BAYESIAN INFERENCE OF CMB GRAVITATIONAL LENSING

ETHAN ANDERES\textsuperscript{1}, BENJAMIN D. WANDELT\textsuperscript{2}, AND GUILHEM LAVAUX\textsuperscript{2}

\textsuperscript{1}Department of Statistics, University of California, Davis, CA 95616, USA
\textsuperscript{2}Sorbonne Universités, UPMC Univ Paris 06 & CNRS, UMR7095, Institut d’Astrophysique de Paris, F-75014, Paris, France

Received 2014 December 12; accepted 2015 June 7; published 2015 July 29

ABSTRACT

The Planck satellite, along with several ground-based telescopes, has mapped the cosmic microwave background (CMB) at sufficient resolution and signal-to-noise so as to allow a detection of the subtle distortions due to the gravitational influence of the intervening matter distribution. A natural modeling approach is to write a Bayesian hierarchical model for the lensed CMB in terms of the unlensed CMB and the lensing potential. So far there has been no feasible algorithm for inferring the posterior distribution of the lensing potential from the lensed CMB map. We propose a solution that allows efficient Markov Chain Monte Carlo sampling from the joint posterior of the lensing potential and the unlensed CMB map using the Hamiltonian Monte Carlo technique. The main conceptual step in the solution is a re-parameterization of CMB lensing in terms of the lensed CMB and the “inverse lensing” potential. We demonstrate a fast implementation on simulated data, including noise and a sky cut, that uses a further acceleration based on a very mild approximation of the inverse lensing potential. We find that the resulting Markov Chain has short correlation lengths and excellent convergence properties, making it promising for applications to high-resolution CMB data sets in the future.

Key words: cosmic background radiation – dark matter – gravitational lensing: weak – methods: statistical

1. INTRODUCTION

Over the past few years, data from ground-based telescopes (ACT, SPT, Polarbear) and the Planck satellite have resulted in an unprecedented detection of weak gravitational lensing of the cosmic microwave background (CMB; Das et al. 2011; Engelen et al. 2012; Planck Collaboration 2014, 2015; the Polarbear Collaboration et al. 2014). Upcoming high-resolution, high-signal-to-noise experiments are poised to make the gravitational lensing distortion a powerful probe of cosmology, dark matter, and neutrino physics. The state-of-the-art estimator of CMB gravitational lensing, the quadratic estimator developed by Hu and Okamoto (Hu 2001; Hu & Okamoto 2002), works in part through a delicate cancellation of terms in an infinite Taylor expansion of the lensing effect on the CMB. The effect of this cancellation is particularly sensitive to foreground contaminants and sky masking, which, if not fully accounted for, limit the statistical inferential power of this new data.

Bayesian lensing is possibly the most promising alternative to the quadratic estimator. It has been known for some time that the quadratic estimator is suboptimal for high-signal-to-noise, high-resolution experiments and that a full Bayesian treatment can overcome this limitation (Hirata 2003a, 2003b). Indeed, Bayesian techniques applied to the lensed CMB observations have the potential to drastically change the way lensing is estimated and used for inference. Current frequentist estimators of the unknown lensing potential treat the unlensed CMB as a source of shape noise that is marginalized out. Conversely, a Bayesian lensing posterior treats the lensing potential and the unlensed CMB as joint unknowns, thereby obtaining scientific constraints jointly rather than marginally. Moreover, the posterior distribution is easier to interpret and sequentially update with additional data. From the geometry of weak lensing, most of the lensing power comes from matter at high redshift $z \sim 2$. At these distances the matter distribution on large scales is well approximated by Gaussian density fluctuations. In addition, the unlensed CMB is at present indistinguishable from an isotropic Gaussian random field.

From a statistical perspective, this is a perfect scenario for Bayesian methods in that both the observations and the unknown lensing potential are physically predicted to be Gaussian random fields.

Physicists have known for some time that Bayesian methods could potentially provide next-generation lensing estimates. In their seminal review Lewis & Challinor (2006) discuss the possibility of obtaining posterior draws from the lensing potential and the unlensed CMB jointly. However, they acknowledge the main obstacle for naive Gibbs implementations:

"... given a particular lensing potential the delensed sky is given essentially by a delta function. This means that naive Gibbs iterations will not converge within a reasonable time. At the time of writing there are no known practical methods for sampling from the full posterior distribution."

In this paper we show that, indeed, there does exist a practical way to obtain Gibbs iterations that converge quickly. The solution is through a re-parameterization of the CMB lensing problem. Instead of treating the lensing potential as unknown we work with inverse lensing, or what we call anti-lensing. Surprisingly, the slowness of naive Gibbs iterations translate to fast convergence of the re-parameterized Gibbs chain.

In Section 3 we motivate our re-parameterization by analyzing a simple two-parameter statistical problem. The concepts are then applied to the Bayesian lensing problem in Section 4. The two conditional distributions in our Gibbs implementation are discussed in Sections 5 and 6. We finish with some simulation examples in Section 7.

All the code presented in this paper is written in the language Julia (Bezanson et al. 2012) and is publicly available through
3. TWO PARAMETER ANALOGY

To motivate our solution to the Bayesian lensing problem we start with a simple two-parameter statistical problem. This system has two unknown parameters $t$, $\varphi$, with a single data point given by

$$\text{data} = t + \varphi + n,$$

where $n$ denotes additive noise. In the Bayesian setting, the posterior distribution is computed as

$$P(t, \varphi|\text{data}) \propto P(\text{data}|t, \varphi)P(t, \varphi),$$

(2)

where $P(\text{data}|t, \varphi)$ denotes the likelihood of the data given $t$, $\varphi$ and $P(t, \varphi)$ denotes the prior on $t$, $\varphi$. The Gibbs sampler is a widely used algorithm for generating (asymptotic) samples from $P(t, \varphi|\text{data})$. The algorithm generates a Markov chain of parameter values $(t^j, \varphi^j), (t^2, \varphi^2), \ldots$ generated by iteratively sampling from the conditional distributions:

$$t^{j+1} \sim P(t | \varphi^j, \text{data})$$

$$\varphi^{j+1} \sim P(\varphi | t^{j+1}, \text{data}).$$

A useful heuristic for determining the convergence rate of a Gibbs chain is the extent to which the two parameters $t$ and $\varphi$ are dependent in $P(t, \varphi|\text{data})$. A highly dependent posterior $P(t, \varphi|\text{data})$ leads to a slow Gibbs chain; near independence leads to a fast Gibbs chain. Indeed, exact independence gives a sample of the posterior after one Gibbs step. A technique for accelerating the convergence of a Gibbs sampler is to find a re-parameterization of $t$ and $\varphi$ in a way that makes the posterior less dependent. In the remainder of this section we discuss a specific re-parameterization that by analogy can be applied to Bayesian lensing.

