

A MARKOV CHAIN MONTE CARLO ALGORITHM FOR ANALYSIS OF LOW SIGNAL-TO-NOISE COSMIC MICROWAVE BACKGROUND DATA

J. B. JEWELL^{1,2}, H. K. ERIKSEN^{3,4}, B. D. WANDELT^{5,6}, I. J. O'DWYER¹, GREG HUEY¹, AND K. M. GÓRSKI^{1,2,7}

¹ Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA; Jeffrey.B.Jewell@jpl.nasa.gov

² California Institute of Technology, Pasadena, CA 91125, USA

³ Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029, Blindern, N-0315 Oslo, Norway; h.k.k.eriksen@astro.uio.no

⁴ Centre of Mathematics for Applications, University of Oslo, P.O. Box 1053, Blindern, N-0316 Oslo, Norway

⁵ Department of Physics, University of Illinois, Urbana, IL 61801, USA

⁶ Astronomy Department, University of Illinois at Urbana-Champaign, IL 61801-3080, USA

⁷ Warsaw University Observatory, Aleje Ujazdowskie 4, 00-478 Warszawa, Poland

Received 2008 July 3; accepted 2009 March 2; published 2009 May 1

ABSTRACT

We present a new Markov Chain Monte Carlo (MCMC) algorithm for cosmic microwave background (CMB) analysis in the low signal-to-noise regime. This method builds on and complements the previously described CMB Gibbs sampler, and effectively solves the low signal-to-noise inefficiency problem of the direct Gibbs sampler. The new algorithm is a simple Metropolis–Hastings sampler with a general proposal rule for the power spectrum, C_ℓ , followed by a particular deterministic rescaling operation of the sky signal, \mathbf{s} . The acceptance probability for this joint move depends on the sky map only through the difference of χ^2 between the original and proposed sky sample, which is close to unity in the low signal-to-noise regime. The algorithm is completed by alternating this move with a standard Gibbs move. Together, these two proposals constitute a computationally efficient algorithm for mapping out the full joint CMB posterior, both in the high and low signal-to-noise regimes.

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

Online-only material: color figures

1. INTRODUCTION

Since the detection of anisotropy in the cosmic microwave background (CMB; Smoot et al. 1992), there has been an emphasis on likelihood or Bayesian methods for the inference of cosmological parameters and their error bars, or more generally, their confidence intervals. CMB analysis is most suitably addressed in a Bayesian, as opposed to frequentist, framework, simply because the observed microwave sky is interpreted as a single realization of a spatial random process.

Early measurements of the CMB were limited to signal-to-noise ratios (S/Ns) of order unity at relatively low angular scales, where direct evaluation of the likelihood for the power spectrum or cosmological parameters is possible. However, the $\mathcal{O}(N^3)$ scaling of computational expense with pixel number N prohibits direct likelihood evaluation for current and future CMB observations. Motivated by the scientific potential of CMB data with increasingly high spatial resolution, yet beset with systematics including partial sky coverage and foregrounds, an iterative method of sampling from the Bayes posterior, using a special case of Markov Chain Monte Carlo (MCMC) known as Gibbs sampling, was introduced by Jewell et al. (2002, 2004). The method was later independently discovered and applied to COBE data (Wandelt et al. 2004), numerically extended to high resolution on the sphere (Eriksen et al. 2004), applied to an analysis of the Wilkinson Microwave Anisotropy Probe (WMAP; Bennett et al. 2003; Hinshaw et al. 2007; Page et al. 2007) data (O'Dwyer et al. 2004; Eriksen et al. 2007a, 2007b), as well as generalized to include inference of foreground model parameters (Eriksen et al. 2008a, 2008b).

While Gibbs sampling provably converges to the Bayes posterior over the entire range of angular scales probed by the data, the run time required to generate enough independent samples at the low signal to noise, small angular scale regime

was found to be prohibitive (Eriksen et al. 2004). The reason for this is that typical variations in the power spectrum from one sample to the next are determined by cosmic variance alone, whereas the posterior itself is given by both cosmic variance and noise. This results in a long correlation length in the sequence of spectra in the low signal-to-noise regime, thus requiring a very long run time to generate a sufficient number of independent samples.

In this paper, we generalize the original Gibbs sampling algorithm to include a new type of MCMC step alternating with standard Gibbs sampling, which solves this problem of slow probabilistic convergence in the low signal-to-noise regime. This method therefore makes possible an exact Bayesian approach to CMB analysis over the entire range of angular scales probed by current and future experiments.

The paper is organized as follows. We first review the CMB Gibbs sampler, and describe the associated numerical difficulties in analysis at small angular scales. We then introduce the new MCMC step to the Markov chain, designed specifically to allow large variations in the high- ℓ CMB spectrum, precisely where the $S/N \leq 1$. We derive the required Metropolis–Hastings acceptance probability correctness in Appendix A, and numerically demonstrate the method in Section 4, for both temperature and polarization. Finally, we summarize and conclude in Section 5.

2. REVIEW OF GIBBS SAMPLING

2.1. The Joint Posterior

We begin by assuming that the observed data may be modeled by a signal and a noise term,

$$\mathbf{d} = \mathbf{A}\mathbf{s} + \mathbf{n}, \quad (1)$$

where \mathbf{d} is a vector containing the data (at every pointing of the detectors), the matrix \mathbf{A} involves both pointing and beam convolution (and where for this paper we will assume symmetric beams and neglect the details of this operation), and \mathbf{n} is additive noise (here in the pixel domain). We assume both the CMB signal and noise to be Gaussian random fields with vanishing mean and covariance matrices \mathbf{S} and \mathbf{N} , respectively. In harmonic space, where $\mathbf{s} = \sum_{\ell,m} a_{\ell m} Y_{\ell m}$, the CMB temperature covariance matrix is given by $C_{\ell m, \ell' m'} = \langle a_{\ell m}^* a_{\ell' m'} \rangle = C_\ell \delta_{\ell \ell'} \delta_{m m'}$, C_ℓ being the angular power spectrum. A generalization to polarization merely requires the replacement of the signal matrix diagonal elements with 3×3 matrices of the form

$$\mathbf{C}_l = \begin{bmatrix} C_l^{\text{TT}} & C_l^{\text{TE}} & C_l^{\text{TB}} \\ C_l^{\text{ET}} & C_l^{\text{EE}} & C_l^{\text{EB}} \\ C_l^{\text{BT}} & C_l^{\text{BE}} & C_l^{\text{BB}} \end{bmatrix}. \quad (2)$$

For the discussion in this section, we focus on the temperature case, but note that the generalization to polarization is straightforward and discussed by Larson et al. (2007).

Given these assumptions, our goal is to quantify what has been learned about the underlying power spectrum of the CMB given the data, or how well the data constrain the cosmological parameters. One proceeds then, in a Bayesian framework, by writing down the posterior given the data,

$$P(C_\ell | \mathbf{d}) \propto \mathcal{L}(\mathbf{d} | C_\ell) P(C_\ell). \quad (3)$$

Here, $\mathcal{L}(\mathbf{d} | C_\ell)$ is the likelihood and $P(C_\ell)$ is a prior on C_ℓ .

In order to derive the functional form of the likelihood, one imagines randomly choosing any relevant model (here a power spectrum drawn from $P(C_\ell)$), and asks what sequence of effects needs to be modeled in order to simulate the data. Here, simulation is understood as conditioning on the chosen model, and leads to a joint density

$$\begin{aligned} P(\mathbf{d}, \mathbf{s}, C_\ell) &= P(\mathbf{d}, \mathbf{s} | C_\ell) P(C_\ell) \\ &= P(\mathbf{d} | \mathbf{s}) P(\mathbf{s} | C_\ell) P(C_\ell) \end{aligned} \quad (4)$$

where the last line follows directly from our data model through the assumption of additive noise. Specifically, the factors in the above are

$$\begin{aligned} -2 \log P(\mathbf{s} | C_\ell) &= \mathbf{s}' \mathbf{C}^{-1} \mathbf{s} - \log |\mathbf{C}| \\ -2 \log P(\mathbf{d} | \mathbf{s}) &= -(\mathbf{d} - \mathbf{s})' \mathbf{N}^{-1} (\mathbf{d} - \mathbf{s}) - \log |\mathbf{N}| \end{aligned} \quad (5)$$

which follow from the assumption that both the signal and noise are independent Gaussian processes.

The idea of a “simulation chain” provides a conceptually clear approach to constructing a joint density, from which we immediately have the Bayesian posterior

$$P(C_\ell | \mathbf{d}) = \int d\mathbf{s} P(C_\ell, \mathbf{s} | \mathbf{d}). \quad (6)$$

The relevance of the above for this paper lies in relating what we refer to as the *joint posterior*, $P(C_\ell, \mathbf{s} | \mathbf{d})$, and the more familiar likelihood $\mathcal{L}(\mathbf{d} | C_\ell) \propto P(C_\ell | \mathbf{d}) / P(C_\ell)$.

Although we can analytically compute the integral of the joint posterior over the signal for the Gaussian signal and noise processes considered here, and therefore simply write down the functional form of the likelihood, it is too expensive to evaluate it for any specified C_l given high-resolution data. Furthermore, for more complicated data models (i.e., including foreground model

uncertainties) we will not be able to perform the integrals over the additional degrees of freedom. Both situations then instead motivate sampling from the joint posterior, and thereby generating samples from $P(C_\ell | \mathbf{d})$ without ever evaluating $P(C_\ell | \mathbf{d})$. We now discuss the original Gibbs sampling approach proposed and implemented by Jewell et al. (2004), Wandelt et al. (2004), and Eriksen et al. (2004), and then introduce a new MCMC step which directly addresses the previously reported slow probabilistic convergence in the low signal-to-noise regime (Eriksen et al. 2004).

2.2. The CMB Gibbs Sampler

As stated above, our goal is to sample from the joint posterior

$$-2 \log P(\mathbf{s}, C_\ell | \mathbf{d}) = \chi^2(\mathbf{d}, \mathbf{s}) + \mathbf{s}' \mathbf{S}^{-1} \mathbf{s} + \log |\mathbf{S}| + \log P(C_\ell). \quad (7)$$

For notational convenience, we have here dropped constant factors of 2π , and also defined

$$\chi^2(\mathbf{s}, \mathbf{d}) = (\mathbf{d} - \mathbf{s})' \mathbf{N}^{-1} (\mathbf{d} - \mathbf{s}). \quad (8)$$

One approach to sample from this posterior is to use an algorithm known as Gibbs sampling, where we can alternately sample from the respective conditional densities,

$$\mathbf{s}^{i+1} \leftarrow P(\mathbf{s} | C_\ell^i, \mathbf{d}) \quad (9)$$

$$C_\ell^{i+1} \leftarrow P(C_\ell | \mathbf{s}^{i+1}, \mathbf{d}). \quad (10)$$

Here, \leftarrow indicates sampling from the distribution on the right-hand side. After some burn-in period, during which all samples must be discarded, the joint samples (\mathbf{s}', C_ℓ') will be drawn from the desired density. Thus, the problem is reduced to that of sampling from the two *conditional* densities $P(\mathbf{s} | C_\ell, \mathbf{d})$ and $P(C_\ell | \mathbf{s}, \mathbf{d})$.

