COSMOLOGICAL PARAMETERS FROM CMB MAPS WITHOUT LIKELIHOOD APPROXIMATION

B. Racine[1], J. B. Jewell[2], H. K. Eriksen[1], and I. K. Wehus[1,2]

[1] Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, NO-0315 Oslo, Norway; benjamin.racine@astro.uio.no
[2] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

ABSTRACT

We propose an efficient Bayesian Markov chain Monte Carlo (MCMC) algorithm for estimating cosmological parameters from cosmic microwave background (CMB) data without the use of likelihood approximations. It builds on a previously developed Gibbs sampling framework that allows for exploration of the joint CMB sky signal and power spectrum posterior, $P(s, C_\ell | d)$, and addresses a long-standing problem of efficient parameter estimation simultaneously in regimes of high and low signal-to-noise ratio. To achieve this, our new algorithm introduces a joint Markov chain move in which both the signal map and power spectrum are synchronously modified, by rescaling the map according to the proposed power spectrum before evaluating the Metropolis–Hastings accept probability. Such a move was already introduced by Jewell et al., who used it to explore low signal-to-noise posteriors. However, they also found that the same algorithm is inefficient in the high signal-to-noise regime, since a brute-force rescaling operation does not account for phase information. This problem is mitigated in the new algorithm by subtracting the Wiener filter mean field from the proposed map prior to rescaling, leaving high signal-to-noise information invariant in the joint step, and effectively only rescaling the low signal-to-noise component. To explore the full posterior, the new joint move is then interleaved with a standard conditional Gibbs move for the sky map. We apply our new algorithm to simplified simulations for which we can evaluate the exact posterior to study both its accuracy and its performance, and find good agreement with the exact posterior; marginal means agree to $\lesssim 0.006\sigma$ and standard deviations to better than $\sim 3\%$. The Markov chain correlation length is of the same order of magnitude as those obtained by other standard samplers in the field.

Key words: cosmic background radiation – cosmological parameters – cosmology: observations – methods: numerical

1. INTRODUCTION

Observations of the cosmic microwave background (CMB) provide a direct image of the early universe (Bennett et al. 2013; Planck Collaboration 2015a), and have revolutionized our understanding of the composition and evolution of the universe as a whole (Hinshaw et al. 2013; Planck Collaboration 2015e). The new insights derive primarily from heroic community-wide efforts in microwave instrumentation, leading to steadily improved detector performance and noise levels. However, with improved data sets follow more stringent requirements on analysis techniques, and optimal statistical CMB analysis has become a rich scientific field in its own right during the last 20 years.

One branch of this community-wide effort has revolved around optimal exploration of the full joint CMB posterior, and among the most successful of such methods is the CMB Gibbs sampler. This framework, which was originally introduced by Jewell et al. (2004), has proved particularly powerful because of its ability to seamlessly and jointly account for astrophysical and instrumental nuisance parameters together with the primary CMB parameters, thereby both mitigating and propagating systematic uncertainties in final science results. This method has already been applied successfully to evaluation of power spectra for *COBE* (Wandelt et al. 2004), *WMAP* (Eriksen et al. 2007; Hinshaw et al. 2013), and *Planck* (Planck Collaboration 2015d), and it has played a central role in separation of astrophysical components for *Planck* (Planck Collaboration 2015b, 2015c).

By virtue of being a Gibbs sampler, this overall method consists of a series of iterated conditional Markov chain moves, in which one or more parameters are changed at a time, leaving all other parameters fixed. Each conditional parameter move is usually fairly simple, and always associated with a well-established sampling algorithm. For instance, the conditional distribution of the CMB power spectrum is an inverse Wishart distribution, that of the CMB sky map is a multivariate Gaussian (Jewell et al. 2004), while astrophysical foreground conditionals can usually be described in terms of some fairly simple χ^2 evaluations (Eriksen et al. 2008). Still, the computational cost per sample is certainly non-trivial, with the primary cost being driven by the multivariate Gaussian sky distribution, and can easily amount to several CPU hours per sample, requiring more advanced algorithmic treatment (Eriksen et al. 2004; Seljebotn et al. 2014; Jasche & Lavaux 2015).

In this paper, we revisit the problem of estimation of cosmological parameters and the CMB power spectrum within the Gibbs sampling framework, with the goal of establishing an efficient algorithm in both high and low signal-to-noise regimes, and thereby reducing the overall computational cost of the method. The same problem has been discussed and addressed repeatedly in the literature already (e.g., Eriksen et al. 2004; Jewell et al. 2009), but no definitive and general solution has been presented until now. In Section 2 we present the intuition behind our new algorithm, and we compare the new approach to existing sampling schemes. Then, in Section 3 we formalize the algorithm in standard mathematical notation, before testing it on simplified simulations in Section 4. We conclude in Section 5.