The relevant situation for Bayesian lensing is the case in which $t$ and $\varphi$ are highly negatively correlated in $P(t, \varphi|\text{data})$. This motivates re-parameterizing $(t, \varphi)$ to $(\tilde{t}, \varphi)$ where $\tilde{t} \equiv t + \varphi$ so that

$$\text{data} = \tilde{t} + n.$$

In the statistics literature, $(t, \varphi)$ has been referred to as an ancillary parameterization, whereas $(\tilde{t}, \varphi)$ is referred to as a sufficient parameterization. We note that the terms ancillary and sufficient parameterization have been used interchangeably with the nomenclature non-centered and centered parameterizations, respectively, in the statistics literature of Roberts et al. (2003), Gelfand et al. (1995), Papaspiropoulos & Roberts (2008), Papaspiropoulos et al. (2007), and Yu & Meng (2011). Figure 1 illustrates the difference between an ancillary versus sufficient posterior distribution for our simple two-parameter model. The left plot shows the posterior density contours for the ancillary parameterization $(t, \varphi)$, along with 20 steps of a Gibbs sampler. Conversely, the right plot shows the posterior density contours for the sufficient chain $(\tilde{t}, \varphi)$ with 20 Gibbs steps. Note that the negative correlation in the ancillary parameterization manifests in near independence for the sufficient chain. Indeed, the slower the ancillary chain, the faster the sufficient chain and vice versa.
4. ANCILLARY VERSUS SUFFICIENT PARAMETERS
FOR THE LENSED CMB

The ancillary parameterization presented in the previous section is analogous to the lensed CMB problem as follows:

\[ \text{data}(x) = T(x + \nabla \phi(x)) + n(x) \]

analogous to \( \text{data} = t + \varphi + n \),

where the unlensed CMB temperature field \( T \) and the lensing potential \( \phi \) are the two unknown parameters. As discussed in Section 2, the Gibbs chain based on the ancillary parameters \( T(x) \) and \( \phi(x) \) is exceedingly slow. This clearly motivates the following re-parameterization to sufficient parameters for the lensed CMB problem

\[ \text{data}(x) = \bar{T}(x) + n(x) \]

where \( \bar{T} \) now denotes the lensed CMB temperature field with no noise or beam. The sufficient chain then proceeds as

\[ \bar{T}^{i+1} \sim P\left(\bar{T}^i|\phi^i, \text{data}\right) \]

and

\[ \phi^{i+1} \sim P\left(\phi|\bar{T}^{i+1}, \text{data}\right) \]

In Section 6 we adapt an iterative message-passing algorithm, originally developed by Elsner \& Wandelt (2013) and Jasche \& Lavaux (2015), for Wiener filtering and sampling from Equation (3). In Section 5 we derive a Hamiltonian Markov Chain algorithm to sample from Equation (4). Our Hamiltonian Markov Chain algorithm relies on an approximation—motivated again by the two-parameter system—we call anti-lensing.

4.1. Anti-lensing Approximation

In the two-parameter analogy from Section 3, the relation between the sufficient parameter \( \bar{t} \) and the ancillary parameter \( t \) is given by \( \bar{t} - \varphi = t \). We refer to the corresponding relation for CMB lensing as anti-lensing:

\[ \bar{T}(x - \nabla \phi(x)) \approx T(x) \]

We distinguish between inverse lensing and anti-lensing. Inverse lensing denotes the true coordinate displacement, which, when applied to \( \bar{T} \), recovers the unlensed \( T \). Conversely, anti-lensing is given by \( -\nabla \phi \) and approximates inverse lensing.

To examine the difference between anti-lensing and inverse lensing note that an extra divergence-free potential is needed to model the inverse lensing displacement field. Indeed, let \( f(x) := x + \nabla \phi(x) \) denote the lensing map. With this notation we have

\[ \bar{T}(x) = T(f(x)) \quad \text{and} \quad T(x) = \bar{T}(f^{-1}(x)) \]

where \( f^{-1} \) is the inverse lensing map that satisfies \( x = f^{-1}(f(x)) \). Now let \( d(x) \) denote the displacement vector field for inverse lensing so that \( f^{-1}(x) = x + d(x) \). Therefore

\[ x = f^{-1}(f(x)) = f(x) + d(f(x)) \]

In particular \( d(f(x)) = x - f(x) = -\nabla \phi(x) \) which gives

\[ d(x) = -\nabla \phi(f^{-1}(x)) \]

This implies that the inverse lensing displacement is modeled as a warped version of the curl-free vector field \( \nabla \phi \) (warped by \( f^{-1} \)). This warping introduces a nonzero divergence-free term (just as lensing adds nonzero B-mode power in the CMB polarization).

To illustrate the expected magnitudes of the divergence-free and curl-free terms, start with a Helmholtz decomposition of the inverse lensing displacement:

\[ d(x) = -\nabla \phi^\text{inv}(x) - \nabla^\bot \psi^\text{inv}(x) \]

where \( \nabla^\bot := \left( -\frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right) \) and \( \psi^\text{inv} \) denotes a stream function potential, which models a field rotation so that

\[ \bar{T}(x - \nabla \phi^\text{inv}(x) - \nabla^\bot \psi^\text{inv}(x)) = T(x) \]

We claim that the expected size of the inverse lensing displacement \( d(x) \) is on the order of arcmin but the correlation length scale of \( \phi \) is on the order of degrees, we claim that \( -\nabla \phi(f^{-1}(x)) \) is well approximated by \( -\nabla \phi(x) \). In particular, the divergence-free term \( -\nabla^\bot \psi^\text{inv} \) is small and

\[ -\nabla \phi \approx -\nabla \phi^\text{inv} \approx -\nabla \phi^\text{inv} - \nabla^\bot \psi^\text{inv} = d \]
Figure 2 illustrates the magnitudes of the above terms. The anti-lensing potential $-\phi$ is shown (upper-left) along with the corresponding inverse lensing potential $-\phi^{\text{inv}}$ (upper-right). The difference $\phi^{\text{inv}} - \phi$ is also shown (bottom left) along with the stream function $\psi^{\text{inv}}$ (bottom right). Clearly, the magnitude of the difference $\phi^{\text{inv}} - \phi$ and $-\psi^{\text{inv}}$ is subdominant to the estimation error expected in current lensing experimental conditions.

5. HAMILTONIAN MONTE CARLO SAMPLER FOR $P(\phi|T, \text{DATA})$

The Hamiltonian Monte Carlo (HMC) algorithm is an iterative sampling algorithm designed to mitigate the low-acceptance rate of the Metropolis–Hastings algorithm when working in a high dimension. A nice review of the HMC can be found in Neal (2011). For applications of the HMC in cosmology see Hajian (2007), Taylor et al. (2008), Elsner & Wandelt (2010), Jasche et al. (2010), and Jasche & Wandelt (2012, 2013a, 2013b). In the present case we utilize the HMC algorithm to produce samples of $\phi$ from $P(\phi|T, \text{data})$. The key to making the HMC work for lensing is to parameterize $\phi$ in terms of its Fourier transform. One can then utilize Claim 1, presented below, to efficiently compute the gradient of the log conditional density of $P(\phi|T, \text{data})$, which is a necessary computation for the HMC algorithm.

