Analyzing weak lensing of the cosmic microwave background using the likelihood function

Christopher M. Hirata* and Uroš Seljak[†]

Physics Department, Princeton University, Princeton, New Jersey 08544
(Received 24 September 2002; published 12 February 2003)

Future experiments will produce high-resolution temperature maps of the cosmic microwave background (CMB) and are expected to reveal the signature of gravitational lensing by intervening large-scale structures. We construct all-sky maximum-likelihood estimators that use the lensing effect to estimate the projected density (convergence) of these structures, its power spectrum, and cross-correlation with other observables. This contrasts with earlier quadratic-estimator approaches that Taylor expanded the observed CMB temperature to linear order in the lensing deflection angle; these approaches gave estimators for the temperature-convergence correlation in terms of the CMB three-point correlation function and for the convergence power spectrum in terms of the CMB four-point correlation function, which can be biased and nonoptimal due to terms beyond the linear order. We show that for sufficiently weak lensing, the maximum-likelihood estimator reduces to the computationally less demanding quadratic estimator. The maximum likelihood and quadratic approaches are compared by evaluating the root-mean-square (rms) error and bias in the reconstructed convergence map in a numerical simulation; it is found that both the rms errors and bias are of order 1 percent for the case of Planck and of order 10–20 percent for a 1 arcminute beam experiment. We conclude that for recovering lensing information from temperature data acquired by these experiments, the quadratic estimator is close to optimal, but further work will be required to determine whether this is also the case for lensing of the CMB polarization field.

I. INTRODUCTION

Gravitational weak lensing of the cosmic microwave background (CMB) has been recognized as a potential indicator of large-scale structure in the universe. Compared to galaxy surveys, weak lensing has the advantage of directly tracing the matter density, thus avoiding the uncertainties associated with the relationship between the distributions of galaxies and of mass [1]. Because the CMB is the most distant background object that can be used for weak lensing studies, it probes the matter distribution at higher redshifts than can be reached by galaxy weak lensing and is sensitive to the largest observable scales in the universe [1–5].

In addition to providing data on the power spectrum of density fluctuations on these large scales, CMB weak lensing may yield constraints on the expansion history of the universe by making possible a measurement of the integrated Sachs-Wolfe (ISW) effect. The ISW effect (the change in temperature of the CMB radiation as it passed through a changing gravitational potential) is smaller than the primary CMB fluctuations produced in the early universe and consequently can be detected only through the cross-correlation of CMB observations with some tracer of the gravitational potential. Because it is sensitive directly to the potential, weak lensing is an ideal candidate for this cross-correlation [6,7].

Because detection of CMB weak lensing may be possible with near-future satellite experiments, such as Planck and possibly even the Microwave Anisotropy Probe (MAP), several algorithms have been proposed for estimating matter distributions, power spectra, and ISW cross-correlations

from CMB temperature maps. Some of these methods are based on local statistics, such as the products of gradients of the temperature field [1]. Recently Hu [2,3], working to linear order in the deflection angle, determined the optimal quadratic estimator (i.e. quadratic in the CMB temperature map) for the deflection field. Within this linear approximation, the corresponding power spectrum estimator makes full use of the information in the CMB four-point correlation function [8,9]. However, the limits to the validity of the linear order approximation have not been well determined, and the possibility of obtaining more information on lensing from higher-order correlation functions has not been studied in detail. Neglect of nonlinear terms may also create a bias in the quadratic estimators of the power spectrum. The nonlinear terms may be important whenever the deflection angle is comparable to the scale of CMB fluctuation used in the reconstruction of lensing potential. The deflection angle is of the order of several arcminutes and for high resolution experiments significant amount of lensing information comes from CMB modes on the same scale, indicating that the nonlinear terms may be important. In order to address these issues, we use the likelihood function to construct estimators rather than assuming an estimator with a particular form (local, quadratic, etc.) and avoid linearizing in the deflection field except to compare our results to previous work and where necessary for computational tractability.

We work principally in position space rather than harmonic space. This is done partly because real data are obtained in position space, and partly to show how the harmonic-space estimators [2] can be derived from position-space arguments; also, the generalization of the position-space analysis to anisotropic instrument noise is more transparent. We also do not consider the reconstruction of matter distributions from CMB polarization; although polarization

---

*Electronic address: chirata@princeton.edu
[†]Electronic address: useljak@princeton.edu

can theoretically yield much better information about lensing than CMB temperature fluctuations [10], it is also computationally more demanding, so we defer a more careful analysis to a future work.