We now describe the sampling algorithms for each of these two conditional distributions, starting with $P(C_\ell | \mathbf{s}, \mathbf{d})$. First, note that $P(C_\ell | \mathbf{s}, \mathbf{d}) = P(C_\ell | \mathbf{s})$ which follows directly from the construction of the joint density of “everything” above. This is also intuitively easy to understand since if we already know the CMB sky signal, the data themselves tell us nothing new about the CMB power spectrum. Next, since the sky is assumed to be Gaussian and isotropic, the distribution reads

$$P(C_\ell | \mathbf{s}) \propto P(C_\ell) \frac{e^{-\frac{1}{2} \mathbf{s}' \mathbf{S}_\ell^{-1} \mathbf{s}_\ell}}{\sqrt{|\mathbf{S}_\ell|}} = P(C_\ell) \frac{e^{-\frac{2\ell+1}{2} \frac{\sigma_\ell}{C_\ell}}}{C_\ell^{\frac{2\ell+1}{2}}}, \quad (11)$$

which, when interpreted as a function of C_ℓ , is known as the inverse Gamma distribution. In this expression, $\sigma_\ell = \frac{1}{2\ell+1} \sum_m |a_{\ell m}|^2$ denotes the observed power spectrum of \mathbf{s} . Fortunately, there exists a simple textbook sampling algorithm for this distribution (e.g., Gupta & Nagar 2000), and we refer the interested reader to the previous papers for details. For an alternative, and more flexible, sampling algorithm, see Eriksen & Wehus (2009).

In order to describe the sky signal sampling step, we first define the mean-field map (or Wiener filtered data) to be $\hat{\mathbf{s}} = (\mathbf{S}^{-1} + \mathbf{N}^{-1})^{-1} \mathbf{N}^{-1} \mathbf{d}$, and note that the conditional sky signal density given the data and C_l can be written as

$$P(\mathbf{s} | C_\ell, \mathbf{d}) \propto e^{-\frac{1}{2} (\mathbf{s} - \hat{\mathbf{s}})' (\mathbf{S}^{-1} + \mathbf{N}^{-1}) (\mathbf{s} - \hat{\mathbf{s}})}. \quad (12)$$

Thus, $P(\mathbf{s} | C_\ell, \mathbf{d})$ is a Gaussian distribution with mean equals $\hat{\mathbf{s}}$ and a covariance matrix equals $(\mathbf{S}^{-1} + \mathbf{N}^{-1})^{-1}$.

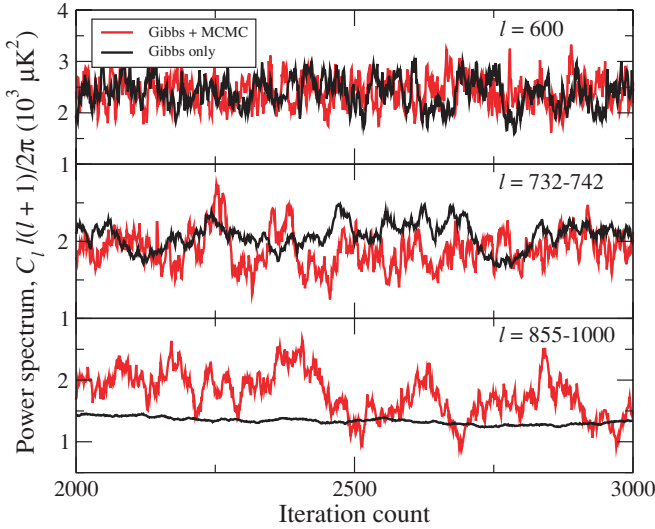


Figure 1. Comparison of C_ℓ chains produced by standard Gibbs sampling (black) and by the Gibbs+MCMC hybrid (red) for three selected multipole bins. The simulation was based on full sky coverage and uniform noise. See the text for full details.

(A color version of this figure is available in the online journal.)

Sampling from this Gaussian distribution is straightforward, but computationally somewhat cumbersome. First, draw two random white noise maps ω_0 and ω_1 with zero mean and unit variance. Then solve the equation

$$[\mathbf{S}^{-1} + \mathbf{N}^{-1}]\mathbf{s} = \mathbf{N}^{-1}\mathbf{d} + \mathbf{S}^{-\frac{1}{2}}\omega_0 + \mathbf{N}^{-\frac{1}{2}}\omega_1 \quad (13)$$

for \mathbf{s} . Since the white noise maps have zero mean, one immediately sees that $\langle \mathbf{s} \rangle = \hat{\mathbf{s}}$, while a few more calculations show that $\langle \mathbf{s}\mathbf{s}^T \rangle = (\mathbf{S}^{-1} + \mathbf{N}^{-1})^{-1}$.

The problematic part about this sampling step is the solution of the linear system in Equation (13). Since this is a $\sim 10^6 \times 10^6$ system for current CMB data sets, it cannot be solved by brute force. Instead, one must use a method called Conjugate Gradients (CG), which only requires multiplication of the coefficient matrix on the left-hand side, not inversion. For details on these computations, together with some ideas on preconditioning, see Eriksen et al. (2004).

2.3. Convergence Issues in the Low Signal-to-Noise Regime

As originally applied to high-resolution CMB data, the Gibbs sampling algorithm as described above has very slow convergence at the high- ℓ , low signal-to-noise part of the spectrum. The reason for the slow convergence is easy to understand in light of the above: when sampling from $P(C_\ell|\mathbf{s})$, the typical step size is given by cosmic variance at all angular scales. In the high signal-to-noise regime, cosmic variance dominates the noise variance, and we are able to explore the full width of the posterior in only a few Gibbs iterations. However, in the low signal-to-noise end, cosmic variance is far smaller than the posterior variance, and it takes a prohibitively long time to converge probabilistically. This problem of “slow mixing” of the Gibbs sampler is illustrated in Figures 1 and 2. The long correlation length starting at signal to noise of unity leads to extremely long run times in order to produce a reasonable number of uncorrelated samples.

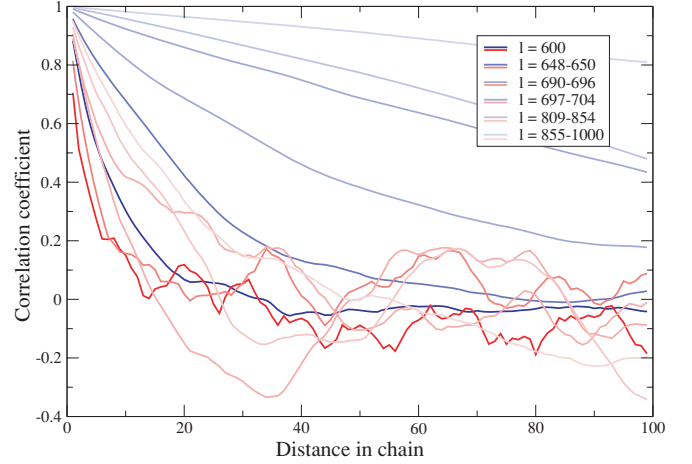


Figure 2. Comparison of chain correlation functions for standard Gibbs sampling (blue) and Gibbs+MCMC (red), computed from the full-sky uniform noise temperature data set. Note that while the correlation length goes to infinity with increasing ℓ (or equivalently, low signal to noise) for standard Gibbs sampling, it is $\lesssim 40$ everywhere for the MCMC hybrid case.

(A color version of this figure is available in the online journal.)

3. A LOW SIGNAL-TO-NOISE MCMC SAMPLER

When sampling from the true posterior, the goal is to produce as many independent samples from $P(C_\ell, \mathbf{s}|\mathbf{d})$ as possible. One might intuitively guess that it should be straightforward to establish good approximations to the posterior in the low signal-to-noise regime, since in the limit of vanishing signal to noise we simply recover the prior. This suggests that we look for a sampling scheme in which we first sample a new spectrum from some approximation to the true posterior independent on the current spectrum and CMB map, followed by sampling the CMB map from the conditional $P(\mathbf{s}|C_\ell, \mathbf{d})$. The problem with such a direct scheme is that the accept probability will involve a ratio of determinants which are too expensive to compute.

We are therefore motivated to look for a sampling scheme in which we can make a large variation in C_ℓ in the low signal-to-noise regime, and make an associated *deterministic change* in the CMB map, while still maintaining a reasonably high acceptance rate. The motivation for a deterministic change is that it will avoid introducing ratios of determinants which we cannot compute.

3.1. Proposal Rule and Acceptance Probability

Assume that we have defined a deterministic sampling scheme for \mathbf{s} , and that our new CMB map is given by some function

$$\mathbf{s}_{n+1} = F(\mathbf{s}_n, C_\ell^{(n+1)}, C_\ell^{(n)}). \quad (14)$$

Then the condition of detailed balance for our MCMC sampler requires that

$$F^{-1}(\mathbf{s}_{n+1}, C_\ell^{(n+1)}, C_\ell^{(n)}) = F(\mathbf{s}_{n+1}, C_\ell^{(n)}, C_\ell^{(n+1)}), \quad (15)$$

or, in other words, that the inverse function is given by exchanging the order of the spectra in the function F . One simple function which has this property is

$$\mathbf{s}_{n+1} = \left(\frac{C_\ell^{(n+1)}}{C_\ell^{(n)}} \right)^{\frac{1}{2}} \mathbf{s}_n. \quad (16)$$

The total proposal matrix is then

$$w(C_\ell^{(n+1)}, s_{n+1} | C_\ell^{(n)}, s_n) = w(C_\ell^{(n+1)} | C_\ell^{(n)}, \mathbf{d}) \delta \left(s_{n+1} - \left(\frac{C_\ell^{(n+1)}}{C_\ell^{(n)}} \right)^{-\frac{1}{2}} s_n \right),$$

and the “reverse” proposal is

$$w(C_\ell^{(n)}, s_n | C_\ell^{(n+1)}, s_{n+1}) = w(C_\ell^{(n)} | C_\ell^{(n+1)}, \mathbf{d}) \delta \left(s_n - \left(\frac{C_\ell^{(n)}}{C_\ell^{(n+1)}} \right)^{-\frac{1}{2}} s_{n+1} \right).$$

The condition of detailed balance including deterministic moves requires the consideration of some technical points which we leave to Appendix A. There we show that the full Metropolis–Hastings accept probability reads

$$A = \min \left[1, \frac{e^{-\chi^2(s_{n+1}, \mathbf{d})} w(C_\ell^{(n)} | C_\ell^{(n+1)}, \mathbf{d})}{e^{-\chi^2(s_n, \mathbf{d})} w(C_\ell^{(n+1)} | C_\ell^{(n)}, \mathbf{d})} \right]. \quad (17)$$

The significance of the above is that we can make relatively large changes to the power spectrum in the low signal-to-noise regime, where \mathbf{N}^{-1} is getting small, since the χ^2 is affected only very mildly by changes in any low signal-to-noise mode.