Figure 2. Difference between anti-lensing and inverse lensing. Upper left: anti-lensing potential $-\phi$. Upper right: the inverse lensing potential $-\phi^{\text{inv}}$. Bottom left: the difference $\phi^{\text{inv}} - \phi$. Bottom right: the inverse lensing stream function $-\psi^{\text{inv}}$.

Notation: throughout the remainder of this paper, the Fourier transform of any function $f(x)$ will be denoted by $f_l$ or $f_k$ so that $f_l = \int_{\mathbb{R}^2} e^{-i\mathbf{x}\cdot\mathbf{l}} f(x) \, dx$ and $f(x) = \int_{\mathbb{R}^2} e^{i\mathbf{x}\cdot\mathbf{l}} f_l \, d\mathbf{l}$, where $l \in \mathbb{R}^2$ is a two-dimensional frequency vector and $x \in \mathbb{R}^2$ is a two-dimensional spatial coordinate.

To describe the HMC algorithm let $\phi$ denote the concatenation of the real and imaginary parts of $\phi_l$ as $l$ ranges through discrete frequencies $l$ ranging up to a pre-specified $l_{\max}$ but excluding half of the Fourier frequencies due to the Hermitian symmetry associated with the Fourier transform of a real field. Note that $\phi$ is a vector of real numbers. Let $P(\phi|T, \text{data})$ denote the density of $\phi$ given $T$ and the data. Let $p$ denote a “momentum” vector and $m$ denote a “mass” vector, which are both the same length as $\phi$. The Hamiltonian is a function of $\phi$ and $p$ and is defined as follows:

$$H(\phi, p) = -\log P(\phi|T, \text{data}) + \sum_l \frac{p_l^2}{2m_l^2}.$$ 

This Hamiltonian generates a time-dependent evolution of $\phi$ and $p$ given by

$$\frac{d\phi^i}{dt} = \nabla_\phi H(\phi^i, p^i)$$

$$\frac{dp^i}{dt} = -\nabla_p H(\phi^i, p^i).$$
The HMC is a discrete version of this time-dynamic equation, using a leapfrog method, which produces a Markov chain $(\phi_1, p_1), (\phi_2, p_2), \ldots$, where the $i$th iteration is given by Algorithm 1 below.

**Algorithm 1.** $i$th step of the Hamiltonian Markov Chain

1. Set $\phi^0 = \phi_{i-1}$ and simulate $p^0 \sim N(0, \Lambda_n)$, where $\Lambda_n$ is diagonal with $\text{diag}(\Lambda_n) = m$.
2. Recursively compute $\phi^{k+1}$ and $p^{k+1}$ for $k = 1, \ldots, n$ using the following equations:
   $$
   \phi^{k+1} := \phi^k + \epsilon \Lambda_n^{1/2} p^k - \frac{\epsilon}{2} \nabla_{\phi} H(\phi^k, p^k),
   $$
   $$
   p^{k+1} := p^k - \epsilon \left[ \nabla_{\phi} H(\phi^k, p^k) + \nabla_{\phi} H(\phi^{k+1}, p^k) \right].
   $$
3. Simulate $u \sim U(0, 1)$ and define $p = \min \left( 1, \frac{1}{e^{-\epsilon(\phi^{k+1})}/e^{-\epsilon(\phi^k)}} \right)$.
4. If $u < p$, set $\phi_i = \phi^{k+1}$, otherwise set $\phi_i = \phi_{i-1}$.

The HMC algorithm is notoriously sensitive to tuning parameters. The prevailing wisdom (see Neal 2011) is that one should set $m$ to match the reciprocal of the posterior variance of $\phi$. For the simulation presented in Section 7 we simply set $m^{-1} = 2 \times 10^3$ to be nearly proportional to $C_l^{\phi\phi}$ with a slight attenuation at low wavenumber. In particular, we set $m^{-1} := 2 \times 10^3 \frac{3}{4} + \frac{1}{4} \tanh \left( \frac{11 - 1500}{200} \right) C_l^{\phi\phi} \delta_0$, where $\delta_0$ denotes the constant obtained by evaluating the discrete Dirac delta function at $l = 0$ (see the appendix for a detailed discussion). This choice was motivated by the fact that the high-frequency terms $\phi_l$ are not well constrained by the posterior distribution, which results in a posterior variance closely matching $C_l^{\phi\phi}$. The remaining parameters of Algorithm 1 are set to $n = 30$ and $\epsilon = 2 \times 10^{-3} u$, where $u$ is a uniform (0, 1) random variable sampled anew at each pass of Algorithm 1 (the use of random $c$ is designed to avoid resonant frequencies, as advocated in Taylor et al. 2008).

The key difficulty in using Algorithm 1 is the computation of the $\nabla_{\phi} H(\phi^k, p^k)$, or equivalently the computation of $\nabla_{\phi} \log P(\phi|T)$, or data). The number of frequencies is extremely large and therefore any slow computation of the gradients will present a serious bottleneck. The following claim shows that the gradient of the log density of $P(\phi|T)$, data), with respect to the Fourier basis of $\phi$, can be computed quickly with Fourier and inverse Fourier transforms.

**Claim 1.** Under the anti-lensing approximation (5) for any nonzero frequency vector $l \equiv (l_1, l_2) \in \mathbb{R}^2$

$$
\begin{align*}
\frac{\partial}{\partial \phi_l} \log P(\phi|T), \text{ data}) &\propto -\frac{\phi_l}{C_l^{\phi\phi}} \\
&\quad - \sum_{q=1,2} i q \int_{\mathbb{R}^2} e^{-i q A^l(x)} B(x) \frac{dx}{2\pi},
\end{align*}
$$

where $\phi_l = \text{re} \phi_l + i \text{im} \phi_l, \frac{\partial}{\partial \phi_l} \equiv \frac{\partial}{\partial \text{re} \phi_l} + i \frac{\partial}{\partial \text{im} \phi_l}$ and

$$
B_l \equiv \frac{1}{C_l^{TT}} \int_{\mathbb{R}^2} e^{-i q A^l(x)} (x - \nabla \phi(x)) \frac{dx}{2\pi},
$$

and

$$
A^l(x) \equiv \frac{\partial T}{\partial x_q} (x - \nabla \phi(x)).
$$

An important fact used in the derivation of Equation (7) is that the lensing and anti-lensing operator is invertible. For example, if $\phi(x)$ and $T(x)$ are known at all pixel locations $x$, then it is possible to perfectly reconstruct $T(x)$. This implies that the anti-lensing operation (which is a linear action on the CMB) can be represented as an infinitesimal permutation matrix. Therefore, the determinant of the anti-lensing operator $\det(dT^\phi/dT)$ equals 1, where $T^\phi(x) \equiv T(x - \nabla \phi(x))$. Now to compute the likelihood surface $P(\phi|T)$, data) as a function of $\phi$ we obtain the following formula:

$$
P(\phi|T, \text{ data}) = P(\phi|T) \times P(T|\phi)P(\phi)
$$

$$
= \frac{\det(dT^\phi/dT)}{P(T^\phi)} P(T^\phi) P(\phi),
$$

where $P(T^\phi|\phi)$ represents the likelihood that $T^\phi$ is statistically unlensed by $\phi$. In other words, $P(T^\phi|\phi)$ measures the likelihood that $T^\phi$ is an isotropic Gaussian random field with spectral density $C_l^{TT}$. This explains the following characterization of the log likelihood of $\phi$ given $T$ and the data:

$$
\log P(\phi|T, \text{ data}) = c - \frac{1}{2} \int_{\mathbb{R}^2} \frac{\left| \frac{T^\phi}{C_k^{TT}} + \frac{\left| \phi_l \right|^2}{C_l^{\phi\phi}} \right|^2}{2\pi},
$$

(10)

where $c$ is a constant that does not depend on $\phi$. The remaining details of the derivation of Claim 1 are provided in the appendix.