We will proceed as follows: Section II introduces our formalism and notation, and defines the basic mathematical operations that will be used in the rest of the paper. Section III considers the likelihood function for the CMB and its dependence on the lensing potential (the potential that generates the deflection field). In Sec. IV we consider the maximum likelihood estimators for the power spectrum of the lensing potential and its cross-correlation with the CMB. In Sec. V, we describe our numerical implementation of the estimators from Secs. III and IV; the performance of the estimators, as determined numerically, is described in Sec. V E. We conclude in Sec. VI.

II. FORMALISM

A. CMB

The cosmic microwave background temperature fluctuation $\tilde{\Theta}$ in a particular direction $\mathbf{n}$ on the unit sphere is defined by $\tilde{\Theta}(\mathbf{n}) = T(\mathbf{n})/T_0 - 1$ where $T(\mathbf{n})$ is the CMB temperature in direction $\mathbf{n}$ and $T_0 = 2.72$ K is the mean temperature of the CMB. This temperature fluctuation can be expressed in harmonic space as

$$\tilde{\Theta}(\mathbf{n}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \tilde{\Theta}_{lm} Y_{lm}(\mathbf{n}), \qquad (1)$$

where the Y_{lm} are spherical harmonics and $\tilde{\Theta}_{lm}$ are the corresponding coefficients. The spherical harmonics are orthogonal and are normalized so that their squared amplitude integrates to one over the sphere: $\int_{S^2} |Y_{lm}|^2 d\Omega = 1$, and the transformation of Eq. (1) can thus be inverted as

$$\tilde{\Theta}_{lm} = \int_{S^2} d^2 \mathbf{n} Y_{lm}^*(\mathbf{n}) \tilde{\Theta}(\mathbf{n}). \qquad (2)$$

Because the statistical average $\langle \tilde{\Theta}_{lm} \rangle = 0$, we extensively use the power spectrum. The power spectrum is defined for a statistically isotropic temperature fluctuation as the variance

$$\langle \tilde{\Theta}_{l'm'}^*, \tilde{\Theta}_{lm} \rangle = C_l^{\tilde{\Theta}\tilde{\Theta}} \delta_{ll'} \delta_{mm'}. \qquad (3)$$

For gravitational lensing work, we distinguish three temperature fluctuations: the unlensed temperature fluctuation Θ; the lensed temperature fluctuation $\tilde{\Theta}$; and the measured temperature fluctuation $\hat{\Theta}$. Throughout this paper, we will take the primary (unlensed) anisotropy Θ to be a Gaussian random field. The measurement is related to the actual temperature fluctuation by the instrument noise, ϵ:

$$\hat{\Theta}(\mathbf{n}) = \tilde{\Theta}(\mathbf{n}) + \epsilon(\mathbf{n}). \qquad (4)$$

We assume that the instrument noise ϵ is independent of $\tilde{\Theta}$.

Occasionally we will use the flat-sky approximation, in which a map Θ can be expanded in Fourier modes, $\tilde{\Theta}(\mathbf{n}) = (1/\sqrt{4\pi}) \Sigma_{\mathbf{l}} \tilde{\Theta}_{\mathbf{l}} e^{-i\mathbf{l}\cdot\mathbf{n}}$. The Fourier modes are normalized over an area of 4π, and populate the $\mathbf{l}$ plane with a two-dimensional density of $1/\pi$; this ensures that the flat-sky and all-sky normalizations are consistent on small scales.

B. Lensing

Gravitational lensing of the CMB by scalar perturbations can be expressed in terms of the lensing potential Φ, defined by

$$\tilde{\Theta}(\mathbf{n}) = \Theta[\mathbf{n} + \nabla\Phi(\mathbf{n})], \qquad (5)$$

where ∇ is the two-dimensional gradient operator on the unit sphere. The lensing potential Φ is the projected gravitational potential along the line of sight (see the Appendix for details),

$$\Phi(\mathbf{n}) = -2 \int_0^{r_{ls}} dr \Psi(r\mathbf{n}, -r) \left(\frac{1}{T(r)} - \frac{1}{T(r_{ls})} \right), \qquad (6)$$