We note the interesting point (discussed more completely in Appendix B) that if one changes variables in the joint posterior from CMB maps, \mathbf{s} , to whitened maps, $\mathbf{x} = C_\ell^{-\frac{1}{2}} \mathbf{s}$, and then Gibbs sample in the new variables (C_ℓ, \mathbf{x}) , the resulting accept probability for changes in C_ℓ conditioned on \mathbf{x} is numerically identical to the above. While at first there might seem to be a conceptual distinction between MCMC schemes in a change of variables or those involving partial deterministic proposals, it turns out that this numerical equivalence holds more generally, and in Appendix B we establish a relation between both approaches. We bring this point to the attention of the reader as there could be other deterministic proposal schemes or another change of variables which lead to improvements over the approach presented in this paper.

For the numerical demonstration of the MCMC algorithm presented in this paper, we use a simple symmetric Gaussian proposal, truncated at $C_\ell > 0$ (or, for polarization, the region where the resulting CMB covariance matrix is positive definite), for the power spectrum,

$$w(C_\ell^{(n+1)} | C_\ell^{(n)}, \mathbf{d}) \propto e^{-\frac{1}{2} \left(\frac{C_\ell^{n+1} - C_\ell^n}{\tau_\ell} \right)^2} I(C_\ell > 0), \quad (18)$$

where τ_ℓ is a measure of the typical step size taken between two samples. Note that because this proposal density is symmetric, the ratio of C_ℓ proposals cancels, and the acceptance probability is entirely determined by the change in χ^2 .

It should be noted that while the above MCMC step satisfies detailed balance, it is not irreducible, in the sense that there is not a nonvanishing probability in reaching any state from any other state in a finite number of MCMC steps; the phases are unchanged in each MCMC step. However, alternating these steps with a traditional Gibbs sampling step gives a combined “two-step” MCMC algorithm which indeed is irreducible, and therefore provably converges to the joint posterior. Once again, the details are left to the Appendices for the interested reader.

3.2. Optimization of the MCMC Sampler

A general advantage of the Gibbs sampler is the fact that it is free of tunable efficiency parameters. The same is not true for the Metropolis–Hastings MCMC algorithm; for satisfactory sampling performance, it typically has to be tuned quite extensively. In this section, we describe three specific features that helps in this task, namely (1) step size tuning, (2) slice sampling, and (3) binning.

First, we have to ensure that the step size of our Gaussian proposal density roughly matches the width of the target distribution, in order to maintain both a reasonable acceptance rate and high mobility (for a general discussion of these issues and technical considerations of how to tune proposals, we refer the interested reader to Roberts & Rosenthal 2001). We do this by performing an initial test run, producing typically a few hundreds C_ℓ samples, and compute the standard deviation of these samples for each ℓ . These are then adopted as the proposal widths for the main run, scaled by some number less than unity, typically between 0.05 and 0.5. For the initial test run, we approximate the posterior width by the noise variance alone,

$$\tau_\ell^2 = \frac{2}{2\ell + 1} \frac{N_\ell}{b_\ell^2}, \quad (19)$$

because the MCMC sampler is used only in the low signal-to-noise regime. In this expression, N_ℓ is the power spectrum of the instrumental noise alone, and b_ℓ is the product of the Legendre transform of the beam and the HEALPix window function.

Next, the Metropolis–Hastings MCMC algorithm is inefficient in spaces with too many free parameters. For this reason, we divide the power spectrum coefficients, C_ℓ , into subsets, each containing typically only 10–20 multipoles. Then we propose changes to one subset at a time, while keeping all other multipoles fixed. Finally, we loop over subsets, and thus effectively implement a multipole slice Gibbs sampler for the full power spectrum.

This is computationally feasible, because a single MCMC proposal only requires a single χ^2 evaluation, which has a computational cost of a single spherical harmonic transform. Since drawing a full sky map from $P(\mathbf{s} | C_\ell, \mathbf{d})$ in the classical Gibbs sampling step requires $\mathcal{O}(10^2)$ spherical harmonic transforms, we can indeed afford to perform many MCMC proposals for each Gibbs step, without dominating the total cost.

Nevertheless, for very high resolution analysis, it is often beneficial to bin several C_ℓ together, both in order to increase the signal to noise of the joint coefficient, and to decrease the number of parameters that needs to be sampled by MCMC. We implement this by defining a new binned spectrum, weighted by $\ell(\ell + 1)/2\pi$, as follows:

$$C_b = \frac{1}{N_b} \sum_{\ell \in b} \frac{\ell(\ell + 1)}{2\pi} C_\ell. \quad (20)$$

Here, $b = [\ell_{\min}, \ell_{\max}]$ denotes the current bin, and $N_b = \ell_{\max} - \ell_{\min} + 1$ is the number of multipoles within the bin. These new (and fewer) coefficients are then sampled with the above MCMC sampler, after which the original spectrum coefficients are given by

$$C_\ell = \frac{2\pi}{\ell(\ell + 1)} C_b. \quad (21)$$

4. TESTING AND VALIDATION

We have implemented the new sampling step described above in the previously Gibbs sampling code called “Commander”

(Eriksen et al. 2004, 2008b), and in this section we demonstrate its advantages compared to the old sampling algorithm. We consider two different cases, namely high- ℓ temperature and low- ℓ polarization analysis. In the former case, we also analyze two cases, with and without a sky cut. The former allows us to verify the results against an analytically known answer, while the latter demonstrates that the sky cut does not degrade the sampling efficiency.

4.1. Temperature Analysis

The high- ℓ temperature simulation is designed to mimic the five-year *WMAP* temperature data (Hinshaw et al. 2009) with one exception, namely that the noise is assumed spatially uniform, in order to facilitate analytic comparison. Specifically, the CMB realization was drawn from the best-fit Λ CDM model derived from *WMAP* alone (Komatsu et al. 2009), including multipoles up to $\ell_{\text{max}} = 1000$, and then smoothed with the instrumental beam of the *WMAP* V1 differencing assembly, and pixelized at HEALPix⁸ resolution $N_{\text{side}} = 512$. Finally, uniform noise of $\sigma_0 = 40 \mu\text{K RMS}$ was added to each pixel. This corresponds to an S/N of unity at $\ell \sim 550$, roughly similar to the five-year *WMAP* data. We analyze this simulation both with and without the *WMAP* KQ85 sky cut (Gold et al. 2009).

In both analyses, we adopted the Gaussian proposal density with tuned variances, as described above. We also bin the power spectrum in progressively wide bins, starting at $\ell = 600$, to maintain a reasonable signal to noise per sampled power spectrum parameter. Ten bins were sampled jointly per proposal, while all others were kept fixed.

In the full-sky case, we produced a total of 31,800 samples over 60 chains, and in the cut-sky case a total of 6800 samples. The cost for producing one sample in the latter, and by far most expensive, set was 2.5 CPU hours, for a total of 17,000 CPU hours. The number of MCMC steps per Gibbs step was one in the former and 20 in the latter. (Since the signal sampler dominates the cut-sky Gibbs chain, one can perform more low S/N steps without slowing down the overall code significantly.) In addition to these two main sample sets, we also produced two longer chains with each 3500 samples for the full-sky case, both with and without the new MCMC step turned on, in order to compare the Markov chain correlation lengths before and after including the MCMC sampler.

We first consider the full-sky data set, and in Figure 1 we show a segment of each of the two longer chains for three selected multipole bins. The top panel shows $\ell = 600$, which is the first bin to be sampled by MCMC, the middle panel shows $\ell = 732\text{--}742$, where there is still some signal in the data, and, finally, the bottom panel shows $\ell = 855\text{--}1000$, which is strongly noise dominated. Starting with the top panel, we see that the red curve (Gibbs+MCMC) scatters significantly faster than the black curve (Gibbs only), implying more efficient sampling. This trend becomes even stronger with lower signal to noise, until the last case, where the Gibbs-only chain essentially does not move at all, while the MCMC sampler does probe the full range. Note, however, that even the MCMC sampler has a significant correlation length in this range, and this implies that there is still some room for improvement to be made in defining our proposals.

Next, these considerations are quantified in Figure 2, where we plot the Markov chain correlation length as a function of distance in the chain, for six bins with and without the MCMC

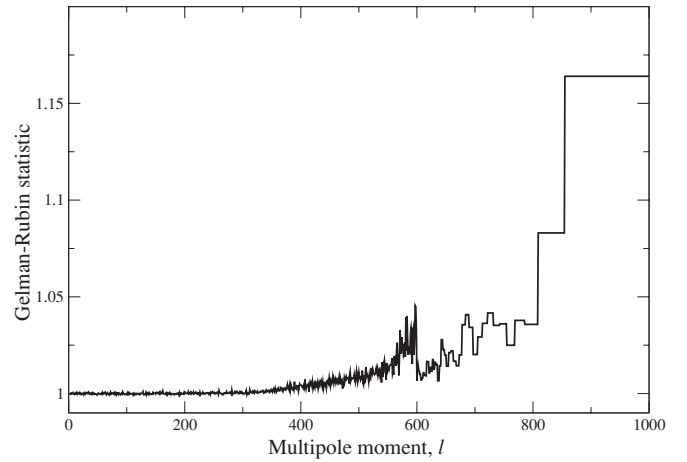


Figure 3. Gelman–Rubin statistic for the full-sky, uniform noise temperature analysis. Note the feature at $\ell = 600$, which marks the transition between standard Gibbs sampling and Gibbs+MCMC.