It is instructive to compare the gradient calculation (Equation (7)) with the quadratic estimate of $\phi$ developed in Hu (2001) and Hu & Okamoto (2002). The quadratic estimate, applied to observations of the form $T(x) + n(x)$, is given by

$$
\hat{\phi}_l = -N_l \sum_{q=1,2} i q \int_{\mathbb{R}^2} e^{-i q A^l(x)} B(x) \frac{dx}{2\pi},
$$

(11)

where

$$
N_l \equiv \left[ C_l^{TT} + C_l^{nn} \right]^{-1} \left[ \tilde{H}_l + n_l \right]
$$

and

$$
A^l \equiv i q \left[ C_l^{TT} \left[ C_l^{TT} + C_l^{nn} \right]^{-1} \right] \left[ \tilde{H}_l + n_l \right]
$$

The term $N_l$ is radially symmetric in frequency $l$ and corresponds to a normalization that makes the quadratic estimate unbiased up to first order. After substituting $C_l^{TT} + C_l^{nn} \rightarrow C_l^{TT}$ and $n_l \rightarrow 0$ in the formula for the
quadratic estimate, one obtains
\[
\frac{\partial}{\partial \phi_i} \log P(\phi|\tilde{T}, \text{data})|_{\phi=0} = \frac{\hat{\phi}_i}{N_l}.
\]
Indeed, an approximate Newton step, using Equation (7), is an accurate approximation to the quadratic estimate \(\hat{\phi}_i\). This illustrates how the parameterization \((\tilde{T}, \phi)\) results in Gibbs iterations that make drastic moves, on the order of the size of the quadratic estimate.

One of the features of the quadratic estimate is that the fast Fourier transform (FFT) and inverse fast Fourier transform (IFFT) can be used to compute \(\hat{\phi}_i\) for all frequencies \(l\). Naively computing the quadratic form of the quadratic estimate requires \(O(n^2)\) flops, rather than the \(O(n \log n)\) flops obtained by the FFT/IFFT method, where \(n\) denotes the number of pixels. We note that Claim 1 establishes that the gradient computation inherits a similar FFT/IFFT characterization to compute \(\frac{\partial}{\partial \phi_i} \log P(\phi|\tilde{T}, \text{data})\) at all frequencies \(l\), in \(O(n \log n)\) flops. Since this gradient computation needs to be embedded in a Hamiltonian Markov step within a Gibbs Chain, the computation efficiency gained by the FFT/IFFT is absolutely crucial.

6. ITERATIVE MESSAGE PASSING ALGORITHM FOR \(P(T|\phi, \text{DATA})\)

There are two natural ways to model the lensed CMB \(T\). If one marginalizes out \(\phi\), then \(T\) is modeled as a non-Gaussian but isotropic random field. Conversely, if one conditions on \(\phi\) the field \(T\) is modeled as a non-isotropic but Gaussian random field. The latter case is relevant for sampling from \(P(T|\phi, \text{data})\), which is therefore simply a Gaussian conditional simulation problem. Unfortunately, the non-isotropic (indeed, non-stationary) nature of the conditional distribution of \(T\) presents serious computational challenges. In what follows we extend a new iterative algorithm developed in Elsner \\& Wandelt (2013) for computing the Gaussian conditional expectation when the signal is diagonalized in harmonic space and the noise is diagonalized in pixel space. The method we present here is similar to the Gibbs sampling adaptation of Jasche \\& Lavaux (2015).

Start by transforming each pixel location \(x\) by the lensing operation \(x + \nabla \phi(x)\) while simultaneously preserving the data associated with that pixel. This effectively de-lenses \(\text{data}(x) = T(x + \nabla \phi(x)) + n(x)\), but produces observations on an irregular grid. In particular, one may switch to the lensed coordinates \(y = x + \nabla \phi(x)\) so that
\[
(x + \nabla \phi(x), \text{data}(x)) = (y, T(y) + \tilde{n}(y)),
\]
where \(\tilde{n}(x + \nabla \phi(x)) = n(x)\). Now the data \((y, T(y) + \tilde{n}(y))\) is arranged on an irregular grid in \(y\). This irregular grid is then embedded into a high-resolution regular grid by nearest neighbor interpolation. The points \(y\) which do not get assigned an observation \(T(y) + \tilde{n}(y)\) under the interpolation, we consider to be masked. Figure 3 illustrates this situation. The left plot shows the irregularly sampled data \((x + \nabla \phi(x), \text{data}(x))\) and the right plot shows the grid embedding. The filled dots represent observations of \(T(y) + \tilde{n}(y)\), whereas the empty dots correspond to a masked observation of \(T(y)\). Finally, we extend the definition of \(\hat{n}(y)\) to have infinite variance over the masked region, thereby producing data \(T(y) + \tilde{n}(y)\) over a dense regular grid in \(y\).

As an intermediate step in producing a sample from \(P(T|\phi, \text{data})\) we produce a conditional sample of \(T(y)\) given the observations \(T(y) + \tilde{n}(y)\). The difficulty of this step is that \(\tilde{n}(y)\) is non-homogeneous noise—from the masking and any inhomogeneity in \(n(x)\)—and therefore it is not decorrelated by the Fourier transform. To handle this situation we adapted a new method for Gaussian conditional expectation developed in Elsner \\& Wandelt (2013). This method works particularly well for observations with large amounts of irregular masking, as in our case. The algorithm utilizes a messenger field that effectively behaves as a latent—signal plus white noise—model that is amenable to Gibbs sampling (Jasche \\& Lavaux 2015).

The delensing algorithm described in this paper requires the capability to do fast constrained realization of non-lensed CMBs. The embedding illustrated in Figure 3 means that each constrained realization needs to be computed on a mask with a great deal of structure on the scale of the pixels of dense grid. Several algorithms exist in the literature to solve the general problem of constrained Gaussian random field on a given mask and power spectrum: the conjugate gradient method (Eriksen et al. 2004; Wandelt et al. 2004); the multiscale conjugate gradient method (Smith et al. 2007); the multigrid method (Seljebotn et al. 2014); and the Messenger algorithm (Elsner \\& Wandelt 2013) and its variant the Gibbs-Messenger (Jasche \\& Lavaux 2015).

Since the lensing potential changes from each iteration to the next, every constrained realization of non-lensed CMB needs to be computed for a different set of active points in the embedding grid. This effectively means that the solution is done for a different mask at every iteration. This rules out linear solvers that require expensive pre-computations, e.g., of pre-conditioners, that depend on the coefficient matrix of the system since that depends on the mask. We also require exact acceptance and higher speed than direct or standard iterative methods. This reduces the possibilities to either the Messenger algorithm or the Gibbs-Messenger.