where r_{ls} is the comoving distance to the last-scatter surface, $\Psi(\mathbf{x}, \tau)$ is the gravitational scalar potential at comoving position $\mathbf{x}$ and conformal time τ, and $T(r)$ is the tangentlike function ($\tan r$, r, or $\tanh r$ depending on whether the universe is closed, spatially flat, or open). The convergence $\kappa = -\frac{1}{2}\nabla^2\Phi$ is positive when structures along the line of sight act as a converging lens (i.e. when they magnify the CMB) and is negative for a diverging lens. Conceptually, we would thus expect κ to be a measure of the projected density perturbation; as shown in the Appendix, this is indeed the case. We define the power spectra $C_l^{\Phi\Phi}$ and $C_l^{\kappa\kappa} = l^2(l+1)^2 C_l^{\Phi\Phi}/4$, and the cross-correlation $C_l^{\Theta\Phi}$, in analogy to Eq. (3).

We will in several instances require use of the lensing operator Λ that performs the operation in Eq. (5):

$$\Lambda[\Phi]\Theta(\mathbf{n}) = \Theta[\mathbf{n} + \nabla\Phi(\mathbf{n})]. \qquad (7)$$

On occasion, we shall refer to the linear approximation to the lensing operator:

$$\tilde{\Theta} = \Lambda\Theta \approx \Theta + \nabla\Theta \cdot \nabla\Phi. \qquad (8)$$

Note that we have used $\tilde{\Theta}$ to represent the lensed CMB temperature and Θ to represent the unlensed temperature; some authors have used this convention [3], while others [1,2,8,10] have used Θ for the lensed and $\tilde{\Theta}$ for the unlensed temperature.

C. Convolutions and integrals

A convolution of a function Θ on the unit sphere with kernel C is written as

$$C\Theta(\mathbf{n}) = \int_{\Omega} d^2 \mathbf{n}' C(\mathbf{n}, \mathbf{n}') \Theta(\mathbf{n}'), \qquad (9)$$

where Ω is the region in which we have data, and the kernel C can be decomposed in multipoles using the Legendre polynomials:

$$C(\mathbf{n}, \mathbf{n}') = \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} C_l P_l(\mathbf{n} \cdot \mathbf{n}'). \qquad (10)$$

We will also need to take the inverse operation C^{-1} such that $CC^{-1}\Theta = \Theta$. In the case of a true full-sky experiment (i.e. one that acquires usable data over the full 4π steradians), the C^{-1} operation is trivial: we apply a convolution with a C^{-1} kernel with multipoles $(C^{-1})_l = (C_l)^{-1}$. The inversion is more difficult on a portion of the sphere, as discussed in Sec. V.

Finally, we make use of the notation derived from linear algebra: our ''column vectors'' are functions on Ω, and our ''matrices'' are linear operators on this set of functions: $Av(\mathbf{x}) = \int_{\Omega} A(\mathbf{x}, \mathbf{y}) v(\mathbf{y}) d^2\mathbf{y}$. Example uses of this notation are $u^T v = \int_{\Omega} uv\, d^2\mathbf{n}$ and $A^T(\mathbf{x}, \mathbf{y}) = A(\mathbf{y}, \mathbf{x})$.

III. LIKELIHOOD ANALYSIS

We analyze the likelihood function for gravitational lensing because this function retains all of the information provided by the observations. In particular, we can compare the ''optimal'' maximum likelihood estimators (MLEs) to previous results. We examine the relationship between the quadratic estimators and the likelihood-based estimators and the criteria for their equivalence, i.e. for optimality of the quadratic estimator. In this section we are attempting to estimate Φ based on the measured temperatures $\hat{\Theta}$, and so we will treat $\hat{\Theta}$ as a random variable whose distribution depends on the fixed parameters $\Phi(\mathbf{x})$. We will allow Φ to be a random field when we consider its statistical properties (e.g. the power spectrum $C^{\Phi\Phi}$) in Sec. IV.