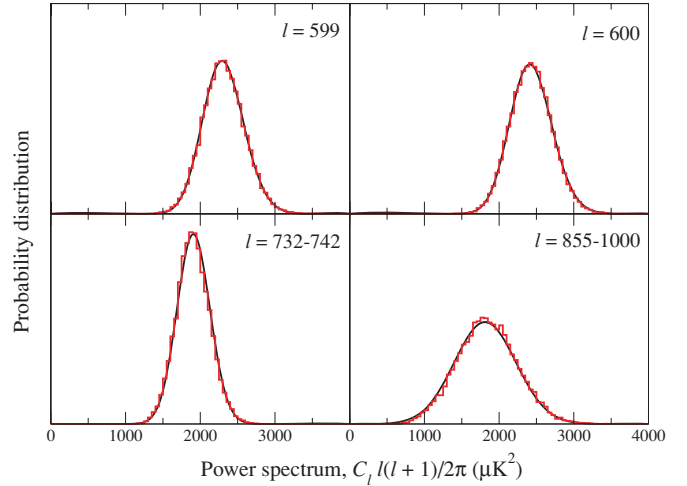


Figure 4. High- ℓ temperature marginal posteriors computed with Gibbs+MCMC from the full-sky, uniform noise temperature data set, compared to analytic results.

(A color version of this figure is available in the online journal.)

sampler. As first reported by Eriksen et al. (2004), we see that the Gibbs-only correlation length increases dramatically with decreasing signal to noise, rendering the algorithm essentially useless in this regime. However, we also see that the new MCMC step effectively resolves this issue, as the correlation length (here defined by having a correlation less than 0.2) now is less than ~ 40 steps. This is a dramatic improvement, and makes the algorithm useful even in this range. Nevertheless, we once again point out that it is possible to make further improvements by establishing better proposal densities.

In Figure 3, we consider the convergence properties of the $\sim 30\text{k}$ samples set, by computing the Gelman–Rubin statistic R (Gelman & Rubin 1992) as a function of ℓ . Typically, one recommends that R should be less than, say, 1.2 in order to claim convergence. We see that this holds everywhere for this sample set, and typically it is even less than 1.05. Note also the step at $\ell = 600$, showing clearly the beneficial effect of the MCMC sampler.

Next, in Figure 4, we compare the marginal distributions derived from this sample set with the analytic result,

⁸ <http://healpix.jpl.nasa.gov>

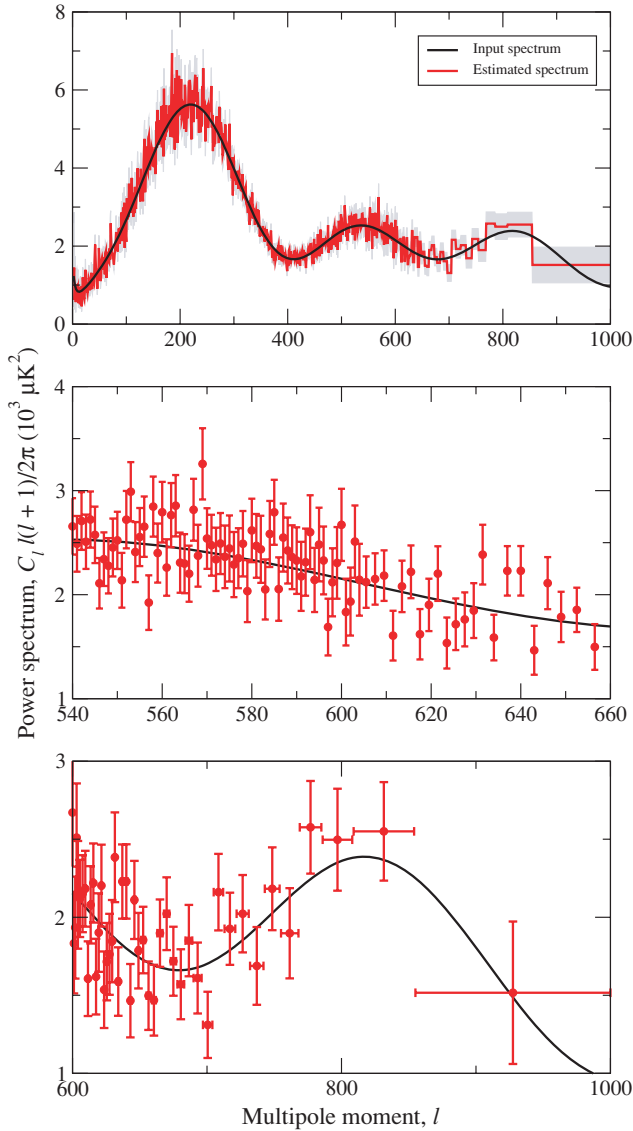


Figure 5. Temperature power spectrum estimated from cut-sky temperature data. The panels show the same spectrum, but emphasizing different multipole ranges (full-range; S/N ~ 1 transition region; and high- ℓ , low S/N).

(A color version of this figure is available in the online journal.)

$$P(C_\ell | \mathbf{d}) \propto \prod_{\ell \in b} \frac{e^{-\frac{2\ell+1}{2} \frac{\sigma_{\ell}^{S+N}}{b_\ell^2 C_\ell + N_\ell}}}{(b_\ell^2 C_\ell + N_\ell)^{\frac{2\ell+1}{2}}}. \quad (22)$$

Here, $b = [\ell_{min}, \ell_{max}]$ indicates a given multipole bin, b_ℓ denotes the product of the instrumental beam and the HEALPix pixel window, and σ_{ℓ}^{S+N} is the power spectrum of the noisy data map. We see that the new algorithm reproduces the analytic distributions very well, and this verifies the overall method.

Finally, the cut-sky power spectrum with one-sigma confidence regions is shown in three panels in Figure 5, focusing on different ℓ -ranges, namely all ℓ , the S/N ~ 1 transition region, and the low S/N region. This completes the high- ℓ temperature analysis validation.

4.2. Polarization Analysis

We now consider polarization analysis, and construct a new low- ℓ simulation for this purpose. This simulation does not mimic any planned experiment, but is rather designed to

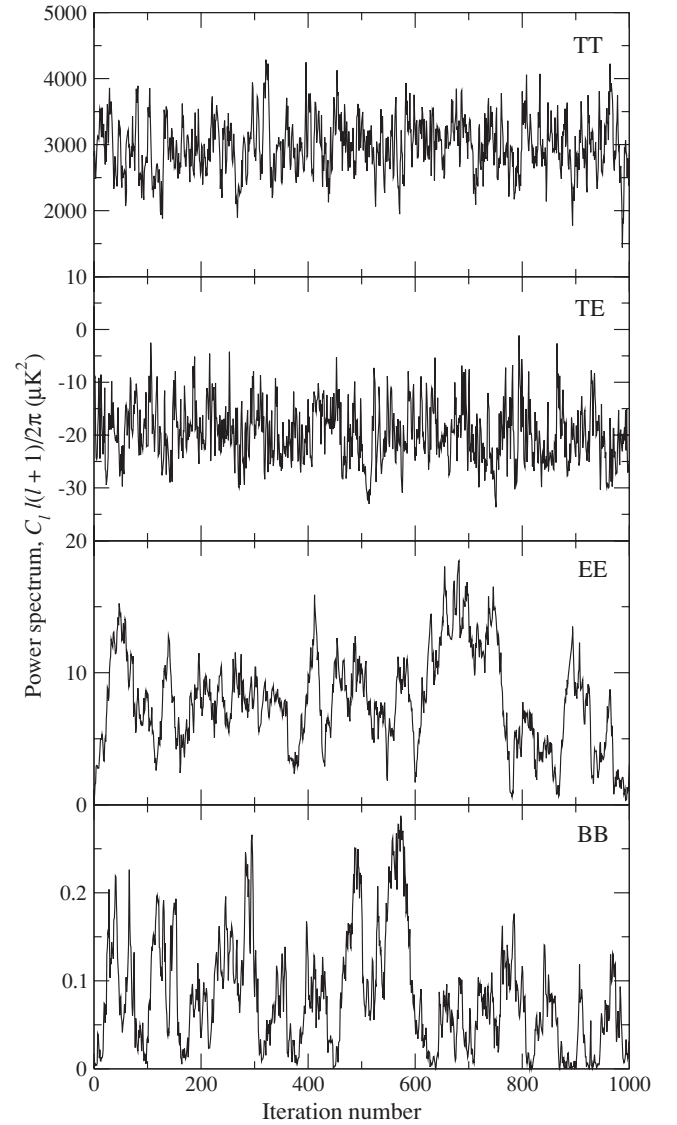


Figure 6. C_ℓ chains generated by Gibbs+MCMC hybrid for the cut-sky polarization data set. Only the highest multipole bin for each spectrum is shown ($\ell = 108$ –150 for TT, $\ell = 88$ –150 for TE, $\ell = 101$ –150 for EE, and $\ell = 61$ –150 for BB).

highlight the analysis method itself. Specifically, we drew a new CMB realization from the best-fit *WMAP* Λ CDM spectrum that includes a nonzero tensor contribution, including multipoles up to $\ell_{max} = 150$, and convolved this with a 3° FWHM Gaussian beam, and pixelized it at $N_{side} = 64$. Uniform noise of $5 \mu\text{K}$ RMS was added to the temperature component, and $1 \mu\text{K}$ RMS to the polarization components. The five-year *WMAP* polarization sky mask was imposed on the data.

We allowed for nonzero C_ℓ^{TT} , C_ℓ^{TE} , C_ℓ^{EE} , and C_ℓ^{BB} spectra, but fixed $C_\ell^{TB} = C_\ell^{EB} = 0$. These spectra were then individually binned to maintain a reasonable signal to noise per bin. (Details on how to introduce individual binning of each power spectrum were recently described by Eriksen & Wehus 2009.) Again, a tuned Gaussian proposal density was used in the MCMC step. A total of 12,000 samples were produced over 12 chains, and the CPU time per sample was 55 s, for a total of ~ 200 CPU hours.