The Gibbs-Messenger generates very fast constrained realizations that converge to the correct distribution in a statistical sense without iterating to a numerical solution (thus obviating the need to specify a cooling schedule in Algorithm 2) for the price of losing independence between subsequent samples. In contrast the Messenger algorithm simulates independent constrained realizations, but requires iteration of the linear system. To ensure a numerically accurate solution implies a conservative choice of cooling schedule in Algorithm 2 (though this is still much faster than the alternatives described in the previous paragraph).

We adopt a hybrid approach where we occasionally generate an independent sample using the Messenger algorithm, and then generate many quick samples using the Gibbs-Messenger approach. The detailed choices for the cooling schedule and the number of samples between full Messenger solutions will be described in Section 7.
In observations of the form \( y = T(x) + \hat{n}(y) \), where \( T(x) \) denotes the unlensed CMB and \( \hat{n}(y) \) denotes the noise. The unobserved locations, represented by open circles, are characterized with an infinite variance for \( m(y) \). The right panel shows a sampled mean zero Gaussian random field \( Z(y) \) that is independent across pixels and has a pointwise variance

\[
\frac{1}{\lambda_i \sigma^2 + 1 \delta^2(\nu_i)}.
\]

3: Return the unobserved locations, \( T(x) = T(x + \nabla \phi(x)) \).

Algorithm 2. Algorithm for sampling from \( P(T|x) \), where data \( y = T(x) + \hat{n}(y) \).

1: Set cooling schedule \( \lambda_1 \geq \ldots \geq \lambda_n \), where \( \lambda_n = 1 \).
2: Decompose \( \hat{n}(y) \) into a homogeneous part with variance \( \sigma^2 \) and a non-homogeneous part with variance \( \sigma^2(y) \) so that

\[
\text{var}(\hat{n}(y)) = \sigma^2 + \sigma^2(y),
\]

where \( \hat{n}(y) \) is \( \infty \) on all masked pixels \( y \). Notice that the spectral density of the homogeneous part is given by \( \delta^2 dy \), where \( dy \) denotes the pixel grid area.
3: Initialize the fields \( y_T \) and \( y_T(y) \) to be zero on all pixel locations \( y \).
4: Recursively update fields \( y_T \) and \( y_T(y) \) by the following steps for \( j = 1, \ldots, n \):
   (a) Simulate a mean zero Gaussian random field \( Z(y) \) that is independent across pixels and has a pointwise variance
   \[
   \frac{1}{\lambda_i \sigma^2 + 1 \delta^2(\nu_i)}.
   \]
   (b) Update \( M(y) \) by
   \[
   M(y) = y_T(y) - \text{data}(y) \frac{\lambda_i \sigma^2}{\lambda_i \sigma^2 + \sigma^2(y)} + T(y) \frac{\sigma^2(y)}{\lambda_i \sigma^2 + \sigma^2(y)} + Z(y).
   \]
   (c) Simulate a mean zero Gaussian random field, \( W(y) \), with spectral density
   \[
   \langle W(y) W(y') \rangle = \delta_{x,y'} \left( \frac{1}{\lambda_i \sigma^2} + \frac{1}{\delta^2(\nu_i)} \right)^{-1}.
   \]
   (d) Update \( T(x) = T(x + \nabla \phi(x)) \).
5: Return \( T(x) \).

Indeed, our goal is to explore the low noise and small beam experimental conditions where the quadratic estimate is known to be suboptimal (see Hirata 2003a, 2003b). That being said, the only change needed to incorporate other experimental details in the Bayesian lensing methodology presented here, including foreground contaminants, is how one samples from \( P(T|\psi, \text{data}) \). Algorithms 2 and 3 take advantage of the special lensed-grid structure of the data when conditioning on \( \phi \) to accomplish this goal. Adding different/new experimental details to the data will still result in a Gaussian constrained realization problem. We acknowledge that more complicated modeling of the data will most certainly introduce additional computational challenges. However, we consider these computational challenges to be sub-dominant to the fundamental bottleneck for Bayesian lensing, which was the extremely slow mixing time of the original Gibbs formulation. Moreover, the structure of the new Gibbs formulation isolates all experimental details to the Gibbs step (Equation (4)), where conditioning on \( \phi \) results in a classic Gaussian constrained realization problem.

7. SIMULATION EXAMPLE

In this section we present a simulation to illustrate the methodology presented above. The simulated lensing potential used in this section, shown at left in Figure 5, is generated on a flat sky with periodic boundary conditions. The data, shown upper-left in Figure 6, are generated on 2 arcmin pixels with independent additive noise and masking. The noise level is set to 8.0 \( \mu \)K arcmin and the masking covers approximately 10% of the pixels. The parameters of the Bayesian lensing procedure are the Fourier modes of \( \phi \) and \( \phi \). For the lensing potential we set \( |l| \) to 460. For this sky coverage the scale-resolution in Fourier space \( \Delta l = 21 \) yields 1500 unknown Fourier coefficients for \( \phi \). The \( |l| \) of 2700 for the unlensed temperature \( T \) is set in Algorithm 2 and corresponds to half of the Nyquist limit at 2 arcmin pixels. CAMB (Lewis et al. 2000) is used to generate the fiducial theoretical power spectra, \( C_l^{TT} \) and \( C_l^{\phi \phi} \), for simulating the data and for the Bayesian priors on \( T \) and \( \phi \) (both fields are assumed to be Gaussian). The fiducial cosmology used in CAMB is based on a flat, power-law CDM cosmological model, with a baryon density of
Figure 4. Figure shows values of the cooling parameter $\lambda_j$ for $j = 1, 100, 300, 500, 700, 900$ used in Algorithm 2. This schedule is applied every 100th step in the Gibbs algorithm. In Algorithm 2, the value of $\lambda_j\sigma^2 dy$ serves as the spectral density of artificial additive white noise in the latent field $M(x)$. Therefore setting $\lambda_j$ to greater than 1 encourages fast mixing of $T_i$ at all frequency vectors $l$ such that $C_{TT} \leq \lambda_j\sigma^2 dy$. The cooling schedule shown above is an attempt to let $\lambda_j$ approach 1 in such a way as to encourage all frequency vectors up to $|l|_{\text{max}}$ to mix quickly.

$\Omega_b = 0.044$; a cold dark matter density of $\Omega_{\text{cdm}} = 0.21$; a cosmological constant density of $\Omega_{\Lambda} = 0.74$; Hubble parameter $h = 0.71$ in units of 100 km s$^{-1}$ Mpc$^{-1}$, primordial scalar fluctuation amplitude $A_s(k = 0.002$ Mpc$^{-1}) = 2.45 \times 10^{-9}$; scalar spectral index $n_s(k = 0.002$ Mpc$^{-1}) = 0.96$; primordial helium abundance $Y_p = 0.24$; and a reionization optical depth of $\tau_e = 0.088$.

