB Gibbs Sampling

First, we assume that our observations, \( \mathbf{d} \), in the direction \( \hat{n} \) may be modeled in terms of a signal, \( \mathbf{s} \), and a noise, \( \mathbf{n} \), component,

\[
\mathbf{d}(\hat{n}) = \mathbf{s}(\hat{n}) + \mathbf{n}(\hat{n}).
\]

Further, we assume that both \( \mathbf{s} \) and \( \mathbf{n} \) are Gaussian distributed with vanishing mean and covariances \( \mathbf{S} \) and \( \mathbf{N} \), respectively.

The CMB is in this paper additionally assumed to be isotropic, such that in spherical harmonic space \( (\mathbf{s}(\hat{n}) = \sum_{\ell,m} a_{\ell m} Y_{\ell m}(\hat{n})) \) the CMB covariance matrix may be written as \( \mathbf{S}_{\ell m,\ell' m'} = C_{\ell} \delta_{\ell \ell'} \delta_{mm'} \), where \( C_{\ell} = \langle a_{\ell m} a_{\ell m}^* \rangle \) is the angular power spectrum. Our goal is now to map out the CMB posterior distribution \( P(s, C_i|\mathbf{d}) \) and the CMB likelihood \( \mathcal{L}(C_i) = P(\mathbf{d}|C_i) \). Note that we in this paper are concerned with the problem of likelihood characterization only, which is a post-processing step relative to the Gibbs sampler. For notational transparency, we therefore neglect issues such as foreground marginalization, instrumental beams, multifrequency observations, etc. For full details on these issues, see, e.g., Eriksen et al. (2008a).

When working with real-world CMB data, there are a number of issues that complicate the analysis. Two important examples are anisotropic noise and Galactic foregrounds. First, because of the scanning motion of a CMB satellite, the pixels in a given data set are observed over unequal amounts of time. This implies that the effective noise is a function of pixel location on the sky. Second, large regions of the sky are obscured by Galactic foregrounds (e.g., synchrotron, free–free and dust emission), and these regions must be rejected from the analysis by masking.

Because of such issues, the total data covariance matrix \( \mathbf{S} + \mathbf{N} \) is dense in both pixel and harmonic space. As a result, it is computationally unfeasible to evaluate and sample directly from \( P(s, C_i|\mathbf{d}) \). Fortunately, this problem was originally solved by Jewell et al. (2004), Wandelt et al. (2004), Eriksen et al. (2004), who developed a particular CMB Gibbs sampler for precisely this purpose. For full details on this method, we refer the interested reader to the original papers, and in the following we only describe the main ideas. The practical implementation of the algorithm used in this paper is called “Commander,” and has been described in detail by Eriksen et al. (2004, 2008b).

The idea behind the CMB Gibbs sampler is to draw samples from the joint posterior by alternately sampling from the two corresponding conditionals. The sampling scheme may thus be written in the symbolic form

\[
\mathbf{s}^{i+1} \leftarrow P(s|C_i^i, \mathbf{d}),
\]

\[
C_i^{i+1} \leftarrow P(C_i|s^{i+1}, \mathbf{d}),
\]

where the left arrow implies sampling from the distribution on the right-hand side. Then, after some burn-in period, \( (\mathbf{s}^i, C_i^i) \) will be drawn from the desired distribution. The only remaining step is to write down sampling algorithms for each of the two above conditional distributions, both of which are readily available for our problem, since the former is simply a multivariate Gaussian, and the second is a product of independent inverse Gamma distributions. For one possible general sampling algorithm for \( P(C_i|s) \), see, e.g., Eriksen & Wehus (2009).

#### 2.2. The Blackwell–Rao Estimator

The Gibbs sampler produces a set of samples drawn from the joint CMB posterior, \( P(s, C_i|\mathbf{d}) \). However, for these samples to be useful for the estimation of cosmological parameters, we have to transform the information contained in this sample set into a smooth approximation to the likelihood, \( \mathcal{L}(C_i) = P(\mathbf{d}|C_i) \). In principle, we could simply generate a multivariate histogram...
and read off corresponding values, but this does not work in practice because of the large dimensionality of the parameter space.