2. INTUITION AND MOTIVATION

As noted already by Eriksen et al. (2004), the most severe complication for estimation of the CMB power spectrum and
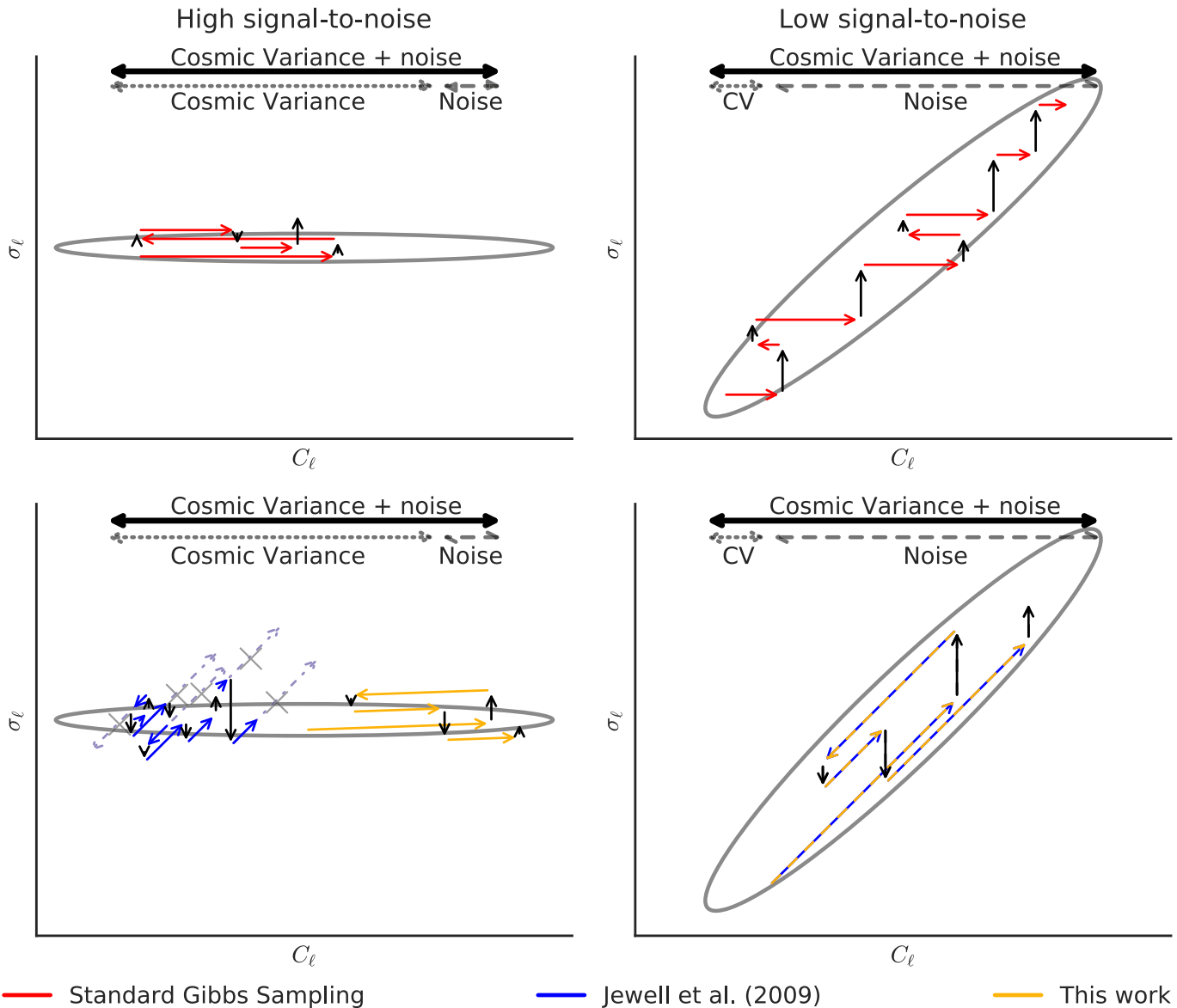
Figure 1. Illustration of the performance of different sampling algorithms in different signal-to-noise regimes. We sketch the exploration of the joint distribution of proposed C_ℓ and the signal map's power spectrum $\sigma_\ell = 1/(2\ell + 1)\sum|s_{\ell m}|^2$. In the high signal-to-noise limit (left column), standard Gibbs sampling steps (proposing C_ℓ according to the cosmic variance (CV) and solving Equation (10)) achieve both good acceptance rates and correlation lengths. In the low signal-to-noise limit (right column), the cosmic variance, and therefore the step length, is much smaller than the noise contribution to the posterior, and standard Gibbs sampling results in a long correlation length. Joint sampling steps in which the sky map is rescaled by the power spectrum, as proposed by Jewell et al. (2009) and illustrated in the bottom right panel, avoid this problem if one proposes C_ℓ's with a variance that includes noise in addition to cosmic variance. Unfortunately, as illustrated in the bottom left panel, the corresponding signal rescaling does not perform well in the high signal-to-noise regime. When proposing C_ℓ according to the cosmic variance and the noise, naive rescaling leads to a large change in the amplitude of the signal map, without correspondingly modifying the phase information of the map, and most steps are rejected in the Metropolis–Hastings acceptance evaluation through a poor effective χ^2. This problem is solved by the sampling algorithm introduced in the present paper, in which we exclude the high signal-to-noise Wiener filter component of the signal from the rescaling operation, and only modify the fluctuations around this mean-field map. The net result is an algorithm that works in both low and high signal-to-noise regimes.

cosmological parameters with the Gibbs sampling framework concerns the relationship between effective signal-to-noise ratio and Markov chain correlation length: while the width of the full power spectrum posterior is given by both cosmic variance and instrumental noise, the step size of the Markov chain power spectrum in the default algorithm (Jewell et al. 2004; Wandelt et al. 2004) is given by cosmic variance alone.

This problem is illustrated in the top two panels of Figure 1. Each move (illustrated by black arrows for sky map parameters and colored for power spectrum parameters) affects only one parameter at a time. Each arrow therefore points parallel to either coordinate axis. In the high signal-to-noise regime (top left panel), the sky signal is highly constrained, and the corresponding marginal posterior is very narrow. The power spectrum marginal, however, still has significant uncertainty due to cosmic variance, even if the noise contribution is small. However, since the sky map distribution essentially converges to a delta function with increasing signal-to-noise ratio, the joint distribution is nearly uncorrelated between the two directions, and pure Gibbs steps (defined by a Gaussian

distribution in the vertical direction and an inverse Wishart distribution in the horizontal direction) are able to navigate the full posterior efficiently.

In the low signal-to-noise regime (top right panel), this is no longer true. In this case, there is significant uncertainty in the true sky signal, and *for each sky map value* the power spectrum conditional follows an inverse Wishart distribution centered on a value given by the sky map. The joint distribution therefore becomes highly degenerate. To move from one end of the joint distribution to the other, a very large number of orthogonal moves are thus required. Such degeneracies are a well-known problem for the Gibbs sampling algorithm in general.

This problem was first identified and studied by Eriksen et al. (2004), and a first proper attempt at solving it was subsequently presented by Jewell et al. (2009). They introduced a new type of joint Markov chain move that consists of three steps. First, one proposes an arbitrary change to the power spectrum, $C_\ell^{i+1} \leftarrow T(C_\ell^i)$, where T is some proposal rule. Second, one rescales the corresponding sky map, $s^{i+1} \leftarrow S_{i+1}^{1/2} S_i^{-1/2} s^i$, where $S_i = S(C_\ell^i) = \langle ss^t \rangle$ is the signal-only covariance matrix. This rescaling operation leaves the quantity $s^t S^{-1} s$ invariant, and one can show that any determinant factors in the corresponding Metropolis–Hastings accept probability cancel, and the final accept ratio is given by χ^2's only (Jewell et al. 2009). The third and final step is therefore to evaluate this accept probability, and apply the Metropolis–Hastings rule.