In Figure 6, we show one C_ℓ chain for each of the four sampled spectra, for the last (and therefore most difficult) bin in each case. Note that the C_ℓ^{EE} and C_ℓ^{BB} spectra have essentially

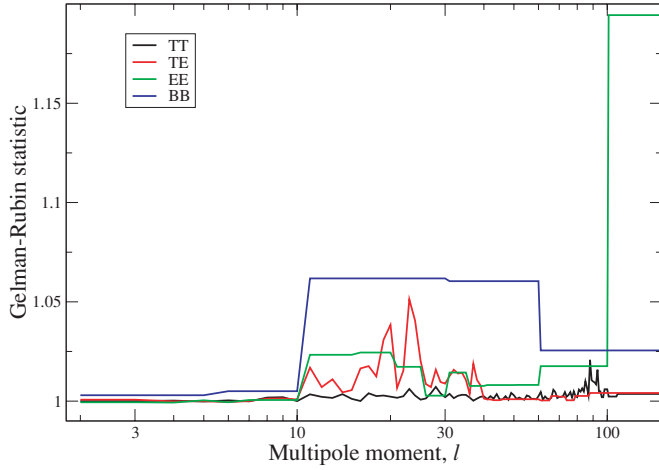


Figure 7. Gelman–Rubin statistic for cut-sky polarization analysis. (A color version of this figure is available in the online journal.)

vanishing signal to noise, and therefore these chains reach zero values. Clearly, we see that mixing properties of these chains are satisfactory, and the correlation lengths are quite short.

In Figure 7, we show the Gelman–Rubin statistics for each of the four power spectra, and with the single exception of the very last bin of C_ℓ^{EE} , all R values are well below 1.1. Thus, all spectra have converged well everywhere.

Finally, in Figure 8, we show the reconstructed marginal power spectra for each polarization component, overplotted on the input spectrum. The agreement is very good. Note, however, that these spectra are direct marginals, and not a joint maximum likelihood estimate. They are therefore not individual unbiased estimators. In particular, the marginal C_ℓ^{EE} power spectrum is biased slightly high because of the combination of the $C_\ell^{TT}C_\ell^{EE} - (C_\ell^{TE})^2 > 0$ positivity constraint and relatively low signal to noise. Consideration of the joint polarization posterior, which is an unbiased estimator, is postponed to a future publication.

5. CONCLUSIONS

We have presented a new MCMC algorithm for the high- ℓ , low signal-to-noise limit of the joint posterior which solves the slow probabilistic convergence of the traditional Gibbs sampler in this regime. This, in principle, allows sampling over the joint posterior $p(C_l, \mathbf{s}|\mathbf{d})$ over the entire range of angular scales probed by current and future CMB experiments. The limiting computational burden is now entirely in the map-making step of Gibbs sampling, for which the cost per Gibbs iteration now scales with the expense of multiplication by the inverse noise matrix \mathbf{N}^{-1} . Assuming pixel uncorrelated (but scan weighted) noise as a good approximation at small angular scales, the cost of an \mathbf{N}^{-1} multiplication is that of a forward and inverse spherical harmonic transform, or $\mathcal{O}(\ell_{\max}^3)$. Future work will attempt to push the generalized Gibbs+MCMC sampling scheme presented here to smaller angular scales, ultimately limited by the degree to which we can compute harmonic transforms.

We acknowledge use of the HEALPix⁹ software (Górski et al. 2005) and analysis package for deriving the results in this paper. H.K.E. acknowledges financial support from the Research Council of Norway. The research described in this paper was

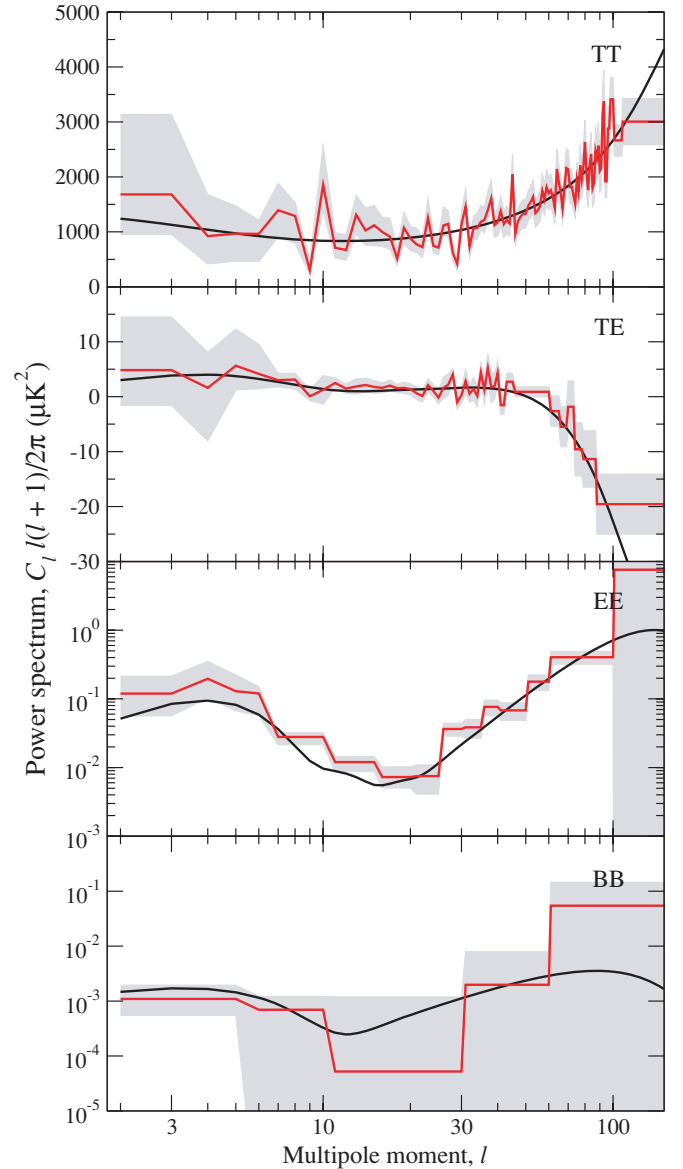


Figure 8. Marginal C_ℓ power spectra (red curves) estimated from cut-sky polarization data. Gray bands indicate 68% confidence regions, and the black lines show the input spectrum. (Note that the marginal spectra shown here are not individually unbiased estimators because of the correlations between TT, TE, and EE. Proper treatment of the full joint polarization density will be considered separately in a future publication.)

(A color version of this figure is available in the online journal.)

carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

APPENDIX A

INCLUDING DETERMINISTIC PROPOSALS IN MCMC

Here we review the derivation of the accept probability in MCMC when using deterministic proposals (or proposals where some of the degrees of freedom are specified as deterministic functions of the past state and/or proposed variations in some other degrees of freedom). We first briefly review the Metropolis–Hastings MCMC algorithm and the proof of its convergence, and then turn to the special case involving deterministic proposals. Much of the review of the MCMC

⁹ <http://healpix.jpl.nasa.gov>

algorithm here follows (Sokal 1989). We also note that similar technical considerations including deterministic elements in proposals are presented in Green (1995) in the context of MCMC algorithms in which the dimension of the state space itself is included as a random variable to be sampled over.

The goal is the construction of a transition matrix $T(C_l, \mathbf{s}|C'_l, \mathbf{s}', \mathbf{d})$ such that after initializing the Markov Chain with a sample from any probability density $p_0(C_l, \mathbf{s}|\mathbf{d})$, we generate samples from a sequence of probability densities

$$p_{n+1}(C_l, \mathbf{s}|\mathbf{d}) \equiv \int d(C'_l, \mathbf{s}') T(C_l, \mathbf{s}|C'_l, \mathbf{s}', \mathbf{d}) p_n(C'_l, \mathbf{s}'|\mathbf{d}) \quad (\text{A1})$$

which eventually converge to an *equilibrium density* $\pi(C_l, \mathbf{s}|\mathbf{d})$

$$\pi(C_l, \mathbf{s}|\mathbf{d}) = \lim_{n \rightarrow \infty} p_n(C_l, \mathbf{s}|\mathbf{d}). \quad (\text{A2})$$

We remind the reader of the sufficient conditions to establish convergence of an MCMC algorithm: *stationarity*, which means that the MCMC transition matrix satisfies

$$\pi(C_l, \mathbf{s}|\mathbf{d}) = \int d(C'_l, \mathbf{s}') T(C_l, \mathbf{s}|C'_l, \mathbf{s}', \mathbf{d}) \pi(C'_l, \mathbf{s}'|\mathbf{d}) \quad (\text{A3})$$

and irreducibility, which means that for any two states, there is a finite number of iterations which give a nonvanishing probability to transition from one state to the other. It is well known that these two properties are sufficient to establish convergence, as can be seen simply from the triangle inequality

$$\begin{aligned} & \int d(C_l, \mathbf{s}) |\pi(C_l, \mathbf{s}|\mathbf{d}) - p_n(C_l, \mathbf{s}|\mathbf{d})| \\ &= \int d(C_l, \mathbf{s}) \left| \int d(C'_l, \mathbf{s}') T(C_l, \mathbf{s}|C'_l, \mathbf{s}') (\pi(C'_l, \mathbf{s}'|\mathbf{d}) - p_{n-1}(C'_l, \mathbf{s}'|\mathbf{d})) \right| \\ &\leq \int d(C_l, \mathbf{s}) \int d(C'_l, \mathbf{s}') T(C_l, \mathbf{s}|C'_l, \mathbf{s}') |\pi(C'_l, \mathbf{s}'|\mathbf{d}) - p_{n-1}(C'_l, \mathbf{s}'|\mathbf{d})| \\ &= \int d(C'_l, \mathbf{s}') \left(\int d(C_l, \mathbf{s}) T(C_l, \mathbf{s}|C'_l, \mathbf{s}') \right) |\pi(C'_l, \mathbf{s}'|\mathbf{d}) - p_{n-1}(C'_l, \mathbf{s}'|\mathbf{d})| \\ &= \int d(C'_l, \mathbf{s}') |\pi(C'_l, \mathbf{s}'|\mathbf{d}) - p_{n-1}(C'_l, \mathbf{s}'|\mathbf{d})|. \end{aligned}$$