In the current literature, the best approach for handling this problem is the BR estimator (Wandelt et al. 2004; Chu et al. 2005), which attempts to smooth the sampled histogram by taking advantage of the known analytic distribution, \( P(\mathbf{C}|s) \): first, we define the observed power spectrum, \( \sigma_\ell \), of the current CMB sky Gibbs sample, \( s(\hat{\mathbf{n}}) = \sum_{m} c_{lm} Y_{lm}(\hat{\mathbf{n}}) \),

\[
\sigma_\ell \equiv \frac{1}{2\ell + 1} \sum_{m=\ell}^{\ell} |a_{\ell m}|^2. \tag{4}
\]

Then, the BR estimator is derived as follows:

\[
P(\mathbf{C}|\mathbf{d}) = \frac{1}{P(\mathbf{C}|\mathbf{s}) P(\mathbf{s}|\mathbf{d})} \int P(\mathbf{C}|\mathbf{s}) P(\mathbf{s}|\mathbf{d}) d\mathbf{s} \tag{5}
\]

\[
= \int P(\mathbf{C}|\mathbf{s}) P(\mathbf{s}|\mathbf{d}) d\mathbf{s} \tag{6}
\]

\[
= \int P(\mathbf{C}|\sigma_\ell) P(\sigma_\ell|\mathbf{d}) D\sigma_\ell \tag{7}
\]

\[
\approx \frac{1}{N^G} \sum_{i=1}^{N^G} P(\mathbf{C}|\sigma_\ell^i). \tag{8}
\]

In other words, the BR estimator is nothing but the average of \( P(\mathbf{C}|\sigma_\ell) \) over the sample set, where \( \sigma_\ell \) refers to the power spectrum of a full-sky noiseless CMB signal Gibbs sample. This distribution has a simple analytic expression (e.g., Chu et al. 2005),

\[
P(\mathbf{C}|\sigma_\ell) \propto \prod_{\ell} \frac{e^{-\frac{\sum_{\ell} C_{\ell}^*}{\sigma_\ell^2}}}{\frac{\sum_{\ell} C_{\ell}^*}{\sigma_\ell^2}}. \tag{9}
\]

While Equation (8) does constitute a computationally convenient and accurate approximation to the full likelihood for some special applications, it suffers badly from poor convergence properties with increasing dimensionality of the sampled space. This behavior may be understood in terms of relative distribution widths: suppose we want to map out an \( \ell_{\text{max}} \)-dimensional distribution, and each of the univariate BR functions, \( [i.e., P(\mathbf{C}|\sigma_\ell)] \), have a standard deviation of, say, 90% of the corresponding marginal distributions. The total volume fraction spanned by a single sample in \( \ell_{\text{max}} \) dimensions is then \( f = 0.9^{\ell_{\text{max}}} \), an exponentially decreasing function with \( \ell_{\text{max}} \). Therefore, it also takes an exponential number of samples in order to build up the full histogram, and this becomes computationally unfeasible for realistic data sets already at \( \ell_{\text{max}} \lesssim 30-50 \) (Chu et al. 2005).

The main problem with this approach is that one attempts to map out all possible \( N \)-point correlation functions between all multipoles. The number of such \( N \)-point functions is obviously overwhelming with increasing dimensionality. But this also hints at a possible resolution of the problem: we know by experience that the CMB likelihood is a reasonably well-behaved function, in that (1) there are only weak correlations between multipoles for data sets with nearly full-sky coverage and (2) that even including just two-point correlations (in transformed variables) produces very reasonable results (e.g., Bond et al. 2000; Verde et al. 2003). This intuition will be used in the next section to define a stable likelihood estimator.

3. THE GAUSSIANIZED BLACKWELL–RAO ESTIMATOR

We now introduce a new Gibbs-based likelihood estimator which we call the “GBR.” The basic idea behind this approach is similar to that employed by, e.g., Bond et al. (2000), Hamimeche 

Explicitly, our approximation is defined by transforming the univariate marginal distributions, \( P(\mathbf{C}|\mathbf{d}) \), into Gaussianized variables, \( x_\ell \), and then assuming a multivariate Gaussian distribution in these transformed variables,

\[
P(\mathbf{C}|\mathbf{d}) = \left( \prod_{\ell} \frac{\partial C_{\ell}}{\partial x_\ell} \right)^{-1} P(\mathbf{x}|\mathbf{d}). \tag{10}
\]