Intuitively, this algorithm corresponds to diagonal moves in Figure 1, as illustrated in the bottom two panels. With an accept rate given by χ^2's alone, this kind of move works very well in the low signal-to-noise regime (bottom right panel), since the effective χ^2 does not change appreciably when changing the amplitude of the map. However, in the high signal-to-noise regime the χ^2 becomes sensitive to the *phase information* in the sky map, and large map rescaling factors are generally associated with very low accept probabilities; only very short moves are allowed in order to stay within the acceptable region.

In the present paper we solve this problem by introducing a small but critical variation of the previous scheme. First, as detailed by Jewell et al. (2004) and Wandelt et al. (2004), we note that the sky signal map may be decomposed into the sum of a Wiener filter component, $\hat{s}$, and a fluctuation term, $\hat{f}$, in the form $s = \hat{s} + \hat{f}$. The high signal-to-noise information is contained in $\hat{s}$, while the noise-dominated component is described by $\hat{f}$. Exploiting the intuition described above, we now note that the optimal Markov chain move should leave $\hat{s}$ invariant, and rescale only $\hat{f}$ by the power spectrum. Specifically, we introduce a partially rescaled proposal rule in the following form

$$s^{i+1} = \hat{s}^{i+1} + S_{i+1}^{1/2} S_i^{-1/2} (s^i - \hat{s}^i) \tag{1}$$

$$= \hat{s}^{i+1} + S_{i+1}^{1/2} S_i^{-1/2} \hat{f}^i \tag{2}$$

where $\hat{s}^i$ is the Wiener filtered sky map evaluated with C_ℓ^i.

This new move is contrasted to the previous joint sampler in the bottom two panels of Figure 1 in terms of yellow versus blue arrows. In the low signal-to-noise regime (right panel), the two perform nearly identically, since $s \approx \hat{f}$. However, they perform very differently in the high signal-to-noise regime: since $\hat{f}$ is very small in the signal-dominated regime, only small changes are proposed to s in the new scheme, maintaining a high net accept rate.

3. ALGORITHMS

In this section, we first define necessary notation and review previous CMB Gibbs sampling algorithms, before formalizing the intuition described above into a well-defined and operational algorithm. The technical derivation of the Metropolis–Hastings accept probability is deferred to the Appendix.

Let us start by considering a data model of the form

$$d = As + n, \tag{3}$$

where d denotes a data vector in pixel space; A represents convolution with an instrumental beam, usually represented by a Legendre transform b_ℓ; s is a Gaussian signal with covariance S; and n denotes Gaussian instrumental noise with covariance N. We further assume that the signal s is statistically isotropic, and define its power spectrum as $C_\ell \equiv \langle s_{\ell m}^* s_{\ell' m'} \rangle = C_\ell \delta_{\ell \ell'} \delta_{mm'}$, where $s = \sum_{\ell m} s_{\ell m} Y_{\ell m}$. The signal covariance matrix is then given as $S_{\ell m, \ell' m'} = C_\ell \delta_{\ell \ell'} \delta_{mm'}$.

The overall goal of this paper is to characterize the marginal power spectrum posterior $P(C_\ell | d)$ somehow. In the literature, this is conventionally done in several different ways, for instance by adopting either single-ℓ or binned estimates of the power spectrum. While the algorithm presented here certainly is suitable for such parameterizations as well, we will in the following instead focus directly on cosmological parameters, which are the ultimate goal of any CMB experiment. Furthermore, adopting a high-level parameterization that depends on both low and high multipoles (and therefore both high and low signal-to-noise regimes) puts maximum pressure on the algorithm itself. Note that this method is also naturally suitable for sampling power spectrum coefficients, C_ℓ, which can be useful to reveal features or anomalies in the data.

For convenience, we adopt a standard six-parameter ΛCDM model in the following, with baryon density $\Omega_b h^2$, cold dark matter density $\Omega_c h^2$, optical depth at reionization τ, amplitude and tilt of the primordial fluctuations A_s and n_s, and the Hubble parameter H_0 as free parameters. To evaluate corresponding power spectra, we employ CAMB[3] (Lewis et al. 2000). In the rest of the paper, we will denote this vector of parameters as θ.

Thus, our goal is to map out $P(\theta | d)$. To do so, we use Bayes' theorem to write the joint density $P(\theta, s | d)$ as

$$P(\theta, s | d) = \frac{P(\theta, s, d)}{P(d)} = P(d|s) P(s|\theta) \frac{P(\theta)}{P(d)}$$

$$= \frac{e^{-\frac{1}{2}(d-As)^t N^{-1}(d-As)}}{\sqrt{|N|}} \frac{e^{-\frac{1}{2} s^t S^{-1} s}}{\sqrt{|S|}} \frac{P(\theta)}{P(d)}. \tag{4}$$

For clarity, we have dropped the dependence of S on θ, as well as any factor of 2π. Optional priors on θ are described by $P(\theta)$, and from now on we will neglect the overall normalization factor, $P(d)$, often called the evidence. We note that our target distribution is obtained by marginalizing the joint distribution over s,

$$P(\theta | d) = \int P(\theta, s | d) ds. \tag{5}$$

3.1. Gibbs Sampling

To recap, the basic idea behind Gibbs sampling is to draw a sample from the joint posterior $P(\theta, s|d)$ by iteratively sampling from the corresponding conditional probabilities,

$$s^{i+1} \leftarrow P(s|\theta^i, d) \qquad (6)$$

$$\theta^{i+1} \leftarrow P(\theta|s^{i+1}, d). \qquad (7)$$

As discussed by Jewell et al. (2004) and Wandelt et al. (2004), the former of these distributions may be recognized as a standard multivariate Gaussian by rewriting the exponent in Equation (4) as follows:

$$(d - As)^t N^{-1}(d - As) + s^t S^{-1} s$$
$$= (s - \hat{s})^t (S^{-1} + A^t N^{-1} A)(s - \hat{s}), \qquad (8)$$

where we defined the mean-field map $\hat{s} \equiv (S^{-1} + A^t N^{-1} A)^{-1} A N^{-1} d$. Given this expression, we sample the sky signal according to

$$P(s|\theta^i, d) = \mathcal{N}(\hat{s}, (S^{-1} + A^t N^{-1} A)^{-1}), \qquad (9)$$

where $\mathcal{N}(\mu, \mathcal{C})$ is a Gaussian multivariate, with a mean vector μ and a covariance $\mathcal{C}$. Specifically, we generate two random vectors, ω_0 and ω_1, drawn from $\mathcal{N}(0, 1)$, and solve the equation

$$[S^{-1} + A^t N^{-1} A]s = A N^{-1} d + S^{-\frac{1}{2}}\omega_0 + A N^{-\frac{1}{2}}\omega_1. \qquad (10)$$