We ran 10 parallel Gibbs chains for a total of 2500 steps. The timings for each Gibbs iteration averaged approximately 200 s using a Dual Intel Xeon E5-2690 2.90 GHz processor. Each chain was initially warmed up by replacing the HMC draws in the first five iterations with a gradient ascent. A burn-in of approximately 550 runs were discarded and the remaining runs were thinned by 100. The result is a total of 200 posterior samples. The cooling schedule for the iterative message-passing algorithm was selected by numerical experimentation. Most of the Gibbs iterations set the cooling terms $(\lambda_1, \ldots, \lambda_{100}) \equiv (1, \ldots, 1)$ in Algorithm 2. However, we did find it advantageous to periodically run a nontrivial 1000-step cooling schedule for every 100th pass of the Gibbs algorithm. This nontrivial cooling schedule for $\lambda_j$ is plotted in Figure 4 and is set in an attempt to encourage fast mixing of the Fourier modes $T_i$ up to $|l|_{\text{max}}$.

The best Bayesian estimate of $\phi(x)$ corresponds to the posterior mean $\tilde{E}(\phi(x)|\text{data})$. This quantity is approximated by the average of the 200 draws from the Gibbs chain and is shown in the middle plot of Figure 5. The right plot of Figure 5 shows the quadratic estimate of $\phi(x)$ for comparison. However, due to the difficulty when using the quadratic estimate in the presence of sky cuts, the quadratic estimate shown uses all of the data—including the pixels that are masked—in producing the estimate of $\phi$. In general, one can see good agreement with $E(\phi(x)|\text{data})$ and $\phi(x)$. Indeed, the effect of masking is visually undetectable as compared to the quadratic estimate. To get a better visualization of the individual draws from the posterior, the left plot in Figure 7 shows a horizontal cross-section of the posterior draws of $\phi(x)$ taken at vertical degree mark 12:7.

The Gibbs methodology presented here yields samples of the lensed CMB $T(x)$ and the lensing potential $\phi(x)$ conditional on the data. Moreover, as a by-product of Algorithm 3, we also obtain samples of the unlensed $T(x)$ given the data. By averaging 200 draws from the Gibbs chain one can construct an approximation to $E(T(x)|\text{data})$, shown in the upper-right plot of Figure 6. In the bottom left plot of Figure 6 we show the difference $T(x) - E(T(x)|\text{data})$. Compared to the nominal difference between lensed and unlensed CMB $T(x) - T(x)$, shown bottom right in Figure 6, one can see that the Gibbs methodology is successful at delensing the observed CMB. To get a better visualization of the individual draws from the posterior, the right plot in Figure 7 shows a horizontal cross-section of the posterior draws of $T(x)$ taken at vertical degree mark 12:7 and is magnified near the masking region for better visual inspection.

In Figures 8 and 9 we summarize the posterior draws for $\phi$ and $T$ in the Fourier domain. The left plot of Figure 8 shows 95% posterior regions for $l^4|\phi_l|^2/(4\delta_0)$ averaged over $l$ in wavenumber bins. For comparison, the simulation true values of $l^4|\phi_l|^2/(4\delta_0)$ are shown in red and the spectral density $l^2C_{TT}^{\text{sim}}/4$ is plotted in the black solid line. The same quantities are shown for $l^4|T_l|^2/(\delta_0)$ in the right plot of Figure 8. Finally, in Figure 9 we show the empirical cross correlation of the posterior draws over wavenumber bins between the simulation truth and the posterior samples of $\phi_l$ and $T_l$. In particular, samples were generated from $\frac{1}{\tau} \sum_{l \in \Delta l} \phi_l^{\text{sim}} \phi_l^*$ and $\frac{1}{\tau} \sum_{l \in \Delta l} T_l^{\text{sim}} T_l^*$, where $\Delta l$ is a frequency wavenumber bin, $\phi_l$ and $T_l$ are the simulation truth, $\phi_l^{\text{sim}}$ and $T_l^{\text{sim}}$ are sampled from the Gibbs algorithm presented here and $c$ is a normalization constant that transforms to a correlation scale. Notice that the plotted correlations trend to 0 for larger wavenumbers. This is what one would expect since larger wavenumbers have correspondingly smaller signal-to-noise ratios, which causes the posterior to revert back to the prior.

In Figure 10 we show the Gibbs chain correlation length scale and the speed of mixing for different statistics of the lensing potential. The left plot shows the Gibbs chain for $\phi(x)$ where $x = (9^\circ, 9^\circ, 13^\circ, 2)$ and $x = (1^\circ, 6, 1^\circ, 6)$ in the same degree coordinates given in Figures 6 and 5. The right plot shows the real and imaginary parts of $\phi_l$ where the frequency vector $l$ is set to $(126.56, 63.28)$. Each dashed line represents the corresponding simulation truth parameters. These plots suggest that the Gibbs chain is mixing well and that the correlation length scale is small enough so that thinning by 100 is sufficient to yield relatively uncorrelated samples.

8. CONCLUDING REMARKS

In this paper we construct a prototype algorithm that establishes that it is possible to construct a fast Gibbs sampler of the Bayesian posterior for the unknown lensing potential and the de-noised CMB temperature map. This prototype solves one of the fundamental obstacles in a Gibbs implementation of the Bayesian lensing problem: the naive parameterization $(T, \phi)$ is extremely slow. We identify the ancillary and sufficient parameterization duality for this problem and notice that the slowness of the Gibbs chain for the ancillary
parameterization \((T, \phi)\) translates to a fast chain for the sufficient parameterization \((\tilde{T}, \phi)\). This observation is one of the main contributions of this paper. The second contribution is the use of the anti-lensing approximation along with Claim 1, which makes feasible the development of a Hamiltonian Markov Chain algorithm for sampling from \(P(\phi|\tilde{T})\). Without the Fourier transform characterization in Claim 1 the HMC would be computationally prohibitive. The third contribution of this paper is to recognize that a new messenger algorithm from Elsner & Wandelt (2013) and Jasche & Lavaux (2015) can be
adapted for high-resolution conditional Gaussian sampling under the irregular sampling scenario needed for $P(T|\phi, \text{data})$.

Notice that both sampling steps $P(\phi|T)$ and $P(T|\phi, \text{data})$ in our algorithm utilize a high-resolution embedding for $T$. This high-resolution embedding is most likely the dominant bottleneck for scaling the current prototype implementation presented here. In this paragraph we discuss what is needed to avoid using this embedding for scaling up this algorithm.

When sampling from the conditional $P(T|\phi, \text{data})$, the main challenge is to compute $A^T(x)$ and $B(x)$, as defined in Claim 1. Within the HMC algorithm, a proposed lensing potential $\phi(x)$ changes iteratively. At each iteration one requires a new computation of $A^T(x)$ and $B(x)$. In our prototype, a spline interpolation performs the task of fast anti-lensing required for $A^T(x)$ and $B(x)$. It is an open problem how to compute this fast anti-lensing without the need for a high-resolution $T$. Simulating from $P(T|\phi, \text{data})$ also requires a high-resolution embedding in our prototype. This simply expresses the fact that given $\phi$ the field $T$ is modeled as a non-stationary random field. To circumvent this difficulty we transform to lensed coordinates as illustrated in Figure 3. The challenge when avoiding this high-resolution embedding, then, is to directly generate conditional simulations of the non-stationary $T$ given $data(x) = T(x) + n(x)$ and the lensing potential $\phi(x)$.