Here \( \frac{\partial C_{\ell}}{\partial x_\ell} \) is the Jacobian of the transformation, and \( \mathbf{x} = \{x_\ell\} \) is a Gaussian random vector with mean \( \mu = \{\mu_\ell\} \) and covariance matrix \( C_{\ell m} = \langle (x_\ell - \mu_\ell)(x_m - \mu_m) \rangle \). Thus, the approximation of our likelihood estimator relies on the assumption that

\[
P(\mathbf{x}|\mathbf{d}) \approx e^{-\frac{1}{2}(\mathbf{x} - \mu)^T C^{-1}(\mathbf{x} - \mu)}. \tag{11}
\]

Note that this is by construction exact for the full-sky uniform noise case, because the covariance matrix in this case is diagonal, and the full expression factorizes in \( \ell \); in that case we are only performing an identity operation.

### 3.1. Transformation to Gaussian Marginal Variables

The first step in our algorithm is to compute the change-of-variables rule from \( C_\ell \) to \( x_\ell \) that transforms the marginal distribution, \( P(\mathbf{C}|\mathbf{d}) \), for each \( \ell \) into a Gaussian distribution, \( P(x_\ell|\mathbf{d}) \). The data used for this process are the \( \sigma_\ell \) samples drawn from the joint posterior \( P(\mathbf{C}|\mathbf{s}) \) by the CMB Gibbs sampler.

We use two different methods of estimating the marginal distributions from these samples. The first approach is to estimate \( P(\mathbf{C}|\mathbf{d}) \) with the BR estimator as defined by Equation (8), over a grid in \( \mathbf{C}_\ell \) for each \( \ell \). Then, a cubic spline is fitted to the resulting distribution. This is the preferred approach for high signal-to-noise or low-\( \ell \) modes.

However, for low signal-to-noise and high-\( \ell \) modes one observes similarly poor convergence properties of this marginal estimator as for the full joint estimator. In these cases, we therefore instead compute a simple histogram directly from the \( \mathbf{C}_\ell \) samples and fit a smooth spline (Green 

The computational expense of the Gibbs sampler is driven by sampling from \( P(\mathbf{s}|\mathbf{C}_\ell, \mathbf{d}) \), not by \( P(\mathbf{s}|\mathbf{C}_\ell) \). Note that this approach naturally supports arbitrary \( \mathbf{C}_\ell \) binning schemes (Eriksen & Wehus 2009) and also interfaces naturally with the hybrid MCMC scheme described by Jewell et al. (2004).

Given these spline approximations to \( P(\mathbf{C}|\mathbf{d}) \) for each \( \ell \), we compute the corresponding cumulative distributions by numerical integration,

\[
F(\mathbf{C}|\mathbf{d}) = \int_0^{\mathbf{C}_\ell} P(\mathbf{C}^\prime|\mathbf{d}) d\mathbf{C}^\prime_\ell.
\]

This is subsequently identified with a standard Gaussian distribution with zero mean and unity variance. Explicitly, we find
where \( x_\ell(C_\ell) \) over a grid in \( C_\ell \) such that
\[
F(C_\ell|d) = F_{\text{Gauss}}(x_\ell) = \frac{1}{2} \left(1 + \text{erf} \left( \frac{x_\ell}{\sqrt{2}} \right) \right),
\]
where \( \text{erf} \) is the error function. This equation is straightforward to solve using standard numerical root-finding routines. The result is a convenient set of look-up tables \( x_\ell(C_\ell) \), again stored in the form of cubic splines, that allows for very efficient transformation from standard to Gaussian variables for arbitrary values of \( C_\ell \). From these splines, it is also easy to compute the derivatives required for the Jacobian in Equation \( (10) \).

3.2. Estimation of the Joint Gaussian Density

Having defined a change-of-variables for each \( \ell \), the remaining task is to estimate the joint distribution, \( P(x|d) \), in the new variables. In this paper, we approximate this distribution by a joint Gaussian, but any parametric function could of course serve this purpose. For example, we implemented support for the skew-Gaussian distribution (e.g., Azzalini & Capitanio 2003) in our codes, but found that the improvement over a simple Gaussian was very small.