Solving Equation (10) in the context of a realistic experiment with anisotropic noise and non-trivial masks can be computationally expensive (Seljebotn et al. 2014; Jasche & Lavaux 2015). However, this is a purely computational problem of algebraic nature, and fully independent of questions regarding Monte Carlo correlation lengths and signal-to-noise levels. In this paper, we therefore circumvent this problem entirely, and consider only an ideal data set in the harmonic domain with uniform noise and full sky coverage. In this particular case, the noise matrix may be described by a noise power spectrum, $N_{\ell m, \ell' m'} = N_\ell \delta_{\ell\ell'} \delta_{mm'}$, and Equation (10) may be solved directly in harmonic space at negligible computational cost:

$$\hat{s}_{\ell m} = d_{\ell m} \frac{b_\ell}{N_\ell} \sqrt{C_\ell} \frac{N_\ell \sqrt{C_\ell}}{N_\ell + b_\ell^2 C_\ell} \qquad (11)$$

$$\hat{f}_{\ell m} = \left(w_{0\ell m} + w_{1\ell m} b_\ell \frac{\sqrt{C_\ell}}{\sqrt{N_\ell}} \right) \frac{N_\ell \sqrt{C_\ell}}{N_\ell + b_\ell^2 C_\ell}. \qquad (12)$$

Finally, the second conditional distribution in Equation (7) reduces to an inverse Wishart distribution with a well-known sampling algorithm. This particular sampling step, however, will not be needed for the purposes of the current paper, and we therefore refer the interested reader to the referenced papers for further details.

3.2. Joint Sampling by Metropolis–Hastings MCMC

Adopting the above notation, Jewell et al. (2009) introduced the following joint sampling step to mitigate the slow convergence rate discussed in Section 2 in the low signal-to-noise regime:

$$C_\ell^{i+1} = C_\ell^i + \delta C_\ell \qquad (13)$$

$$s_{\ell m}^{i+1} = \sqrt{\frac{C_\ell^{i+1}}{C_\ell^i}} s_{\ell m}^i, \qquad (14)$$

where δC_ℓ denotes a random fluctuation drawn from a normal distribution with zero mean and variance given by the sum of cosmic variance and instrumental noise. They then showed that the corresponding Metropolis–Hastings accept probability for such a joint move reduced to the ratio of exponentiated χ^2's,

$$A = \min\left(1, \frac{e^{-\frac{1}{2}(d - As^{i+1})^t N^{-1}(d - As^{i+1})}}{e^{-\frac{1}{2}(d - As^i)^t N^{-1}(d - As^i)}}\right). \qquad (15)$$

Although efficient in the low signal-to-noise regime, this particular move performs very poorly in the high signal-to-noise regime.

To solve this, we propose in this paper a slight—but critical—variation of the above scheme, as discussed in Section 2. Specifically, rather than rescaling the full signal map, we propose to rescale only the fluctuation component, such that

$$\hat{f}_{\ell m}^{i+1} = \sqrt{\frac{C_\ell^{i+1}}{C_\ell^i}} \hat{f}_{\ell m}^i, \qquad (16)$$

or written in terms of s^{i+1} as in Equation (2),

$$s^{i+1} = \hat{s}^{i+1} + S_{i+1}^{1/2} S_i^{-1/2}(s^i - \hat{s}^i). \qquad (17)$$

Again, the underlying intuition behind this proposal is to leave the high signal-to-noise component of the sky signal invariant, and modify only the low signal-to-noise component, to which the total χ^2 is largely insensitive.

We derive the Metropolis–Hastings accept rate for this new proposal in the Appendix, and find it to be given by

$$A = \min\left[1, \frac{\pi(\theta^{i+1})}{\pi(\theta^i)} \frac{w(\theta^i|\theta^{i+1})}{w(\theta^{i+1}|\theta^i)} \frac{P(\theta^{i+1})}{P(\theta^i)}\right], \qquad (18)$$

where

$$\pi(\theta^{i+1}) = e^{-\frac{1}{2}(d - A\hat{s})^t N^{-1}(d - A\hat{s})} e^{-\frac{1}{2}\hat{s}^t S^{-1}\hat{s}} e^{-\frac{1}{2}\hat{f}^t A^t N^{-1} A\hat{f}}, \qquad (19)$$

and $w(\theta^{i+1}|\theta^i)$ denotes the proposal distribution for θ in Equation (13). Note that, for clarity, we dropped the θ^{i+1} dependence of $S(\theta^{i+1})$, $\hat{s}(\theta^{i+1})$, and $\hat{f}(\theta^{i+1})$ in the above expression.

The proposal rule w is in principle arbitrary. However, for standard cosmological parameter estimation, as implemented for instance in CosmoMC (Lewis & Bridle 2002), overall faster convergence is achieved when adopting a proposal rule that is close to the underlying marginal posterior distribution. Following CosmoMC and other samplers, we therefore adopt a multivariate Gaussian for w of the form:

$$w(\theta^{i+1}|\theta^i) = e^{-\frac{1}{2}(\theta^{i+1} - \theta^i)^t C_\theta^{-1}(\theta^{i+1} - \theta^i)}, \qquad (20)$$

with a covariance matrix derived by some earlier analysis. (If no such earlier analysis is available, we generate a short chain with a diagonal covariance matrix, and compute a first covariance matrix from that run.)

As in the case of the original method introduced by Jewell et al. (2009), the sampling step introduced above explores by itself only a very limited subspace of the full volume of the sky signal posterior, namely that spanned by a single amplitude rescaling. To explore the full posterior volume, this step must

therefore be interleaved with a standard Gibbs step, as described in Equation (8). Overall, the full sampler therefore works as follows:

1. Propose some initial parameter vector θ^0, and generate a power spectrum C_ℓ^0 with CAMB. Solve Equations (11) and (12) to obtain the mean-field map $\hat{s}(\theta^0)$ and the fluctuation map $\hat{f}(\theta^0)$.
2. Propose a new parameter vector θ^1 according to the proposal rule w, and compute the corresponding power spectrum C_ℓ^1.
3. Compute the deterministically rescaled fluctuation map $\hat{f}_{\ell m}(\theta^1) = \sqrt{C_\ell^1/C_\ell^0}\,\hat{f}_{\ell m}(\theta^0)$, and evaluate the accept probability according to Equation (18); accept or reject the proposal according to the usual Metropolis–Hastings rule.
4. Given the most recent parameter sample, make a standard conditional Gibbs step for the sky map, according to Equation (8), computing both the mean-field and fluctuation maps.
5. Iterate 2–4.