We will see that the lensing potential Φ is detectable because its presence breaks spherical symmetry and thus causes correlations between the different spherical harmonic modes of the temperature field $\tilde{\Theta}$, i.e. it creates off-diagonal elements in the covariance $C^{\tilde{\Theta}\tilde{\Theta}} = \langle \tilde{\Theta}\tilde{\Theta}^T \rangle$ when expressed in the spherical harmonic basis; this is manifested in real space by an anisotropic correlation function $C^{\tilde{\Theta}\tilde{\Theta}}(\mathbf{x}, \mathbf{y})$. Since these off-diagonal elements are, in the linear approximation, proportional to the lensing potential Φ, we could take $\tilde{\Theta}\tilde{\Theta}^T$ as a crude estimate of the covariance $C^{\tilde{\Theta}\tilde{\Theta}}$ and form linear combinations of the off-diagonal elements to construct an estimator for Φ; this is the essence of the quadratic estimator methods [2]. (In the presence of instrument noise we measure $\hat{\Theta}$ and not $\tilde{\Theta}$ but the idea is the same.) Note that while some quadratic estimators (e.g. Ref. [1]) have been derived from considering the magnification and shear of small-scale CMB features by larger-scale lensing modes, in analogy to the weak lensing of galaxies, such a picture is not essential to the quadratic estimation framework—quadratic estimation is possible whenever the linear approximation to $C^{\tilde{\Theta}\tilde{\Theta}}$ is valid. The likelihood method, while somewhat more involved, is

useful to investigate for two reasons: first, unlike quadratic estimators, MLE's are guaranteed to be asymptotically efficient (i.e. it is impossible to achieve lower error than the MLE in the limit of an infinite amount of data); and second, the likelihood approach retains its validity even when higher-order [e.g. $O(\Phi^2)$] terms in the covariance are important.

This section will be organized as follows. In Sec. III A, we introduce the likelihood function and its basic properties and give a formal expression for it. We maximize the likelihood function using the calculus of variations (Sec. III B) and proceed to show that within the linear approximation [Eq. (8)] the maximum likelihood estimator reduces to the optimally weighted quadratic estimator (Sec. III C). We examine our ability to reconstruct the primary CMB anisotropy Θ in Sec. III D. We conclude in Secs. III E and III F by examining the limits of validity of the linear approximation.

A. Likelihood function

Likelihood maximization is a generally applicable method to statistical estimation problems. A statistical estimation problem involves a data set, in this case the measured CMB temperature fluctuation $\hat{\Theta}(\mathbf{n}_i)$ at N points $\{\mathbf{n}_1, \ldots, \mathbf{n}_N\}$, which has a probability distribution determined by a set of parameters, in this case the values of the lensing potential Φ. The problem is to estimate the unknown parameters Φ from the observations $\hat{\Theta}$. We represent the probability distribution by a density function P, which is related to the differential probability $d\Pi$ for obtaining temperature measurements between $\hat{\Theta}(\mathbf{n}_i)$ and $\hat{\Theta}(\mathbf{n}_i) + d\hat{\Theta}(\mathbf{n}_i)$:

$$d\Pi = P(\hat{\Theta}|\Phi) d\hat{\Theta}(\mathbf{n}_1) \cdots d\hat{\Theta}(\mathbf{n}_N). \qquad (11)$$

The maximum likelihood estimation method simply selects the value of Φ that yields the largest value of P, i.e. the value of Φ that would have been most likely to generate the observed $\hat{\Theta}$. While this method is very general and can be applied to a wide range of problems, maximum likelihood estimators (MLE's) are frequently very difficult to compute, as is the case here.

For convenience, we will work not with the likelihood function but with its negative logarithm $\mathcal{L}$, which is defined by the relation

$$\mathcal{L}[\Phi] = -\ln P(\hat{\Theta}|\Phi). \qquad (12)$$

If we assume Gaussian instrument noise of covariance $C^{\epsilon\epsilon}$, we find that for fixed Φ, $\hat{\Theta}$ is a Gaussian random field with covariance

$$C^{\hat{\Theta}\hat{\Theta}}[\Phi] = \Lambda[\Phi] C^{\Theta\Theta} \Lambda[\Phi]^T + C^{\epsilon\epsilon}, \qquad (13)$$

where the transpose Λ^T of the linear operator Λ is defined by $\Lambda^T(\mathbf{x}, \mathbf{y}) = \Lambda(\mathbf{y}, \mathbf{x})$. The probability density of $\hat{\Theta}$ is then related to its covariance via the usual relation for a Gaussian:

$$P(\hat{\Theta}|\Phi) = \frac{1}{(2\pi)^{N/2} \sqrt{\det C^{\hat{\Theta}\hat{\Theta}}}} \exp\left(-\frac{1}{2} \hat{\Theta}^T C^{\hat{\Theta}\hat{\Theta}^{-1}} \hat{\Theta} \right).$$

$$(14)$$

Combining this with the definition in Eq. (12) and the standard Gaussian probability density formula, we find that

$$\mathcal{L}[\Phi] = \frac{1}{2} \hat{\Theta}^T (C^{\hat{\Theta}\hat{\Theta}}[\Phi])^{-1} \hat{\Theta} + \frac{1}{2} \ln \det C^{\hat{\Theta}\hat{\Theta}}[\Phi].$$