The only free parameters in this multivariate Gaussian distribution are the mean, \( \mu_\ell \), and the covariance, \( C \). These are again estimated from the samples produced by the Gibbs sampler. First, we draw \( N \sim \mathcal{O}(10^6) C_\ell \) samples from \( P(C_\ell|\sigma_\ell) \), as described above, but this time including all \( \ell \)s for each sample. Then, we Gaussianize these \( \ell \)-by-\( \ell \), by evaluating \( x_\ell(C_\ell) \) for each sample and multipole moment. Finally, we compute the corresponding means and standard deviations,
\[
\mu_\ell = \frac{1}{N} \sum_{i=1}^{N} x_\ell^i,
\]
\[
C_{\ell\ell} = \frac{1}{N} \sum_{i=1}^{N} (x_\ell^i - \mu_\ell)(x_\ell^i - \mu_\ell),
\]
where the sums run over sample index.

4. APPLICATION TO SIMULATED DATA

Before applying the machinery described in the previous section to the five-year WMAP data, we verify the method by analyzing a simulated low-resolution data set. The reason for considering a low-resolution simulation is that only in this case is it possible to evaluate the exact likelihood by brute force in pixel space, without making any approximations.

The simulation is made by drawing a Gaussian realization from the best-fit five-year WMAP \( \Lambda \)CDM power spectrum (Komatsu et al. 2008), smoothing this with a 10\( \degr \) FWHM Gaussian beam, and projecting it on an \( N_{\text{side}} = 16 \) HEALPix\(^5\) grid. Finally, 20\( \mu \)K root mean square (rms) white noise is added to each pixel, and the (degraded) WMAP KQ85 sky cut (Gold et al. 2008) is applied to the data. The maximum multipole considered in this analysis was \( \ell_{\text{max}} = 47 \), and the spectrum was binned with a bin size of \( \Delta \ell = 5 \) from \( \ell = 20 \). The signal to noise is unity at \( \ell = 19 \) and negligible beyond \( \ell \geq 30 \).

We now compute slices for each \( \ell \) through the full multivariate likelihood, both with the method described in Section 3 and by brute-force pixel space evaluation (e.g., Eriksen et al. 2007b), fixing all other \( \ell \)s at the input \( \Lambda \)CDM spectrum. For comparison, we also compute the marginal distributions for each \( \ell \).

The results from this exercise are shown in Figure 1. The black lines indicate the brute-force likelihoods, and the red lines show the GBR likelihoods. The green lines show the marginal distributions, visualizing the effect of mode coupling due to the sky cut.

First, we see that all distributions agree very closely at \( \ell \leq 8 \). In this very large scale regime, all harmonic modes are sufficiently well sampled with the KQ85 sky cut so that mode coupling is negligible. However, from \( \ell \geq 10 \) the marginal distributions are noticeably different from the likelihood slices, with a typical shift in peak position of \( \sim 100 \mu K^2 \)s. We also see that these correlations are accurately captured by the Gaussian approximation implemented in the GBR estimator, as the GBR

\(^{5}\) http://healpix.jpl.nasa.gov.
likelihoods are essentially identical to the brute-force slices up to $\ell = 18$.

At the very high $\ell$ and low signal-to-noise end, we see slight differences between the GBR and the pixel space slices, and in fact, the agreement is better with the marginal distributions. This is caused by poor convergence of the covariance matrix in this particular run and is included here for pedagogical purposes only: in a real analysis, one must always make sure that all distributions have converged well, typically by analyzing different chain sets separately. Note also that with sufficiently wide bins, the correlations to neighboring bins eventually vanish, and in this case it may be better to remove these correlations by hand from the covariance matrix, rather than trying to estimate them by sampling. Whether this is the case or not for a given set can again be estimated by jackknife tests. Finally, for the five-year WMAP analysis presented in this paper, we will only use the GBR estimator in the high signal-to-noise regime, and in that case the distributions converge very quickly.

5. FIVE-YEAR WMAP TEMPERATURE ANALYSIS

We now apply the tools described in Section 3 to the five-year WMAP temperature data. We only consider $\ell \leq 200$ in this paper, to avoid issues with error propagation for unresolved point sources and beam estimation. However, we do correct for the mean spectrum of unresolved point sources, as described below.