3.3. Brute-force Direct Sampling

To validate and benchmark our new sampling scheme, we compare it to a case for which we can evaluate the exact posterior at negligible cost, namely a data set with uniform noise and full sky coverage. In this case, the exact marginal parameter posterior, $P(\theta|\mathbf{d})$, reads

$$\Pi(\theta|\mathbf{d}) = \frac{e^{-\frac{1}{2}\mathbf{d}^t(\mathbf{A}^t\mathbf{S}(\theta)\mathbf{A}+\mathbf{N})^{-1}\mathbf{d}}}{\sqrt{|(\mathbf{A}^t\mathbf{S}(\theta)\mathbf{A})+\mathbf{N}|}}. \tag{21}$$

To map out this distribution, we use a standard Metropolis sampler with the same proposal distribution, w, as for the joint Gibbs sampler.

4. VALIDATION AND BENCHMARKS

To validate and benchmark our method, we now apply it to a simplified simulation generated as follows. We draw a random Gaussian CMB sky realization from the *Planck* 2015 best-fit ΛCDM power spectrum (TT + lowP; Planck Collaboration 2015e), and convolve this with a 13′ FWHM Gaussian beam. Finally, we add white noise with a power spectrum amplitude of $N_\ell = 1.84 \times 10^{-3}\,\mu\mathrm{K}^2$. Both the beam and noise level are chosen to mimic *WMAP*, in order to probe both the high and low signal-to-noise regimes within our effective multipole range. To be specific, with these parameter choices we find a signal-to-noise ratio per multipole of unity at $\ell = 900$.

Next, since we consider only temperature observations in the following, our constraints on τ are very loose. To produce more realistic results, we therefore additionally impose an informative prior of $\tau = 0.07 \pm 0.02$.

As our first proposal matrix, we adopt the covariance matrix obtained from the *Planck* 2015 TT + lowP Markov chains, which are publicly available from the *Planck* Legacy Archive.[4] However, due to its higher signal-to-noise ratio and different effective sky realization, this proposal distribution is quite poor. We therefore first generate a set of chains with 50k samples each, and compute the parameter covariance from the

---
[4] http://www.cosmos.esa.int/web/planck/pla
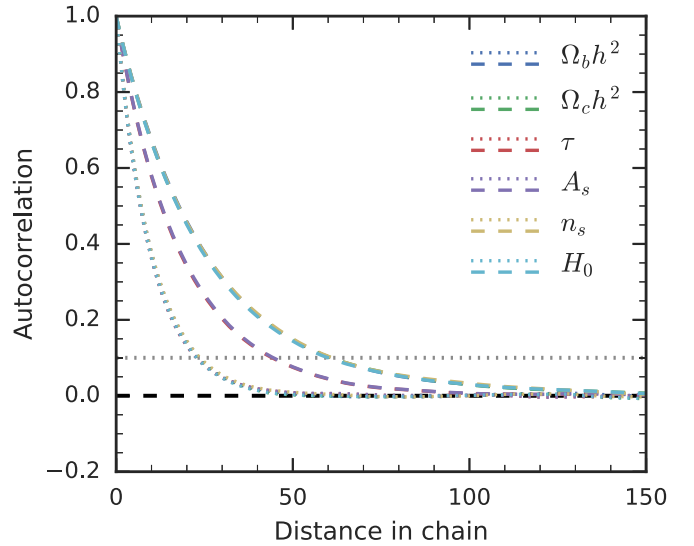
Figure 2. Autocorrelation function for each cosmological parameter. The correlation length, defined as the distance in the chain above which the autocorrelation function drops below 10%, is around 50 for all chains using the joint method (dashed), whereas with the direct method (dotted) chains have a factor-of-two shorter correlation length.

latter 40k samples. We additionally rescale the resulting distribution by $C_\theta \rightarrow 2.4/\sqrt{6}\,C_\theta$, as proposed in Dunkley et al. (2005), for further optimization. Based on this proposal, we run 40 chains with 30k samples each, for both the new joint sampler and the exact sampler. After conservatively removing burn-in, we retain a total of 1M samples for analysis. Note that this is far more than is strictly required, but since each sample is cheap, and our primary concerns here are of algorithmic nature, optimization of chain lengths is not an issue.

We first consider the overall efficiency of the algorithm, as measured in terms of Markov chain correlation lengths. These are shown in Figure 2 for each of the six cosmological parameters, and for both the new (dashed) and the exact (dotted) samplers. Overall, we see that the correlation lengths for the new sampler are roughly twice as long as for the standard sampler, which translates into a computational cost roughly double that of standard samplers. For comparison, we typically find a correlation length of roughly 30–50 with CosmoMC. Thus, the new sampler performs similarly to existing samplers in terms of overall correlation length, within a very small factor. This has never before been achieved within the CMB Gibbs sampling framework.