The authors wish to thank the anonymous referee for constructive comments and suggestions. B.W. acknowledges funding through his Chaire d’Excellence from the Agence Nationale de la Recherche (ANR-10-CEXC-004-01). This work has been performed within the Labex ILP (reference ANR-10-LABX-63) part of the Idex SUPER, and received financial state aid managed by the Agence Nationale de la Recherche.

![Figure 7](image-url)  
**Figure 7.** Here we plot one-dimensional slices of the posterior draws from $P(\phi(x)|\text{data})$ and $P(T(x)|\text{data})$. For reference, these slices are taken from the horizontal regions at vertical degree mark 12.7 in Figures 6 and 5.

![Figure 8](image-url)  
**Figure 8.** Estimates of $l^4|\phi|^2/4$ and $l^2|T|^2$ (shown in blue), scaled to the units of the corresponding spectral density. The red dots show $l^4|\phi|^2/4$ and $l^2|T|^2$ for the simulation truth (similarly scaled). The discrepancy between the red dots and the spectral densities, shown in black, is exclusively due to cosmic variance. The confidence bars show 95% probability regions from the posterior distributions $P(l^4|\phi|^2/4|\text{data})$ and $P(l^2|T|^2|\text{data})$.
Recherche, as part of the programme Investissements d’avenir under the reference ANR-11-IDEX-0004-02. E.A. acknowledges grant support from NSF CAREER DMS-1252795.

APPENDIX

Before we proceed to the proofs we briefly discuss notation. First, we do not differentiate, notionally, a random field with periodic boundary conditions on \((-L/2, L/2)^2\) and the case where \(L \to \infty\) so that the Fourier series \(\sum_{q \in \mathbb{Z}^2} \hat{f}(q/L) e^{2\pi i q \cdot x/L}\) converges to the continuous Fourier transform \(\int_{\mathbb{R}^2} e^{2\pi i f(t \cdot x)} dt\). For example, at times we will refer to an infinitesimal area element \(drl\) or \(dkt\) in Fourier space, which simply equals \((2\pi/L)^2\) for large \(L\). In this case, \(\delta_l\) denotes a discrete Dirac delta function that we equate with \(1/dl\) when \(l = 0\) and zero otherwise. Indeed, throughout the paper it will be convenient to use \(\delta_0 \equiv \delta_1 \mid_{l=0}\) to denote the constant \(1/dl\). Second, for any function \(f(x)\) let \(f^\phi(x) = f(x - \nabla \phi(x))\) denote anti-lensing of \(f\) and \(f^\phi l\) denote the Fourier transform of \(f^\phi(x)\).

Proof of Claim 1. Since \(\mathcal{T}\) is sufficient for the unknown \(\phi\) we have that

\[
P(\phi | \mathcal{T}, \text{data}) = P(\phi | \mathcal{T}) \propto P(\mathcal{T} | \phi) P(\phi).
\]

Since \(\phi(x)\) is an isotropic random field with spectral density \(C_l^\phi\) we have that \(E(\phi_l \phi^*_l) = \delta_{l-l'} C_l^\phi\). Therefore \(E(\phi_l \phi_l^*) = \delta_0 C_0^\phi\) and \(E(\phi_l \phi_l) = 0\) implies that the random variables \(\text{re } \phi_l\) and \(\text{im } \phi_l\) are independent \(\mathcal{N}(0, \frac{1}{2} \delta_0 C_0^\phi)\) for each fixed \(l\). Moreover, \(\phi(x)\) takes values in \(\mathbb{R}\) so that \(\phi_l = \phi_l^*\).

This implies that \(\phi_l\) are independent random variables over all \(l\), which are restricted to the Hermitian half of the Fourier grid, denoted here as \(\mathbb{H}\). In particular, if we exclude the zero frequency \(l = 0\) we get

\[
\log P(\phi) - c_1 = -\frac{1}{2} \sum_{k \in \mathbb{H}\setminus \{0\}} \left[ \frac{\text{re } \phi_k}{2} \delta_0 C_k^\phi + \frac{\text{im } \phi_k^*}{2} \delta_0 C_k^\phi \right]
= -\frac{1}{2} \int_{\mathbb{R}^2} |\phi_k|^2 C_k^\phi dk
\]

and

\[
\log P(\mathcal{T} | \phi) - c_2 = -\frac{1}{2} \sum_{k \in \mathbb{H}\setminus \{0\}} \left[ \frac{\text{re } T_k}{2} \delta_0 C_{kT} + \frac{\text{im } T_k^*}{2} \delta_0 C_{kT} \right]
= -\frac{1}{2} \int_{\mathbb{R}^2} |T_k|^2 C_{kT} dk.
\]

where \(c_1\) and \(c_2\) are constants and \(T^\phi(x) \equiv T(x - \nabla \phi(x))\). Taking derivatives in (12) gives

\[
\frac{\partial}{\partial \phi_l} \log P(\phi) = -2\langle d(l) \rangle C_l^\phi.
\]

Taking derivatives in (13) gives

\[
\frac{\partial}{\partial \phi_l} \log P(\mathcal{T} | \phi) = -\text{re} \int_{\mathbb{R}^2} \frac{\partial T_k}{\partial \text{re } \phi_l} \frac{T_k^*}{C_{kT}} dk
\]

and

\[
\frac{\partial}{\partial \phi_l} \log P(\mathcal{T} | \phi) = -\text{im} \int_{\mathbb{R}^2} \frac{\partial T_k}{\partial \text{im } \phi_l} \frac{T_k^*}{C_{kT}} dk.
\]

Taking linear combinations of the two equalities in Lemma 1 below we get

\[
\frac{\partial T_k^\phi}{\partial \text{re } \phi_l} = \frac{1}{2} \frac{\partial T_k}{\partial \text{re } \phi_l} + \frac{1}{2} \frac{\partial T_k}{\partial \phi_l^*}
= \frac{dk}{2\pi} \sum_{q=1,2} t_q \left[ \left| (\nabla_q T)^\phi \right|_{l-i} - \left| (\nabla_q T)^\phi \right|_{l+i} \right]
\]

and

\[
\frac{\partial T_k^\phi}{\partial \text{im } \phi_l} = \frac{i}{2} \frac{\partial T_k}{\partial \text{re } \phi_l} + \frac{i}{2} \frac{\partial T_k}{\partial \phi_l^*}
= \frac{dk}{2\pi} \sum_{q=1,2} t_q \left[ \left| (\nabla_q T)^\phi \right|_{l-i} - \left| (\nabla_q T)^\phi \right|_{l+i} \right].
\]
real, which implies
\[
\frac{\partial}{\partial \phi_l} \log P(\tilde{T}|\phi) = -\int_{\mathbb{R}^2} \tilde{T}_k^\phi \left( \frac{\partial \tilde{T}_k^\phi}{\partial \phi_l} \right) dk
\]
\[
= -\frac{dk}{\pi} \sum_{q=1,2} i l_q \int_{\mathbb{R}^2} \left( \nabla q \tilde{T} \right)^q_{k+l} \tilde{T}_k^{\phi^*} dk
\]
\[
= -i2(dk) \sum_{q=1,2} l_q \int_{\mathbb{R}^2} e^{-i\pi l} A^q(x) B(x) \frac{dx}{2\pi},
\]
by Lemma 3 below,

where \( A^q(x) \equiv (\nabla q \tilde{T})^q(x) \) and \( B_k \equiv (T_k^{\phi^*})/C_k^{TT} \).