5.1. Data

We analyze the foreground reduced five-year WMAP $V$-band temperature sky maps, which are available from Legacy Archive for Microwave Background Data Analysis (LAMBDA). The $V$-band data were chosen because these are considered to be the cleanest in terms of foregrounds out of the five WMAP bands (Gold et al. 2008). Further, at $\ell \leq 200$ the $V$-band alone is strongly cosmic variance dominated, and therefore one does not gain any significant statistical power by co-adding with other bands. Instead, one only increases the chance of introducing foreground biases by adding more frequencies. We work with the individual differing assembly (DA) maps (Hinshaw et al. 2003) and take into account the beam and noise pattern for each map separately.

The WMAP sky maps are pixelized at an HEALPix resolution of $N_{\text{side}} = 512$, corresponding to a pixel size of $7^\circ$, and the instrumental beam of the two $V$-band channels has an FWHM of $21^\circ$. We therefore impose an upper harmonic mode limit of $\ell_{\text{max}} = 700$ in the Gibbs sampling (Commander) step, probing deeply into the noise-dominated regime. Note, however, that we only use $\ell \leq 200$ in the GBR estimator, to avoid high-$\ell$ complications, such as beam and point source error propagation, in the cosmological parameter stage.

We correct the spectrum for unresolved point sources using the WMAP model. Explicitly, the mean spectrum due to unresolved point sources in a single frequency, $\nu$, for the five-year WMAP data is modeled as (Hinshaw et al. 2003, 2007; Nolta et al. 2008)

$$C_{\ell}^{\text{ps}} = A_{\text{ps}}a(\nu)^2 \left( \frac{\nu}{\nu_0} \right)^{2\beta}, \quad (14)$$

where $A_{\text{ps}} = 0.011 \pm 0.001$ is the point source amplitude relative to the $Q$-band channel ($\nu_0 = 41$ GHz), $\beta = -2.1$ is the best-fit spectral index of the point sources, and $a(\nu)$ is the conversion factor between antenna and thermodynamic temperature units. To correct for this in our analysis, we subtract $C_{\ell}^{\text{ps}}$, evaluated at $\nu = 61$ GHz, from each $\sigma_\ell$ sample before computing the GBR estimator.

Finally, we impose the WMAP KQ85 sky cut (Gold et al. 2008) on the data that masks point sources, removing 18% of the sky. Note that we adopt the template-corrected maps provided by the WMAP team in this analysis and postpone an internal Gibbs sampling based foreground analysis to a future paper; for now our main focus is the new likelihood approximation, not the impact of foregrounds.

5.2. Analysis Overview

The analysis consists of the following steps:

1. Generate 4000$\sigma_\ell$ samples with Commander from the five-year V1 and V2 DAs, including $\ell$s up to $\ell_{\text{max}} = 700$, divided over eight chains.
2. Generate 500,000$C_{\ell}$ samples from these $\sigma_\ell$s, including $\ell$s between $\ell = 2$ and 250.
3. Compute the corresponding GBR parameters, i.e., transformation tables, means $\mu$, and covariance matrix $C$.
4. Modify the five-year WMAP temperature likelihood by replacing the existing low-$\ell$ part with Equation (10), with the parameters given in Equation (3). The transition multipole between the low-$\ell$ and high-$\ell$ is increased from $\ell = 32$ to 200. Multipoles between $\ell = 201$ and 250 are included in the GBR estimator to avoid truncation effects, but the spectrum in this range is kept fixed at a fiducial spectrum, in order not to count these multipoles twice.
5. Cosmological parameters are estimated using CosmoMC (Lewis & Bridle 2002).

5.3. Convergence Analysis

Before presenting the results from the WMAP analysis, we consider the question of convergence. First, we compute the Gelman–Rubin statistic (Gelman & Rubin 1992) for each $\sigma_\ell$ using the eight chains computed with Commander and removing the first 20 samples for burn-in. We find that $R - 1$ is less than 0.01 for $\ell \lesssim 300$ and less than 1.1 for $\ell \lesssim 500$, indicating very good convergence in terms of power spectra.