Next, in Figure 3 we compare the 1D and 2D marginal distributions derived using the two samplers. At least at a visual level, all distributions agree very well. This agreement is quantified in Table 1 in terms of posterior mean and standard deviations for each of the two methods, and the relative difference between the two. We find that the posterior means are identical up to a few thousandths of a σ, while somewhat larger differences are observed for the posterior standard deviation—up to 3% for τ and A_s. The cause of this small discrepancy is still under investigation, although we observe that it vanishes if we loosen or remove the prior on τ. It is thus not an intrinsic feature of the method as such, but rather related to the use of external priors. Since most applications of this type are anyway made without such external priors, and the difference is in either case very small, we defer further discussion and resolution of the issue to a future publication.
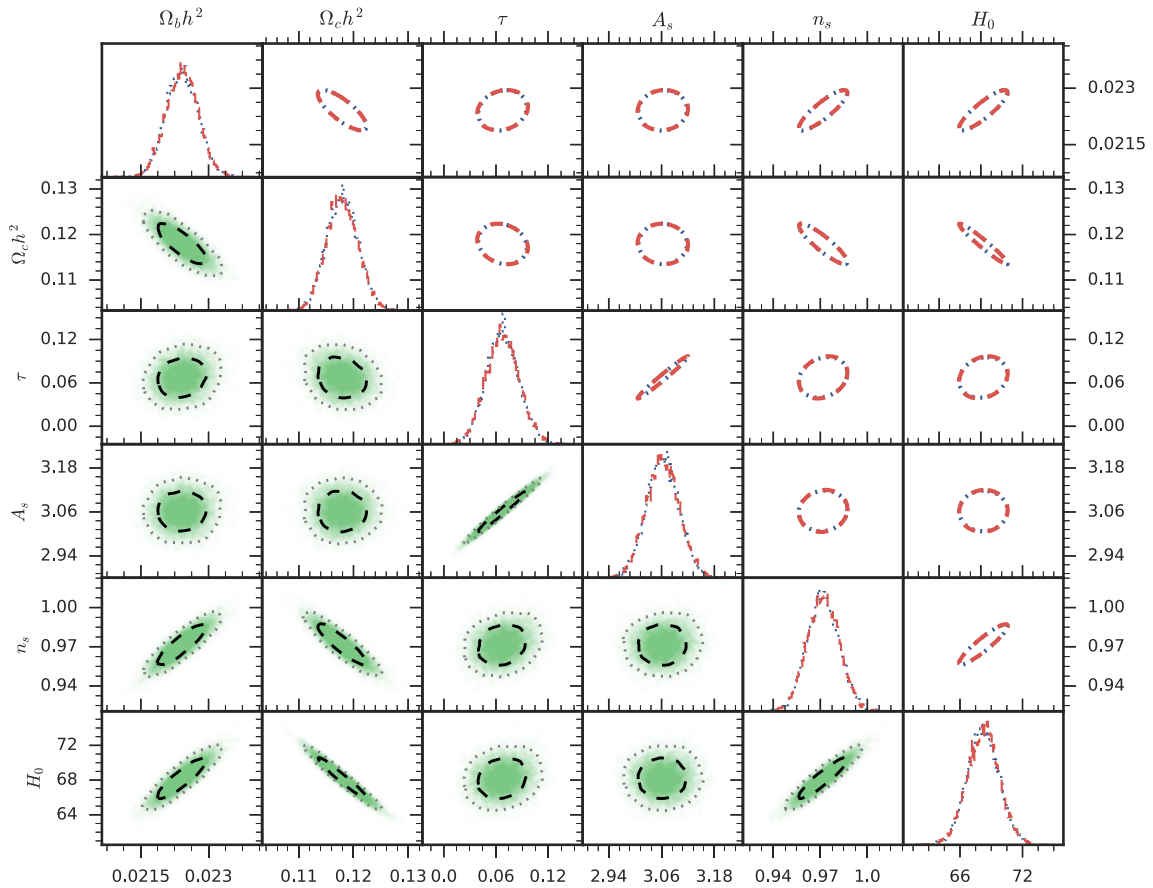
Figure 3. Comparison of the recovered posterior distribution for the two sampling methods described in the main text. On the diagonal we show the marginal posterior histograms for the direct method (red dashed) and the joint method (blue dotted). In the upper triangle, we show the 1σ 2D-Gaussian ellipses, whose mean value and covariance are estimated from the chains. In the lower triangle, we show scatter plots for the joint method (green), as well as the 1σ (black dashed) and 2σ (gray dotted) contours, computed from a Gaussian kernel density estimation on the chains.

Table 1
Summary of Cosmological Parameters

	$\Omega_b h^2$	$\Omega_c h^2$	τ	A_s	n_s	H_0
New joint sampler	0.02241 ± 0.00035	0.1180 ± 0.0029	0.067 ± 0.019	3.063 ± 0.037	0.972 ± 0.01	68.2 ± 1.5
Exact posterior	0.02241 ± 0.00035	0.1180 ± 0.0029	0.068 ± 0.019	3.063 ± 0.038	0.972 ± 0.01	68.2 ± 1.5
Posterior mean bias $\left(\frac{\Delta\langle\mu\rangle}{\sigma_1}\right)$	-0.002	-0.002	-0.005	-0.006	-0.001	0.003
Posterior rms bias (%)	0.0	0.0	3.2	3.1	0.0	0.1

Note. Posterior mean and standard deviations for each of the six cosmological parameters are shown for the new sampling algorithm in the top row and for the exact calculation in the second row. The third row shows the difference in the posterior means measured in units of σ_1, the rms derived from the new method. The bottom row shows relative difference in the derived posterior RMSs as a percentage. Note that an external prior was applied to τ.

5. CONCLUSIONS

We have introduced a new sampling step for exploring the joint CMB sky signal and power spectrum posterior that is suitable for both high and low signal-to-noise regimes. This new step is very closely related to an already introduced rescaling algorithm (Jewell et al. 2009), but with a very important difference: rather than rescaling the entire sky signal in each iteration, we now propose to rescale only the component with low signal-to-noise fluctuation. As a result, high Metropolis–Hastings accept rates are maintained even in the high signal-to-noise regime.

Our focus in this paper has been on purely statistical–algorithmic aspects of the method, not real-world applications. To make the algorithm useful for such, it needs to be combined with state-of-the-art constrained realization samplers (Eriksen et al.

2004; Seljebotn et al. 2014), in order to solve the computationally expensive map-making step efficiently. Once that task has been completed, it will finally be possible to go from raw sky maps to cosmological parameters without the use of any likelihood approximations whatsoever. In addition, full physical marginalization over astrophysical foreground contamination will be straightforward (Eriksen et al. 2008; Planck Collaboration 2015c).

Finally, we note that while we have considered only CMB temperature analysis in the current paper, the method generalizes naturally to all other fields that employ joint Gibbs sampling as their basic algorithm. Three specific examples include CMB polarization analysis (Larson et al. 2007), large-scale structure analysis (Jasche & Wandelt 2013), and weak lensing analysis (Alsing et al. 2016).

APPENDIX
TECHNICAL DETAILS OF THE PARAMETER SAMPLING MCMC STEP

We review the technical details of constructing an MCMC algorithm to sample from the Bayes joint posterior of cosmological parameters θ and CMB signal maps s given data $p(\theta, s|d)$, ideally striking a balance between *overall computational expense* and *number of posterior samples*. Transition matrices for MCMC algorithms are constructed to be both *stationary* and *irreducible*. The latter simply means that any "state" is reachable with a finite (non-vanishing) probability after enough iterations, while the former means that the target distribution is invariant under repeated iterations of the algorithm. These two properties are sufficient to establish convergence in measure to the target distribution when started from some initial probability density.

In this Appendix we concentrate on just the step involving variations in cosmological parameters and deterministic changes in the CMB map—this step allows large jumps in the parameters, and represents a *stationary* step leaving the Bayes posterior invariant. It is not, however, irreducible—but we interleave standard Gibbs steps in between to produce an overall transition matrix given by the product of the two that is both stationary and irreducible. We now discuss the details of the joint parameter and CMB map step, and derive its accept probability.