(15)

In some cases, we will use a Gaussian random prior for Φ, in which case we will use the negative log posterior probability $\mathcal{P}$ in place of the negative log likelihood $\mathcal{L}$. The Gaussian prior for Φ is

$$P(\Phi|C^{\Phi\Phi}) = \frac{1}{(2\pi)^{N/2}\sqrt{\det C^{\Phi\Phi}}} \exp\left(-\frac{1}{2}\Phi^T C^{\Phi\Phi-1}\Phi\right),$$

(16)

where N is the number of pixels in the map. From this the negative log posterior probability can be determined (up to an irrelevant constant) to be

$$\mathcal{P}[\Phi;C^{\Phi\Phi}] = \mathcal{L}[\Phi] - \ln P(\Phi|C^{\Phi\Phi}) = \frac{1}{2}\hat{\Theta}^T(C^{\hat{\Theta}\hat{\Theta}}[\Phi])^{-1}\hat{\Theta}$$

$$+ \frac{1}{2}\ln\det C^{\hat{\Theta}\hat{\Theta}}[\Phi] + \frac{1}{2}\Phi^T(C^{\Phi\Phi})^{-1}\Phi$$

$$+ \frac{1}{2}\ln\det C^{\Phi\Phi},$$

(17)

where $C^{\Phi\Phi}$ is the covariance of the prior for Φ.

B. Likelihood-based estimators

We construct estimators for the lensing potential Φ using $\mathcal{L}$ and $\mathcal{P}$ by setting their functional derivatives with respect to $\Phi(\mathbf{n})$ equal to zero. Differentiating Eq. (15) gives

$$\frac{\delta\mathcal{L}[\Phi]}{\delta\Phi} = -\frac{1}{2}\hat{\Theta}^T(C^{\hat{\Theta}\hat{\Theta}}[\Phi])^{-1}\frac{\delta C^{\hat{\Theta}\hat{\Theta}}[\Phi]}{\delta\Phi}(C^{\hat{\Theta}\hat{\Theta}}[\Phi])^{-1}\hat{\Theta}$$

$$+ \frac{1}{2}\mathrm{Tr}\left[(C^{\hat{\Theta}\hat{\Theta}}[\Phi])^{-1}\frac{\delta C^{\hat{\Theta}\hat{\Theta}}[\Phi]}{\delta\Phi}\right].$$

(18)

Using Eq. (13), we calculate the functional derivative of $C^{\hat{\Theta}\hat{\Theta}}[\Phi]$:

$$\frac{\delta C^{\hat{\Theta}\hat{\Theta}}[\Phi](\mathbf{y},\mathbf{z})}{\delta\Phi(\mathbf{x})} = \int_\Omega (\Lambda[\Phi]C^{\Theta\Theta})(\mathbf{y},\mathbf{y}')\frac{\delta\Lambda[\Phi](\mathbf{z},\mathbf{y}')}{\delta\Phi(\mathbf{x})}d^2\mathbf{y}'$$

$$+ \text{transpose}.$$

(19)

We differentiate Λ using Eq. (7):

$$\frac{\delta}{\delta\Phi(\mathbf{x})}(\Lambda[\Phi]v)(\mathbf{w})$$

$$= \int_\Omega d^2\mathbf{x}'\frac{\delta\nabla\Phi(\mathbf{x}')}{\delta\Phi(\mathbf{x})}\cdot\frac{\delta}{\delta\nabla\Phi(\mathbf{x}')}(\Lambda[\Phi]v)(\mathbf{w})$$

$$= [\nabla_\mathbf{w}\delta^{(2)}(\mathbf{w}-\mathbf{x})]\cdot(\Lambda[\Phi]\nabla v)(\mathbf{w}).$$