**Lemma 1.**
\[
\frac{\partial \tilde{T}_k^\phi}{\partial \phi_l} = \frac{dk}{\pi} \sum_{q=1,2} i l_q \left[ (\nabla q \tilde{T})^q \right]_{k+l}
\]
and
\[
\frac{\partial \tilde{T}_k^\phi}{\partial \phi_l} = -\frac{dk}{\pi} \sum_{q=1,2} i l_q \left[ (\nabla q \tilde{T})^q \right]_{k-l}.
\]

**Proof.** First notice
\[
\frac{\partial}{\partial \text{re } \phi_l} \frac{\partial \phi(x)}{\partial \text{re } \phi_l} = \int_{\mathbb{R}^2} i k_q e^{i\pi k} \frac{\partial \phi_k}{\partial \text{re } \phi_l} \frac{dk}{2\pi}
\]
\[
= \frac{dk}{\pi} \sum_{q=1,2} i l_q e^{i\pi l} - il_q e^{-i\pi l} \frac{dk}{2\pi}.
\]

This implies
\[
\frac{\partial \tilde{T}_k^\phi}{\partial \phi_l} = \frac{dk}{\pi} \int_{\mathbb{R}^2} e^{-i\pi l} \tilde{T}(x - \nabla \phi(x)) \frac{dx}{2\pi}
\]
\[
= \sum_{q=1,2} i l_q \int_{\mathbb{R}^2} e^{-i\pi l} \nabla q \tilde{T}(x - \nabla \phi(x)) \frac{dx}{2\pi},
\]
by (10) and (11),

Similarly
\[
\frac{\partial \tilde{T}_k^\phi}{\partial \phi_l} = \frac{dk}{\pi} \sum_{q=1,2} i l_q \left[ (\nabla q \tilde{T})^q \right]_{k-l}.
\]

**Lemma 2.** If \( A(x) \) and \( B(x) \) are real scalar fields then the two integrals, \( \int_{\mathbb{R}^2} \{A_{k-1} - A_{k+1}\} B^*_k dk \) and \( \int_{\mathbb{R}^2} \{A_{k-1} + A_{k+1}\} B^*_k dk \), are both real numbers.
Proof. By a simple change of variables it is clear that
\[
\int_{\mathbb{R}^3} \left( i \{ A_{k-l} - A_{k+l} \} B_k^* \right) dk = \int_{\mathbb{R}^3} i \{ A_{k' - l} - A_{k' + l} \} B_k'^* dk'
\]
and
\[
\int_{\mathbb{R}^3} \left( i \{ A_{k-l} + A_{k+l} \} B_k^* \right) dk = \int_{\mathbb{R}^3} i \{ A_{k' - l} + A_{k' + l} \} B_k'^* dk'.
\]

The following lemma is equivalent to the so-called Convolution Theorem. We state it here for reference.

**Lemma 3.** If \( A(x) \) and \( B(x) \) are real scalar fields then
\[
\int_{\mathbb{R}^3} A_{k-l} B_k^* \frac{dk}{2\pi} = \int_{\mathbb{R}^3} e^{-i\omega t} A(x) B(x) \frac{dx}{2\pi}.
\]

REFERENCES

Bezanson, J., Karpinski, S., Shah, V., et al. 2012, arXiv:1209.5145
Das, S., Sherwin, B., Aguirre, P., et al. 2011, PhRvL, 107, 021301
Dodelson, S. 2003, Modern Cosmology (New York: Academic)
Elsner, F., & Wandelt, B. 2010, ApJ, 724, 1262
Elsner, F., & Wandelt, B. D. 2013, A&A, 549, A111
Engelen, V., Keisler, A., Zahn, O., et al. 2012, ApJ, 756, 142
Eriksen, H. K., O' Dwyer, I. J., Jewell, J. B., et al. 2004, ApJS, 155, 227
Gelfand, A., Sahu, S., & Carlin, B. 1995, Biometrika, 82, 479
Hajian, A. 2007, PhRvD, 75, 083525
Hirata, C. M., & Seljak, U. c. v. 2003a, PhRvD, 67, 043001

Hirata, C. M., & Seljak, U. c. v. 2003b, PhRvD, 68, 083002
Hu, W. 2001, ApJL, 557, L79
Hu, W., & Okamoto, T. 2002, ApJ, 574, 566
Jasche, J., Kitaura, F. S., Li, C., & Enßlin, T. A. 2010, MNRAS, 409, 355
Jasche, J., & Lavaux, G. 2015, MNRAS, 447, 1204
Jasche, J., & Wandelt, B. 2012, MNRAS, 425, 1042
Jasche, J., & Wandelt, B. 2013a, MNRAS, 432, 894
Jasche, J., & Wandelt, B. 2013b, ApJ, 779, 15
Lewis, A. & Challinor, A. 2006, PhR, 429, 1
Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473
Neal, R. M. 2011, in Handbook of Markov Chain Monte Carlo, ed. S. Brooks et al. (Boca Raton, FL: Chapman & Hall/CRC), 113
Papaspiliopoulos, O., & Roberts, G. 2008, AnSta, 36, 95
Papaspiliopoulos, O., Roberts, G., & Skold, M. 2007, StaSe, 22, 59
Planck Collaboration 2014, A&A, 571, A17
Planck Collaboration 2015, arXiv:1502.01591
Roberts, G., Papaspiliopoulos, O., & Skold, M. 2003, in Bayesian Statistics 7: Proc. 7th Valencia Int. Meeting Organization, ed. J. M. Bernardo et al. (New York: Oxford Univ. Press), 307
Seljebotn, D. S., Mardal, K.-A., Jewell, J. B., Eriksen, H. K., & Bull, P. 2014, ApJS, 210, 24
Smith, K. M., Zahn, O., & Doré. O. 2007, PhRvD, 76, 043510
Taylor, J. F., Ashdown, M. A. J., & Hobson, M. P. 2008, MNRAS, 389, 1284
The Polarbear Collaboration, Ade, P. A. R., Akiba, Y., et al. 2014, ApJ, 794, 171
Wandelt, B. D., Larson, D. L., & Lakshminarayanan, A. 2004, PhRvD, 70, 083511
Yu, Y., & Meng, X.-L. 2011, J. Comput. Graphical Stat., 20, 531

The Astrophysical Journal, 808:152 (13pp), 2015 August 1
Anderes, Wandelt, & Lavaux

13