The Metropolis–Hastings MCMC algorithm is one method of constructing such a transition matrix. We choose *any* proposal matrix $w(C_l, \mathbf{s}|C'_l, \mathbf{s}', \mathbf{d})$ and then accept the proposed move with a probability

$$0 \leq A(C_l, \mathbf{s}|C'_l, \mathbf{s}', \mathbf{d}) \leq 1, \quad (\text{A4})$$

while rejecting the proposed move with probability $1 - A$ leads to a “null transition” where the next state in the Markov Chain remains the same. Application of this algorithm then leads to the sequence of probability densities which satisfy

$$\begin{aligned} p_{n+1}(C_l, \mathbf{s}|\mathbf{d}) &= \left(1 - \int d(C'_l, \mathbf{s}') A(C'_l, \mathbf{s}'|C_l, \mathbf{s}, \mathbf{d}) \right. \\ &\quad \times w(C'_l, \mathbf{s}'|C_l, \mathbf{s}, \mathbf{d}) \left. \right) p_n(C_l, \mathbf{s}|\mathbf{d}) \\ &\quad + \int d(C'_l, \mathbf{s}') A(C_l, \mathbf{s}|C'_l, \mathbf{s}', \mathbf{d}) \\ &\quad \times w(C_l, \mathbf{s}|C'_l, \mathbf{s}', \mathbf{d}) p_n(C'_l, \mathbf{s}'|\mathbf{d}), \quad (\text{A5}) \end{aligned}$$

where the first term is the contribution to the probability density p_{n+1} if we reject any proposed move, while the second term is the contribution from accepting the proposed move from any possible previous state. If we demand that, for a chosen proposal matrix, the accept probability satisfies

$$\begin{aligned} & \pi(C'_l, \mathbf{s}'|\mathbf{d}) w(C_l, \mathbf{s}|C'_l, \mathbf{s}', \mathbf{d}) A(C_l, \mathbf{s}|C'_l, \mathbf{s}', \mathbf{d}) \\ &= A(C'_l, \mathbf{s}'|C_l, \mathbf{s}, \mathbf{d}) w(C'_l, \mathbf{s}'|C_l, \mathbf{s}, \mathbf{d}) \pi(C_l, \mathbf{s}|\mathbf{d}) \quad (\text{A6}) \end{aligned}$$

then we see that the Metropolis–Hastings MCMC algorithm satisfies stationarity, i.e., denoting by $T \circ \pi$ the density resulting from one application of the transition matrix to π , we have directly from detailed balance

$$\begin{aligned} T \circ \pi &= \left(1 - \int d(C'_l, \mathbf{s}') A(C'_l, \mathbf{s}'|C_l, \mathbf{s}, \mathbf{d}) w(C'_l, \mathbf{s}'|C_l, \mathbf{s}, \mathbf{d}) \right) \\ &\quad \times \pi(C_l, \mathbf{s}|\mathbf{d}) + \int d(C'_l, \mathbf{s}') A(C_l, \mathbf{s}|C'_l, \mathbf{s}', \mathbf{d}) \\ &\quad \times w(C_l, \mathbf{s}|C'_l, \mathbf{s}', \mathbf{d}) \pi(C'_l, \mathbf{s}'|\mathbf{d}) \\ &= \left(1 - \int d(C'_l, \mathbf{s}') A(C'_l, \mathbf{s}'|C_l, \mathbf{s}, \mathbf{d}) w(C'_l, \mathbf{s}'|C_l, \mathbf{s}, \mathbf{d}) \right) \\ &\quad \times \pi(C_l, \mathbf{s}|\mathbf{d}) + \pi(C_l, \mathbf{s}|\mathbf{d}) \int d(C'_l, \mathbf{s}') A(C'_l, \mathbf{s}'|C_l, \mathbf{s}, \mathbf{d}) \\ &\quad \times w(C'_l, \mathbf{s}'|C_l, \mathbf{s}, \mathbf{d}) \\ &= \pi(C_l, \mathbf{s}|\mathbf{d}). \quad (\text{A7}) \end{aligned}$$

We now turn to the case where our proposal is of the form

$$w(\mathbf{s}', C'_l|\mathbf{s}, C_l) = \delta[\mathbf{s}' - F(\mathbf{s}, C'_l, C_l)] w(C'_l|C_l, \mathbf{d}), \quad (\text{A8})$$

where we randomly propose a new power spectrum, possibly in a manner conditionally dependent on the current spectrum and the data, and then deterministically compute a new CMB map with some function

$$\mathbf{s}' = F(\mathbf{s}, C'_l, C_l). \quad (\text{A9})$$

To satisfy detailed balance with a nonvanishing accept probability, our function must satisfy

$$\begin{aligned} \mathbf{s}' &= F(\mathbf{s}, C'_l, C_l) \\ \mathbf{s} &= F(\mathbf{s}', C_l, C'_l) \quad (\text{A10}) \end{aligned}$$

or, that the inverse function is equivalent to interchanging the order of the power spectrum arguments

$$F(\mathbf{s}', C_l, C'_l) = F^{-1}(\mathbf{s}', C'_l, C_l). \quad (\text{A11})$$

In this paper, we have chosen one such function, given by

$$F(\mathbf{s}, C'_l, C_l) = [\mathbf{C}']^{1/2} [\mathbf{C}]^{-1/2} \mathbf{s}, \quad (\text{A12})$$

where interchanging the spectra in the function above does in fact give the inverse function itself.

Our job now is to *derive* the accept probability such that we satisfy stationarity (as discussed above). For the proposal with deterministic changes to some of the degrees of freedom, stationarity is satisfied if

$$\begin{aligned} (T \circ \pi)(C_l, \mathbf{s}|\mathbf{d}) &= \left[1 - \int d(C_l'', \mathbf{s}'') A[C_l'', \mathbf{s}''|C_l, \mathbf{s}] \delta[\mathbf{s}'' - F(\mathbf{s}, C_l'', C_l)] w(C_l''|C_l, \mathbf{d}) \right] \pi(C_l, \mathbf{s}|\mathbf{d}) \\ &\quad + \int d(C'_l, \mathbf{s}') A[\mathbf{s}, C_l|\mathbf{s}', C'_l] \delta[\mathbf{s} - F(\mathbf{s}', C_l, C'_l)] w(C_l|C'_l, \mathbf{d}) \pi(C'_l, \mathbf{s}'|\mathbf{d}). \end{aligned}$$

In order to determine the integral over the δ -function in the accept term above, we recall the identity for $\delta[G(\mathbf{x})]$, where $G(\mathbf{a}) = 0$,

$$\delta[G(\mathbf{x})] = \frac{\delta(\mathbf{x} - \mathbf{a})}{|\partial G / \partial \mathbf{x}|_a} \quad (\text{A13})$$

In our case, we can identify

$$G(\mathbf{s}') = \mathbf{s} - F(\mathbf{s}', C_l, C_l') \quad (\text{A14})$$

which vanishes at $F^{-1}(\mathbf{s}, C_l, C_l') = F(\mathbf{s}, C_l', C_l)$. We also have the Jacobian

$$\left| \frac{\partial G}{\partial \mathbf{s}'} \right|_{\mathbf{s}'=F^{-1}(\mathbf{s}, C_l, C_l')} = \left| \frac{\partial F}{\partial \mathbf{s}'} \right|_{\mathbf{s}'=F^{-1}(\mathbf{s}, C_l, C_l')} \quad (\text{A15})$$

(i.e., $G(\mathbf{s}')$ is considered a function of \mathbf{s}' with the other CMB map \mathbf{s} considered fixed) which therefore gives

$$\delta[\mathbf{s} - F(\mathbf{s}', C_l, C_l')] = \delta[\mathbf{s}' - F^{-1}(\mathbf{s}, C_l, C_l')] \left| \frac{\partial F}{\partial \mathbf{s}'} \right|_{\mathbf{s}'=F^{-1}(\mathbf{s}, C_l, C_l')}^{-1}. \quad (\text{A16})$$

Inserting this into the condition for stationarity, we have

$$\begin{aligned} (T \circ \pi)(C_l, \mathbf{s}|\mathbf{d}) &= \left[1 - \int d(C_l, '' s'') A[C_l, '' s''|C_l, \mathbf{s}] \delta[\mathbf{s}'' \right. \\ &\quad \left. - F(\mathbf{s}, C_l, '' C_l)] w(C_l''|C_l, \mathbf{d}) \right] \pi(C_l, \mathbf{s}|\mathbf{d}) \\ &\quad + \int d(C_l', s') A[\mathbf{s}, C_l|s', C_l'] \left(\delta[\mathbf{s}' \right. \\ &\quad \left. - F^{-1}(\mathbf{s}, C_l, C_l')] \left| \frac{\partial F}{\partial \mathbf{s}'} \right|_{\mathbf{s}'=F^{-1}(\mathbf{s}, C_l, C_l')}^{-1} \right) \\ &\quad \times w(C_l|C_l', \mathbf{d}) \pi(C_l', \mathbf{s}'|\mathbf{d}) \\ &= \left[1 - \int d(C_l, '' s'') A[C_l, '' s''|C_l, \mathbf{s}] \delta[\mathbf{s}'' \right. \\ &\quad \left. - F(\mathbf{s}, C_l, '' C_l)] w(C_l''|C_l, \mathbf{d}) \right] \pi(C_l, \mathbf{s}|\mathbf{d}) \\ &\quad + \int d(C_l', s') A[\mathbf{s}, C_l|s', C_l'] \left(\delta[\mathbf{s}' \right. \\ &\quad \left. - F(\mathbf{s}, C_l', C_l)] \left| \frac{\partial F}{\partial \mathbf{s}'} \right|_{\mathbf{s}'=F^{-1}(\mathbf{s}, C_l, C_l')}^{-1} \right) \\ &\quad \times w(C_l|C_l', \mathbf{d}) \pi(C_l', \mathbf{s}'|\mathbf{d}) \end{aligned}$$

where in the second line we again used the property that the inverse F^{-1} is equivalent to F with the spectra arguments interchanged. We see from the above that a sufficient condition for stationarity is

$$\begin{aligned} \pi(C_l, \mathbf{s}|\mathbf{d}) w(C_l'|C_l, \mathbf{d}) A[\mathbf{s}', C_l'|\mathbf{s}, C_l] &= A[\mathbf{s}, C_l|s', C_l'] \\ &\times \left(\left| \frac{\partial F}{\partial \mathbf{s}'} \right|_{\mathbf{s}'=F^{-1}(\mathbf{s}, C_l, C_l')}^{-1} \right) w(C_l|C_l', \mathbf{d}) \pi(C_l', \mathbf{s}'|\mathbf{d}). \end{aligned} \quad (\text{A17})$$

An accept probability which satisfies this condition therefore gives cancellation of the integrals over the δ -functions for both the reject and accept contributions, leaving us exactly with $T \circ \pi = \pi$. We therefore have the accept

probability

$$\begin{aligned} A[\mathbf{s}', C_l'|\mathbf{s}, C_l] &= \min \left[1, \frac{\pi(C_l', \mathbf{s}'|\mathbf{d}) w(C_l|C_l', \mathbf{d})}{\pi(C_l, \mathbf{s}|\mathbf{d}) w(C_l'|C_l, \mathbf{d})} \right. \\ &\quad \left. \times \left(\left| \frac{\partial F}{\partial \mathbf{s}'} \right|_{\mathbf{s}'=F^{-1}(\mathbf{s}, C_l, C_l')}^{-1} \right) \right]. \end{aligned} \quad (\text{A18})$$

We give the expression above for the general case of any deterministic change in the CMB map with a function which satisfies $F(\mathbf{s}, C_l, C_l') = F^{-1}(\mathbf{s}, C_l', C_l)$. We now explicitly evaluate this accept probability for the functional form chosen for this paper.