However, the fact that each $\sigma_\ell$ individually is well converged does not automatically imply that the full likelihood is well converged, since the latter depends crucially on the correlations between $\sigma_\ell$s. To assess the convergence in terms of cosmological parameters, we therefore analyze a toy model, by fitting a simple two-parameter amplitude and tilt, $q$ and $n$, model,

$$C_{\ell} = q \left( \frac{\ell}{\ell_{\text{pivot}}} \right)^n C_{\ell}^{\text{fid}}, \quad (15)$$

to the WMAP data between $\ell = 2$ and 250 with the GBR likelihood. Here $C_{\ell}^{\text{fid}}$ is a fiducial power spectrum, which is chosen to be the best-fit five-year WMAP $\Lambda$CDM power spectrum (Komatsu et al. 2008), and $\ell_{\text{pivot}} = 150$. We then map out the likelihood in a grid over $q$ and $n$. This is repeated twice, first including samples from chain numbers 1 to 4 and then from chain numbers 5 to 8.

The results from this exercise are shown in Figure 2 in terms of two sets of likelihood contours, corresponding to each of the two chain sets, respectively. The agreement between the two is excellent, indicating that we also have good convergence in

---

6 http://lambda.gsfc.nasa.gov.
terms of cosmological parameters with the existing sample set. Note also that the point \((q, n) = (1, 0)\) lies well inside the 1σ confidence region, indicating that the best-fit WMAP model, which is obtained including all \(\ell\)s, and this is found to lie well inside the 1σ contour.

Third, as described in Section 3, we construct the GBR covariance matrix from \(N = \mathcal{O}(10^6)\) \(C_\ell\) samples drawn from the (smaller) set of \(\sigma_\ell\) samples. An outstanding question is how large \(N\) should be in order for this covariance matrix to reach convergence, as a function of \(\ell_{\text{max}}\). To settle this question, we carry out the following simple exercise: first we produce two \(C_\ell\) sample sets, each containing \(N\) samples, and all drawn from a single \(\sigma_\ell\) sample. Second, we compute the two corresponding covariance matrices, invert these, then subtract them from each other, and finally compute the standard deviation of all elements. Third, we define the inverse covariance matrix to be converged if the rms MCMC noise is less than 0.005, corresponding to 0.5% of the diagonal elements. (We have checked that this produces robust parameter estimates.) We then find the smallest \(N\) such that this is satisfied, as a function of \(\ell_{\text{max}}\).

The results from this exercise are shown in Figure 3. Here, we see that the number of samples required for convergence rises rapidly up to \(\ell \sim 30\), reaching a maximum of \(~8 \times 10^4\) samples, and then flattens to a plateau. To be on the safe side, we therefore always use either \(5 \times 10^5\) or \(10^6\) samples in the WMAP analysis.

The reason for this behavior becomes intuitive when considering the structure of the actual matrix. This is shown in Figure 4, in the form of a correlation matrix

\[
\hat{C}_{\ell\ell} = \frac{C_{\ell\ell}}{\sqrt{C_{\ell\ell}C_{\ell\ell}}} - \delta_{\ell\ell}.
\]

The main features of this matrix are negative correlations around the diagonal, with the largest amplitudes observed between \(\ell\) and \(\ell \pm 2\). This is expected: first, two modes separated by \(\Delta \ell = 1\) have different parity and can therefore not easily mimic each other. On the other hand, modes separated by \(\Delta \ell = 2\) have both identical parity and similar angular scale, and it is therefore possible to add power to one mode and subtract it from the other, and still maintain an essentially unchanged image. The result is a noticeable anticorrelation between \(\ell\) and \(\ell \pm 2\).

At larger separations in \(\ell\), the correlations die off rapidly, since it is difficult for a large-scale mode to mimic a small-scale mode with a reasonably small sky cut. And this explains the convergence behavior seen in Figure 3: the covariance matrix is strongly band-limited. Therefore, once one has a sufficiently large number of \(C_\ell\) samples for a sub-block to converge, there is also enough samples for a sub-block further away to converge. These are essentially uncorrelated.

5.4. Results

We now present the main results derived from the five-year WMAP temperature data with the GBR estimator between \(\ell = 2\) and 200. First, in the top panel of Figure 5 we plot the power spectrum obtained by maximizing the GBR likelihood together with the official five-year WMAP power spectrum. The bottom panel shows the difference between these two, and the gray band indicates the standard deviation of \(\sigma_\ell\), i.e., the uncertainty due to noise and sky cut but not to cosmic variance. Clearly, the agreement between the two power spectra is very good.