A.1. Joint Proposal for Cosmological Parameters and CMB Signal Map

We first assume joint proposals for *both* CMB maps and parameters of the form

$$w(\theta^{i+1}, s^{i+1}|\theta^i, s^i, d)$$
$$= w(s^{i+1}|\theta^{i+1}, \theta^i, s^i, d)w(\theta^{i+1}|\theta^i, s^i, d). \quad (22)$$

In words, we first generate a proposal for cosmological parameters possibly dependent on the current state of the MCMC chain, followed by a proposal for a new CMB map given the current and proposed parameters θ^i and θ^{i+1} respectively, the current CMB map, and the data (in the numerical examples presented in this paper, a *symmetric* proposal for the parameters was used for them, but we proceed with the general case for now).

For the proposal for the new map s^{i+1}, we consider variations about the new *mean-field map*

$$\hat{s}^{i+1} = (A^t N^{-1} A + S^{-1}(\theta^{i+1}))^{-1} A N^{-1} d \quad (23)$$

(note that the mean-field map is a function of information conditioned on when proposing s^{i+1}, specifically θ^{i+1} and the data d). We consider, for some linear filter $F(\theta^{i+1}, \theta^i)$, general

proposals of the form

$$s^{i+1} = \hat{s}^{i+1} + F(\theta^{i+1}, \theta^i)(s^i - \hat{s}^i) + \beta^{1/2}\xi, \quad (24)$$

with ξ Gaussian-distributed with unit variance and β a scaling factor controlling its variance. Our family of proposals for $w(s^{i+1}|\theta^{i+1}, \theta^i, s^i, d)$ is therefore

$$w_\beta(s^{i+1}|\theta^{i+1}, \theta^i, s^i, d)$$
$$= \frac{1}{(2\pi\beta)^{(\ell_{max}+1/2)}} e^{-\frac{\|s^{i+1}-\hat{s}^{i+1}-F(\theta^{i+1},\theta^i)(s^i-\hat{s}^i)\|^2}{2\beta}}. \quad (25)$$

We take the limit $\beta \to 0$, which reduces to a δ-function for the map about the *deterministic* proposal

$$w_\beta(s^{i+1}|\theta^{i+1}, \theta^i, s^i, d) \xrightarrow{\beta \to 0}$$
$$\times \delta[s^{i+1} - \hat{s}^{i+1} - F[\theta^{i+1}, \theta^i](s^i - \hat{s}^i)]. \quad (26)$$

Finally, we focus on one choice of filter (used for the numerical results in this paper):

$$F(\theta^{i+1}, \theta^i) = S^{1/2}(\theta^{i+1})S^{-1/2}(\theta^i) \quad (27)$$

(see the discussion below for a generalization of this choice of filter and its impact on the MCMC algorithm). To satisfy detailed balance, note that we must have

$$s^i - \hat{s}^i = F[\theta^i, \theta^{i+1}](s^{i+1} - \hat{s}^{i+1}), \quad (28)$$

which means that the filter must satisfy

$$F(\theta^i, \theta^{i+1}) = F^{-1}(\theta^{i+1}, \theta^i) \quad (29)$$

(which it does, as can be readily verified).

A.2. Functional Form of the Accept Probability

As reviewed in Jewell et al. (2009), stationary MCMC transition matrices can be constructed by demanding *detailed balance*, from which the probabilistic rule of accepting or rejecting proposed changes in (θ, s) can be derived. We will require that the probability of *rejecting a proposed move when starting from state* (θ^2, s^2) is equal to the probability of *accepting a transition TO the state* (θ^2, s^2) *from some other state*:

Prob. of rej.
$$= \left[\int d(s^1, \theta^1) A[\theta^1, s^1|\theta^2, s^2] \delta[(s^1 - \hat{s}^1) \right.$$
$$\left. - F(\theta^1, \theta^2)(s^2 - \hat{s}^2)] w(\theta^1|\theta^2, s^2, d)]P(\theta^2, s^2|d) \right.$$
Prob. of acc.
$$= \left[\int d(s^1, \theta^1) A[\theta^2, s^2|\theta^1, s^1] \delta[(s^2 - \hat{s}^2) \right.$$
$$\left. - F(\theta^2, \theta^1)(s^1 - \hat{s}^1)] w(\theta^2|\theta^1, s^1, d)]P(\theta^1, s^1|d). \quad (30)$$

The integration over the δ-function in the latter term is equivalent to

Prob. of acc.
$$= \left[\int d(s^1, \theta^1) A[\theta^2, s^2|\theta^1, s^1] \right.$$
$$\times \frac{\delta[(s^1 - \hat{s}^1) - F(\theta^1, \theta^2)(s^2 - \hat{s}^2)]}{|F(\theta^2, \theta^1)|}$$
$$\left. \times w(\theta^2|\theta^1, s^1, d)]P(\theta^1, s^1|d) \quad (31) \right.$$

where we explicitly note that

$$|F(\theta^2, \theta^1)|^{-1} = \frac{|S(\theta^1)|^{1/2}}{|S(\theta^2)|^{1/2}} \qquad (32)$$

and where the last line follows from $F^{-1}(\theta^2, \theta^1) = F(\theta^1, \theta^2)$. In order to have the two integrals for the reject and accept probabilities be equal, we can demand that in detail the accept probability satisfies

$$P(\theta^2, s^2|d)w(\theta^1|\theta^2, s^2, d) A[\theta^1, s^1|\theta^2, s^2]$$
$$= A[\theta^2, s^2|\theta^1, s^1] \left(\frac{1}{|F(\theta^2, \theta^1)|}\right) w$$
$$\times (\theta^2|\theta^1, s^1, d) P(\theta^1, s^1|d), \qquad (33)$$

which equivalently leads to the usual rule $A(\theta^2, s^2|\theta^1, s^1) = \min[1, R(\theta^2, s^2|\theta^1, s^1)]$ with the ratio defined as

$$R(\theta^2, s^2|\theta^1, s^1) = \left(\frac{P(\theta^2, s^2|d)}{P(\theta^1, s^1|d)}\right)\left(\frac{|S(\theta^2)|^{1/2}}{|S(\theta^1)|^{1/2}}\right)$$
$$\times \left(\frac{w(\theta^1|\theta^2, s^2, d)}{w(\theta^2|\theta^1, s^1, d)}\right). \qquad (34)$$