Since we have $F(\mathbf{s}', C_l, C_l') = [\mathbf{C}]^{1/2}[\mathbf{C}']^{-1/2}\mathbf{s}'$, we have

$$\left(\left| \frac{\partial F}{\partial \mathbf{s}'} \right|_{\mathbf{s}'=F^{-1}(\mathbf{s}, C_l, C_l')} \right) = \frac{|\mathbf{C}|^{1/2}}{|\mathbf{C}'|^{1/2}}. \quad (\text{A19})$$

Reminding the reader of the functional form of the joint posterior in Equation (7), we have the accept probability given by

$$\begin{aligned} A[\mathbf{s}', C_l'|\mathbf{s}, C_l] &= \min \left[1, \frac{\pi(C_l', \mathbf{s}'|\mathbf{d}) w(C_l|C_l', \mathbf{d})}{\pi(C_l, \mathbf{s}|\mathbf{d}) w(C_l'|C_l, \mathbf{d})} \right. \\ &\quad \left. \times \left(\left| \frac{\partial F}{\partial \mathbf{s}'} \right|_{\mathbf{s}'=F^{-1}(\mathbf{s}, C_l, C_l')}^{-1} \right) \right] \\ &= \min \left[1, \frac{e^{-\chi^2(\mathbf{s}', \mathbf{d})} e^{-\mathbf{s}'[\mathbf{C}]^{-1}\mathbf{s}'}}{e^{-\chi^2(\mathbf{s}, \mathbf{d})} e^{-\mathbf{s}[\mathbf{C}]^{-1}\mathbf{s}}} \frac{|\mathbf{C}|^{1/2}}{|\mathbf{C}'|^{1/2}} \frac{w(C_l|C_l', \mathbf{d})}{w(C_l'|C_l, \mathbf{d})} \right. \\ &\quad \left. \times \left(\left| \frac{\partial F}{\partial \mathbf{s}'} \right|_{\mathbf{s}'=F^{-1}(\mathbf{s}, C_l, C_l')}^{-1} \right) \right] \\ &= \min \left[1, \frac{e^{-\chi^2(\mathbf{s}', \mathbf{d})} e^{-\mathbf{s}'[\mathbf{C}]^{-1}\mathbf{s}'}}{e^{-\chi^2(\mathbf{s}, \mathbf{d})} e^{-\mathbf{s}[\mathbf{C}]^{-1}\mathbf{s}}} \frac{|\mathbf{C}|^{1/2}}{|\mathbf{C}'|^{1/2}} \frac{w(C_l|C_l', \mathbf{d})}{w(C_l'|C_l, \mathbf{d})} \right. \\ &\quad \left. \times \left(\frac{|\mathbf{C}'|^{1/2}}{|\mathbf{C}|^{1/2}} \right) \right] \\ &= \min \left[1, \frac{e^{-\chi^2(\mathbf{s}', \mathbf{d})} w(C_l|C_l', \mathbf{d})}{e^{-\chi^2(\mathbf{s}, \mathbf{d})} w(C_l'|C_l, \mathbf{d})} \right], \end{aligned} \quad (\text{A20})$$

where the last line follows from the invariance of the quadratic form under the functional mapping $\mathbf{s}'[\mathbf{C}]^{-1}\mathbf{s}' = \mathbf{s}[\mathbf{C}]^{-1}\mathbf{s}$. Finally, we note that for the special case of a symmetric proposal matrix where $w(C_l'|C_l, \mathbf{d}) = w(C_l|C_l', \mathbf{d})$, the accept probability is completely determined by the (exponentiated) change in χ^2

$$A[\mathbf{s}', C_l'|\mathbf{s}, C_l] = \min \left[1, \frac{e^{-\chi^2(\mathbf{s}', \mathbf{d})}}{e^{-\chi^2(\mathbf{s}, \mathbf{d})}} \right]. \quad (\text{A21})$$

As emphasized earlier in the main part of the text, the above allows large changes to the spectrum precisely where the signal to noise is getting small, as χ^2 does not change much in this regime.

APPENDIX B

RELATION TO GIBBS SAMPLING IN A CHANGE OF VARIABLES

We note here another interesting approach to an MCMC algorithm in a *different set of variables* which in fact allows for

large moves in the spectrum in the low signal-to-noise regime. We define the CMB map

$$\mathbf{x} = \mathbf{C}^{-1/2} \mathbf{s}. \quad (\text{B1})$$

We therefore have the joint posterior *in the new variables* according to

$$p(C_l, \mathbf{s} | \mathbf{d}) d(C_l, \mathbf{s}) = p(C_l, \mathbf{x} | \mathbf{d}) \left| \frac{\partial \mathbf{s}}{\partial \mathbf{x}} \right| d(C_l, \mathbf{x}) \quad (\text{B2})$$

which is explicitly, up to a normalization constant

$$-2 \log p(C_l, \mathbf{x} | \mathbf{d}) = (\mathbf{d} - \mathbf{C}^{1/2} \mathbf{x})^T \mathbf{N}^{-1} (\mathbf{d} - \mathbf{C}^{1/2} \mathbf{x}) - \|\mathbf{x}\|^2. \quad (\text{B3})$$

Then *traditional Gibbs sampling in the new variables* leads to an accept probability when changing the spectrum given the change of variable map \mathbf{x} as

$$A(C_l', \mathbf{x} | C_l, \mathbf{x}) = \min \left[1, \frac{e^{-(\mathbf{d} - [\mathbf{C}']^{1/2} \mathbf{x})^T \mathbf{N}^{-1} (\mathbf{d} - [\mathbf{C}']^{1/2} \mathbf{x})}}{e^{-(\mathbf{d} - \mathbf{C}^{1/2} \mathbf{x})^T \mathbf{N}^{-1} (\mathbf{d} - \mathbf{C}^{1/2} \mathbf{x})}} \times \frac{w(C_l | \mathbf{x}, C_l', \mathbf{d})}{w(C_l' | \mathbf{x}, C_l, \mathbf{d})} \right], \quad (\text{B4})$$

where in the above the proposed variation in the spectrum can now be conditionally dependent on the current change of variable map \mathbf{x} . Assuming a symmetric proposal, or one conditionally independent of \mathbf{x} leads to an accept probability which is *numerically the same as* (A20), and also has the same property—large moves in the spectrum are possible in the low signal-to-noise regime. As a side note, we can see that $\log p(C_l | \mathbf{x}, \mathbf{d})$ is quadratic in $C_l^{1/2}$, and suggests a proposal given by a Gaussian in $C_l^{1/2}$. However, there are two problems with this scheme—sampling in $C_l^{1/2}$ will result in reintroducing a Jacobian factor given by the ratio of $|C'|^{1/2} / |C|^{1/2}$ which results typically in low acceptance probabilities, and furthermore we cannot afford to exactly compute the local “Fisher” covariance matrix for each \mathbf{x} . Because of these difficulties, we in general need to produce a proposal for C_l and then compute the accept probability above.

We now explore the relation between MCMC steps which involve deterministically changing some of the degrees of freedom in the original variables and conditional steps in a new set of variables. It turns out that a change of variables can always be used to generate MCMC steps in the original variables where some of the degrees of freedom are changed deterministically, and where the accept probabilities are numerically equivalent.

For notational convenience, we will assume the state space is separated into two sets of variables (\mathbf{x}, \mathbf{y}) , i.e., for the CMB sampling context we have (\mathbf{s}, C_l) . Now, consider a “global” change of variables of the form

$$\begin{aligned} \mathbf{u} &= F(\mathbf{x}, \mathbf{y}) \\ \mathbf{v} &= \mathbf{y} \end{aligned} \quad (\text{B5})$$

with Jacobian given in block form (i.e., see Gantmakher 1959)

$$\left| \begin{array}{cc} \frac{\partial \mathbf{u}}{\partial \mathbf{x}} & \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \\ \frac{\partial \mathbf{u}}{\partial \mathbf{y}} & \frac{\partial \mathbf{v}}{\partial \mathbf{y}} \end{array} \right| = \left| \begin{array}{cc} \frac{\partial F}{\partial \mathbf{x}} & 0 \\ \frac{\partial F}{\partial \mathbf{y}} & \mathbf{1} \end{array} \right| = \left| \frac{\partial F}{\partial \mathbf{x}} \right|. \quad (\text{B6})$$

A Gibbs sampling step varying \mathbf{v} with \mathbf{u} fixed, has accept probability

$$A(\mathbf{v}_{n+1}, \mathbf{u}_n | \mathbf{v}_n, \mathbf{u}_n) = \min \left[1, \frac{\pi(\mathbf{v}_{n+1}, \mathbf{u}_n | \mathbf{d})}{\pi(\mathbf{v}_n, \mathbf{u}_n | \mathbf{d})} \frac{w(\mathbf{v}_n | \mathbf{v}_{n+1}, \mathbf{u}_n, \mathbf{d})}{w(\mathbf{v}_{n+1} | \mathbf{v}_n, \mathbf{u}_n, \mathbf{d})} \right],$$

where in the last line we have expressed the numerical value of the accept probability in terms of the original variables (\mathbf{x}, \mathbf{y}) , and where in the above we have the constraint

$$\begin{aligned} \mathbf{x}_{n+1} &= F^{-1}(\mathbf{u}_n, \mathbf{y}_{n+1}), \\ \mathbf{x}_n &= F^{-1}(\mathbf{u}_n, \mathbf{y}_n). \end{aligned} \quad (\text{B7})$$

Now consider an MCMC step in the original variables of the form

$$w(\mathbf{x}_{n+1}, \mathbf{y}_{n+1} | \mathbf{x}_n, \mathbf{y}_n) = w(\mathbf{y}_{n+1} | \mathbf{y}_n, \mathbf{x}_n, \mathbf{d}) \times \delta(\mathbf{x}_{n+1} - H(\mathbf{x}_n, \mathbf{y}_{n+1}, \mathbf{y}_n)) \quad (\text{B8})$$

with general accept probability, according to the discussion above

$$\begin{aligned} A(\mathbf{y}_{n+1}, \mathbf{x}_{n+1} | \mathbf{y}_n, \mathbf{x}_n) &= \min \left[1, \frac{\pi(\mathbf{y}_{n+1}, \mathbf{x}_{n+1} | \mathbf{d})}{\pi(\mathbf{y}_n, \mathbf{x}_n | \mathbf{d})} \times \frac{w(\mathbf{y}_n | \mathbf{x}_n, \mathbf{y}_{n+1}, \mathbf{d})}{w(\mathbf{y}_{n+1} | \mathbf{x}_{n+1}, \mathbf{y}_n, \mathbf{d})} \right. \\ &\quad \left. \times \left(\left| \frac{\partial H(\mathbf{x}, \mathbf{y}_n, \mathbf{y}_{n+1})}{\partial \mathbf{x}} \right|^{-1} \right)_{\mathbf{x}_{n+1} = H^{-1}(\mathbf{x}_n, \mathbf{y}_n, \mathbf{y}_{n+1})} \right]. \end{aligned} \quad (\text{B9})$$