Next, in Figure 6 we compare a few selected slices through the GBR likelihood with slices through the WMAP likelihood. All \(\ell\)s other than the one currently considered are kept fixed at the best-fit \(\Lambda\)CDM spectrum. Here, we see that there are small shifts in peak positions, corresponding to the small differences seen in the power spectra in Figure 5. However, a main point...
in this plot is that the GBR likelihood slices are well behaved even at the highest $\ell$s, and this is not the case for the standard BR estimator (Chu et al. 2005).

Finally, in Table 1 and Figure 7 we summarize the marginal cosmological parameter posteriors obtained with the two likelihood codes from CosmoMC. Interestingly, there are some notable differences at the 0.3–0.6$\sigma$ level, with the most striking example being the spectral index of scalar perturbations, $n_s = 0.973 \pm 0.014$. This is only 1.9$\sigma$ away from unity and 0.6$\sigma$ higher than the official WMAP values.

### Table 1

| Parameter | WMAP | GBR | Shift in $\sigma$ |
|-----------|------|-----|--------------------|
| $\Omega_m h^2$ | $0.0228 \pm 0.0006$ | $0.0230 \pm 0.0006$ | 0.4 |
| $\Omega_b h^2$ | $0.109 \pm 0.006$ | $0.0108 \pm 0.006$ | -0.3 |
| $\log(10^{10} A_s)$ | $3.06 \pm 0.04$ | $3.06 \pm 0.04$ | 0.0 |
| $h$ | $0.722 \pm 0.03$ | $0.732 \pm 0.03$ | 0.3 |
| $n_s$ | $0.965 \pm 0.014$ | $0.973 \pm 0.014$ | 0.6 |
| $\tau$ | $0.090 \pm 0.02$ | $0.090 \pm 0.02$ | 0.0 |

**Notes.** Comparison of cosmological parameters obtained with the standard five-year WMAP likelihood code (second column) and with the new GBR estimator at $\ell \leq 200$ (third column), given in terms of marginal means and standard deviations. The shift between the two in units of $\sigma$ is listed in the fourth column.

### 6. CONCLUSIONS

We have presented a new likelihood approximation to be used within the CMB Gibbs sampling framework. This approximation is defined by Gaussianizing the observed marginal power spectrum posteriors, $P(C_\ell | d)$, through a specific change-of-variables, and then coupling these univariate posteriors into a joint distribution through a multivariate Gaussian in the new variables. This process is exact, i.e., an identity operation, in the uniform and full-sky coverage case, and it is also an excellent approximation for the moderate sky cuts relevant to satellite missions such as WMAP and Planck.

Our new approach relies on the previously described CMB Gibbs sampling framework (Jewell et al. 2004; Wandelt et al. 2004; Eriksen et al. 2004), and thereby inherits many important advantages from that. First and foremost, this framework allows for seamless propagation of uncertainties from various systematic effects (e.g., foregrounds, beam uncertainties, calibration or noise estimation errors) to the final cosmological parameters. This is not straightforward in the hybrid scheme used by the WMAP code. Second, this new approach corresponds to the exact low-$\ell$ pixel space likelihood part of the WMAP code, not the approximate high-$\ell$ MASTER part. Still, our method can handle arbitrary high $\ell$s. Third, once the one-time pre-processing step has been completed, the computational expense of our estimator...
Figure 6. Comparison of likelihood slices from the standard WMAP likelihood code (dashed lines) and the new GBR likelihood (solid lines).

Figure 7. Comparison of marginal cosmological parameter posteriors obtained with the standard WMAP likelihood code (dashed lines) and with the modified GBR likelihood code up to $\ell = 200$ (solid line) for five-year WMAP data.

is determined by the cost of $\ell_{\text{max}}$ spline evaluations, while a pixel space approach requires a matrix inversion, and therefore scales as $O(N_{\text{pix}}^3)$. For the cases considered in this paper, the CPU time required for the GBR WMAP estimator up to $\ell = 200$ was $\sim 0.2$ ms, while it was $\sim 20$ s for the pixel space approach up to $\ell = 32$, for a map with 2500 pixels.

In order to validate our estimator, we applied it to a low-resolution simulated data set and compared it to slices through the exact joint likelihood as computed by brute-force evaluation in pixel space. The agreement between the two approaches was excellent. We then applied the same estimator to the five-year WMAP temperature data and estimated both a new power spectrum and new cosmological parameters within a standard six-parameter $\Lambda$CDM model.