Explicitly substituting the form of the joint posterior $P(\theta, s|d)$, and using the fact that $\hat{f}^1 S^{-1}(\theta^1)\hat{f}^1 = \hat{f}^2 S^{-1}(\theta^2)\hat{f}^2$, we have

$$R(\theta^2, s^2|\theta^1, s^1) = \left(\frac{P(\theta^2)}{P(\theta^1)}\right)\left(\frac{e^{-\frac{1}{2}\chi^2[\hat{s}^2] - \frac{1}{2}\hat{s}^2 S^{-1}(\theta^2)\hat{s}^2}}{e^{-\frac{1}{2}\chi^2[\hat{s}^1] - \frac{1}{2}\hat{s}^1 S^{-1}(\theta^1)\hat{s}^1}}\right)$$
$$\times \left(\frac{e^{-\frac{1}{2}\hat{f}^2 A^t N^{-1} A\hat{f}^2}}{e^{-\frac{1}{2}\hat{f}^1 A^t N^{-1} A\hat{f}^1}}\right)\left(\frac{w(\theta^1|\theta^2, s^2, d)}{w(\theta^2|\theta^1, s^1, d)}\right), \quad (35)$$

where $\chi^2[\hat{s}^2] = (d - A\hat{s})^t N^{-1}(d - A\hat{s})$. Above and from now on, we drop the $(.)^t$ notation for the transpose vectors. For the choice of a parameter proposal matrix that is *independent of the CMB map and data and symmetric* (for example $-\log w$ $(\theta^{i+1}|\theta^i, s^i, d) = -\log w(\theta^{i+1}|\theta^i) \sim (\theta^{i+1} - \theta^i) C_\theta^{-1}(\theta^{i+1} - \theta^i))$ the ratio of parameter proposals drops out and we are left with

$$R(\theta^2, s^2|\theta^1, s^1) = \left(\frac{P(\theta^2)}{P(\theta^1)}\right)\left(\frac{e^{-\frac{1}{2}\chi^2[\hat{s}^2] - \frac{1}{2}\hat{s}^2 S^{-1}(\theta^2)\hat{s}^2}}{e^{-\frac{1}{2}\chi^2[\hat{s}^1] - \frac{1}{2}\hat{s}^1 S^{-1}(\theta^1)\hat{s}^1}}\right)$$
$$\times \left(\frac{e^{-\frac{1}{2}\hat{f}^2 A^t N^{-1} A\hat{f}^2}}{e^{-\frac{1}{2}\hat{f}^1 A^t N^{-1} A\hat{f}^1}}\right). \qquad (36)$$

A.3. Generalization to Unity Accept Transitions versus Computational Expense

We note here the interesting generalization of the scheme outlined above which uses deterministic proposals for the CMB signal maps and which leads to unity accept proposals.

We look for a filtering operation $F(\theta^{i+1}, \theta^i)$ that leaves invariant

$$\hat{f}^{i+1}(N^{-1} + S^{-1}(\theta^{i+1}))\hat{f}^{i+1} = \hat{f}^i(N^{-1} + S^{-1}(\theta^i))\hat{f}^i. \quad (37)$$

We can therefore set

$$F(\theta^{i+1}, \theta^i) = (N^{-1} + S^{-1}(\theta^{i+1}))^{-1/2}(N^{-1} + S^{-1}(\theta^i))^{+1/2}. \qquad (38)$$

This has the desired invariance of the fluctuation map, as well as satisfying the condition required for detailed balance, $F^{-1}(\theta^{i+1}, \theta^i) = F(\theta^i, \theta^{i+1})$. The determinant factor appearing in the ratio for the accept probability is

$$|F(\theta^{i+1}, \theta^i)| = \frac{|N^{-1} + S^{-1}(\theta^{i+1})|^{-1/2}}{|N^{-1} + S^{-1}(\theta^i)|^{-1/2}}$$
$$= \left(\frac{|S(\theta^{i+1})|^{1/2}}{|S(\theta^i)|^{1/2}}\right)\left(\frac{|N + S(\theta^i)|^{1/2}}{|N + S(\theta^{i+1})|^{1/2}}\right) \qquad (39)$$

(with generalizations for the case when N^{-1} is singular). The significance of the above is the following—everything in the accept probability ratio R cancels *if* the proposals for the cosmological parameters are from the "exact" form as discussed in this paper.

While it may appear somewhat paradoxical to consider *proposals* in cosmological parameters from the exact posterior marginal (therefore making MCMC unnecessary), the idea provides insight into the correctness of the algorithm as well as intuition if *approximations to the inverse noise* can be made. A well-known approximation to the functional form of the likelihood in CMB analysis has been to take only the diagonal elements in a spherical harmonic basis. The above comments simply show that in the limit that approximations to the noise converge to the true instrument noise (and we can generate proposals to parameters from the approximate posterior) we will have high accept probabilities.

REFERENCES

Alsing, J., Heavens, A., Jaffe, A. H., et al. 2016, MNRAS, 455, 4452
Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, ApJS, 208, 20
Dunkley, J., Bucher, M., Ferreira, P. G., Moodley, K., & Skordis, C. 2005, MNRAS, 356, 925
Eriksen, H. K., Huey, G., Banday, A. J., et al. 2007, ApJL, 665, L1
Eriksen, H. K., Jewell, J. B., Dickinson, C., et al. 2008, ApJ, 676, 10
Eriksen, H. K., O'Dwyer, I. J., Jewell, J. B., et al. 2004, ApJS, 155, 227
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19
Jasche, J., & Lavaux, G. 2015, MNRAS, 447, 1204
Jasche, J., & Wandelt, B. D. 2013, ApJ, 779, 15
Jewell, J., Levin, S., & Anderson, C. H. 2004, ApJ, 609, 1
Jewell, J. B., Eriksen, H. K., Wandelt, B. D., et al. 2009, ApJ, 697, 258
Larson, D. L., Eriksen, H. K., Wandelt, B. D., et al. 2007, ApJ, 656, 653
Lewis, A., & Bridle, S. 2002, PhRvD, 66, 103511
Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473
Planck Collaboration 2015a, arXiv:1502.01582
Planck Collaboration 2015b, arXiv:1502.05956
Planck Collaboration 2015c, arXiv:1502.01588
Planck Collaboration 2015d, arXiv:1507.02704
Planck Collaboration 2015e, arXiv:1502.01589
Seljebotn, D. S., Mardal, K.-A., Jewell, J. B., Eriksen, H. K., & Bull, P. 2014, ApJS, 210, 24
Wandelt, B. D., Larson, D. L., & Lakshminarayanan, A. 2004, PhRvD, 70, 083511