Interestingly enough this suggests that we can set H to be the function

$$H(\mathbf{x}, \mathbf{y}_{n+1}, \mathbf{y}_n) = F^{-1}(F(\mathbf{x}, \mathbf{y}_n), \mathbf{y}_{n+1}). \quad (\text{B10})$$

Does this function have the correct properties for its inverse? Assuming we have computed in the forward direction $\mathbf{x}' = H(\mathbf{x}, \mathbf{y}_{n+1}, \mathbf{y}_n)$, we can invert to find \mathbf{x} by computing sequentially

$$\begin{aligned} F(\mathbf{x}', \mathbf{y}_{n+1}) &= F(\mathbf{x}, \mathbf{y}_n) \\ \mathbf{x} &= F^{-1}(F(\mathbf{x}', \mathbf{y}_{n+1}), \mathbf{y}_n) \\ &\equiv H(\mathbf{x}', \mathbf{y}_n, \mathbf{y}_{n+1}) \end{aligned} \quad (\text{B11})$$

where the last line follows from definition of the forward H . Since we have, by definition

$$\begin{aligned} \mathbf{x}' &= H(\mathbf{x}, \mathbf{y}_{n+1}, \mathbf{y}_n), \\ \mathbf{x} &\equiv H^{-1}(\mathbf{x}', \mathbf{y}_{n+1}, \mathbf{y}_n), \end{aligned}$$

we therefore have shown that

$$H^{-1}(\mathbf{x}', \mathbf{y}_{n+1}, \mathbf{y}_n) = H(\mathbf{x}', \mathbf{y}_n, \mathbf{y}_{n+1}) \quad (\text{B12})$$

as required for a nonvanishing accept probability. To numerically evaluate the accept probability involving the deterministic move, we need to compute the Jacobian $H(\mathbf{x}, \mathbf{y}_n, \mathbf{y}_{n+1}) = F^{-1}(F(\mathbf{x}, \mathbf{y}_{n+1}), \mathbf{y}_n)$, which is

$$\begin{aligned} \left| \frac{\partial H(\mathbf{x}, \mathbf{y}_n, \mathbf{y}_{n+1})}{\partial \mathbf{x}} \right|^{-1} &= \left(\left| \frac{\partial F^{-1}}{\partial \mathbf{u}} \right|_{(\mathbf{u}(\mathbf{x}_{n+1}, \mathbf{y}_{n+1}), \mathbf{y}_n)} \left| \frac{\partial F}{\partial \mathbf{x}} \right|_{(\mathbf{x}_{n+1}, \mathbf{y}_{n+1})} \right)^{-1} \\ &= \left(\left| \frac{\partial(\mathbf{y}, \mathbf{x})}{\partial(\mathbf{u}, \mathbf{v})} \right|_{(\mathbf{u}_{n+1}, \mathbf{v}_n)} \left| \frac{\partial(\mathbf{u}, \mathbf{v})}{\partial(\mathbf{y}, \mathbf{x})} \right|_{(\mathbf{x}_{n+1}, \mathbf{y}_{n+1})} \right)^{-1} \end{aligned}$$

$$\begin{aligned}
&= \left(\left| \frac{\partial(\mathbf{y}, \mathbf{x})}{\partial(\mathbf{u}, \mathbf{v})} \right|_{(\mathbf{u}_n, \mathbf{v}_n)} \left| \frac{\partial(\mathbf{y}, \mathbf{x})}{\partial(\mathbf{u}, \mathbf{v})} \right|_{(\mathbf{u}_n, \mathbf{v}_{n+1})}^{-1} \right)^{-1} \\
&= \left| \frac{\partial(\mathbf{y}, \mathbf{x})}{\partial(\mathbf{u}, \mathbf{v})} \right|_{(\mathbf{u}_n, \mathbf{v}_n)}^{-1} \left| \frac{\partial(\mathbf{y}, \mathbf{x})}{\partial(\mathbf{u}, \mathbf{v})} \right|_{(\mathbf{u}_n, \mathbf{v}_{n+1})}.
\end{aligned}$$

Substituting into the accept probability in the original variables with the deterministic move gives

$$\begin{aligned}
&A(\mathbf{y}_{n+1}, \mathbf{x}_{n+1} | \mathbf{y}_n, \mathbf{x}_n) \\
&= \min \left[1, \frac{\pi(\mathbf{y}_{n+1}, \mathbf{x}_{n+1} | \mathbf{d})}{\pi(\mathbf{y}_n, \mathbf{x}_n | \mathbf{d})} \left(\left| \frac{\partial(\mathbf{y}, \mathbf{x})}{\partial(\mathbf{u}, \mathbf{v})} \right|_{\mathbf{u}_n, \mathbf{v}_{n+1}} \left| \frac{\partial(\mathbf{y}, \mathbf{x})}{\partial(\mathbf{u}, \mathbf{v})} \right|_{\mathbf{u}_n, \mathbf{v}_n}^{-1} \right) \right. \\
&\quad \left. \times \frac{w(\mathbf{y}_n | \mathbf{y}_{n+1}, \mathbf{x}_{n+1}, \mathbf{d})}{w(\mathbf{y}_{n+1} | \mathbf{y}_n, \mathbf{x}_n, \mathbf{d})} \right] \\
&= \min \left[1, \frac{\pi(\mathbf{v}_{n+1}, \mathbf{u}_n | \mathbf{d})}{\pi(\mathbf{v}_n, \mathbf{u}_n | \mathbf{d})} \frac{w(\mathbf{y}_n | \mathbf{y}_{n+1}, \mathbf{x}_{n+1}, \mathbf{d})}{w(\mathbf{y}_{n+1} | \mathbf{y}_n, \mathbf{x}_n, \mathbf{d})} \right].
\end{aligned}$$

But under the change of variables $\mathbf{v} = \mathbf{y}$, and provided that the ratio of proposal densities is numerically the same

$$\frac{w(\mathbf{y}_n | \mathbf{y}_{n+1}, \mathbf{x}_{n+1}, \mathbf{d})}{w(\mathbf{y}_{n+1} | \mathbf{y}_n, \mathbf{x}_n, \mathbf{d})} = \frac{w(\mathbf{v}_n | \mathbf{u}_n, \mathbf{v}_{n+1}, \mathbf{d})}{w(\mathbf{v}_{n+1} | \mathbf{u}_n, \mathbf{v}_n, \mathbf{d})} \quad (\text{B13})$$

the accept probability using a deterministic change in \mathbf{x} is numerically equivalent to the accept probability for the move in the change of variables holding \mathbf{u} fixed! We therefore see a numerical equivalence in the accept probability between a conditional step in a change of variables and an associated move in the original variables, where \mathbf{x} is deterministically changed to hold $F(\mathbf{x}, \mathbf{y})$ constant. For the case studied in this paper, we have explicitly

$$F^{-1}(F(\mathbf{s}, C_l), C'_l) = [\mathbf{C}']^{1/2}(\mathbf{C}^{-1/2}\mathbf{s}) \quad (\text{B14})$$

which is exactly the functional form used for the deterministic MCMC steps.

REFERENCES

- Bennett, C. L., et al. 2003, [ApJS](#), **148**, 1
- Eriksen, H. K., Dickinson, C., Jewell, J. B., Banday, A. J., Górski, K. M., & Lawrence, C. R. 2008a, [ApJ](#), **672**, L87
- Eriksen, H. K., Huey, G., Banday, A. J., Górski, K. M., Jewell, J. B., O'Dwyer, I. J., & Wandelt, B. D. 2007a, [ApJ](#), **665**, L1
- Eriksen, H. K., Jewell, J. B., Dickinson, C., Banday, A. J., Górski, K. M., & Lawrence, C. R. 2008b, [ApJ](#), **676**, 10
- Eriksen, H. K., & Wehus, I. K. 2009, [ApJS](#), **180**, 30
- Eriksen, H. K., et al. 2004, [ApJS](#), **155**, 227
- Eriksen, H. K., et al. 2007b, [ApJ](#), **656**, 641
- Gantmakher, F. R. 1959, *The Theory of Matrices*, Vols. 1–2 (New York: Chelsea)
- Gelman, A., & Rubin, D. 1992, [Stat. Sci.](#), **7**, 457
- Gold, B., et al. 2009, [ApJS](#), **180**, 265
- Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., & Bartelmann, M. 2005, [ApJ](#), **622**, 759
- Green, P. 1995, [Biometrika](#), **82**, 711
- Gupta, A. K., & Nagar, D. K. 2000, *Matrix Variate Distributions*
- Hinshaw, G., et al. 2007, [ApJS](#), **170**, 288
- Hinshaw, G., et al. 2009, [ApJS](#), **180**, 225
- Jewell, J., Levin, S., & Anderson, C. H. 2002, arXiv:[astro-ph/0209560v1](#)
- Jewell, J., Levin, S., & Anderson, C. H. 2004, [ApJ](#), **609**, 1
- Komatsu, E., et al. 2009, [ApJS](#), **180**, 330
- Larson, D. L., Eriksen, H. K., Wandelt, B. D., Górski, K. M., Huey, G., Jewell, J. B., & O'Dwyer, I. J. 2007, [ApJ](#), **656**, 653
- O'Dwyer, I. J., et al. 2004, [ApJ](#), **617**, L99
- Page, L., et al. 2007, [ApJS](#), **170**, 335
- Roberts, G. O., & Rosenthal, J. S. 2001, [Stat. Sci.](#), **16**, 351
- Smoot, G. F., et al. 1992, [ApJ](#), **396**, L1
- Sokal, A. D. 1989, *Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms*, Lecture notes Cours de Troisième Cycle de la Physique en Suisse Romande, Lausanne
- Wandelt, B. D., Larson, D. L., & Lakshminarayanan, A. 2004, [Phys. Rev. D](#), **70**, 083511