The results from these calculations are interesting. First, our power spectrum is statistically very similar to the official WMAP spectrum, with no visible biases seen and relative fluctuations within the level predicted by noise and sky cut. Nevertheless, we do find significant differences in terms of cosmological parameters, and most notably in the spectral index of scalar perturbations, $n_s$. Specifically, we find $n_s = 0.973 \pm 0.014$, which is only $1.9\sigma$ away from unity and $0.6\sigma$ higher than the official WMAP result, $n_s = 0.965 \pm 0.014$. 
This result resembles very much the outcome of a re-analysis we did with the three-year WMAP temperature data (Eriksen et al. 2007b), for which we found a bias of 0.4σ in $n_s$ compared to the official WMAP results. This bias was due to the suboptimal MASTER-based likelihood approximation (Hivon et al. 2002; Verde et al. 2003) used by the WMAP team between $\ell = 12$ and 30, whereas we used an exact estimator in the same range. This study later prompted the WMAP to change their codes to use an exact likelihood evaluator up to $\ell = 30$.

In the same study, we also tried to increase the $\ell$-range for our exact estimator to $\ell = 50$, but found small differences. We therefore concluded, perhaps somewhat prematurely, that an exact estimator up to $\ell = 30$ was sufficient for obtaining accurate results. In contrast, in this paper we still find significant changes when increasing the exact estimator up to $\ell = 200$.

In retrospect, this should perhaps not come as a complete surprise, when realizing that the impact on a particular cosmological parameter typically depends logarithmically on $\ell$. For instance, Hamimeche & Lewis (2008) considered a simple power spectrum model with a single free amplitude, $C_\ell = q C_\ell^{\text{fid}}$, and found that, for a given likelihood estimator to be "statistically unbiased," the systematic errors in that same estimator must fall off faster than $\sim 1/\ell$.

A similar consideration holds for $n_s$. Intuitively, $n_s$ is as much affected by $\ell = 2–10$ as it is between $\ell = 20$ and 100. In the previous three-year WMAP re-analysis paper, we increased the range of the accurate likelihood estimator from $\ell = 12$ to 30, corresponding to a factor of 2.5 in $\ell$, and removed a bias of $\sim 0.4\sigma$ in $n_s$. In this paper, we increase the range from $\ell = 30$ to 200, corresponding to a factor of 6.7 in $\ell$, and find an additional bias of $0.6\sigma$. However, increasing $\ell$ from 30 to 50 corresponds only to a factor of 1.7 in $\ell$, and this appears to be too small to produce a statistically significant result.

The main conclusions from this work are twofold. First, it seems that an accurate likelihood description is required to higher $\ell$s than previously believed, and at least up to $\ell = 200$, in order to obtain unbiased results. By extrapolation, it also does not seem unlikely that even higher multipoles should be included. This issue will be re-visited in a future publication.

Our second main conclusion is that we find a spectral index only 1.9σ away from unity, namely $n_s = 0.973 \pm 0.014$. To us, it therefore seems premature to make strong claims concerning $n_s \neq 1$; the statistical significance of this is rather low, and there are likely still unknown systematic errors in this number.

In a future publication, we will generalize the GBR estimator to polarization. Once completed, this will enable a fully Gibbs-based CMB likelihood analysis at low $\ell$s and remove the need for likelihood techniques based on matrix operations, i.e., inversion and determinant evaluation. The computational cost of a standard cosmological parameter MCMC analysis (e.g., CosmoMC) will then once again be driven by the required Boltzmann codes (e.g., CAMB or CMBFast) and not by the likelihood evaluation. In turn, this will increase the importance of fast interpolation codes such as Pico (Fendt & Wandelt 2007) or COSMONET (Auld et al. 2007). With such fast algorithms for both spectrum and likelihood evaluations ready at hand, the CPU requirements for cosmological parameter estimation may possibly be reduced by orders of magnitude.

We thank Tony Banday, Ben Wandelt, and Graca Rocha for useful and interesting discussions. We acknowledge use of the HEALPix software (Górski et al. 2005) and analysis package for deriving the results in this paper. We acknowledge use of the LAMBDA. This work was partially performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Ø.R., N.E.G., and H.K.E. acknowledge financial support from the Research Council of Norway.

REFERENCES