(20)

Using this relation and integration by parts, we convert Eq. (19) into

$$\frac{\delta C^{\hat{\Theta}\hat{\Theta}}[\Phi](\mathbf{y},\mathbf{z})}{\delta\Phi(\mathbf{x})} = \int_\Omega d^2\mathbf{y}'(\Lambda[\Phi]C^{\Theta\Theta})(\mathbf{y},\mathbf{y}')$$

$$\times[\nabla_\mathbf{z}\delta^{(2)}(\mathbf{z}-\mathbf{x})]\cdot(\Lambda[\Phi]\nabla)(\mathbf{z},\mathbf{y}')$$

$$+ \text{transpose}$$

$$= [\nabla_\mathbf{z}\delta^{(2)}(\mathbf{z}-\mathbf{x})]\cdot(\Lambda[\Phi]\nabla C^{\Theta\Theta}\Lambda[\Phi]^T)$$

$$\times(\mathbf{z},\mathbf{y}) + \text{transpose}.$$

(21)

We also express the trace as an expectation value using the identity $\mathrm{Tr}(X) = \langle uXC^{-1}u\rangle$ with u drawn from a Gaussian distribution of covariance C, and integrate by parts again to yield

$$\frac{\delta\mathcal{L}[\Phi]}{\delta\Phi} = \nabla\cdot[\hat{\Theta}(C^{\hat{\Theta}\hat{\Theta}}[\Phi])^{-1}\Lambda[\Phi]\nabla C^{\Theta\Theta}$$

$$\times\Lambda[\Phi]^{-1}(C^{\hat{\Theta}\hat{\Theta}}[\Phi])^{-1}\hat{\Theta}]$$

$$- \langle\nabla\cdot[\hat{\Theta}(C^{\hat{\Theta}\hat{\Theta}}[\Phi])^{-1}\Lambda[\Phi]\nabla C^{\Theta\Theta}$$

$$\times\Lambda[\Phi]^{-1}(C^{\hat{\Theta}\hat{\Theta}}[\Phi])^{-1}\hat{\Theta}]\rangle.$$

(22)

The functional derivative of $\mathcal{P}$ differs by the addition of a $C^{\Phi\Phi-1}\Phi$ term. The maximum-likelihood estimator $\hat{\Phi}$ for Φ is then the solution to $\delta\mathcal{L}/\delta\Phi = 0$. If we then define the likelihood gradient $G[\Phi]$ by

$$G[\Phi] \equiv \frac{\delta\mathcal{L}}{\delta\Phi} = \nabla\cdot[(C^{\hat{\Theta}\hat{\Theta}}[\hat{\Phi}]^{-1}\hat{\Theta})\Lambda[\hat{\Phi}]\nabla C^{\Theta\Theta}$$

$$\times\Lambda[\hat{\Phi}]^{-1}(C^{\hat{\Theta}\hat{\Theta}}[\hat{\Phi}])^{-1}\hat{\Theta}$$

$$- \langle(C^{\hat{\Theta}\hat{\Theta}}[\hat{\Phi}])^{-1}\hat{\Theta})\Lambda[\hat{\Phi}]\nabla C^{\Theta\Theta}$$

$$\times\Lambda[\hat{\Phi}]^{-1}(C^{\hat{\Theta}\hat{\Theta}}[\hat{\Phi}])^{-1}\hat{\Theta}\rangle],$$

(23)

then the maximum likelihood estimator becomes

$$G[\hat{\Phi}] = 0,$$

(24)

whereas the mode of the posterior probability distribution (i.e. maximum of $e^{-\mathcal{P}}$) is the solution $\hat{\Phi}$ to

$$\hat{\Phi} = -C^{\Phi\Phi}G[\hat{\Phi}].$$

(25)

In deriving Eq. (23), we have dropped boundary terms. In our implementation (Sec. V) we simply do not work near the survey boundaries, however, the formalism can be generalized to include these by setting $C^{\epsilon\epsilon} = \infty$ in the unscanned

regions. This does not cause numerical difficulties because the infinite eigenvalues of $C^{\epsilon\epsilon}$ (and hence $C^{\hat{\Theta}\hat{\Theta}}$) become null eigenvalues of $C^{\hat{\Theta}\hat{\Theta}-1}$ [11].

As a final note, the expectation value in Eq. (23), which derives ultimately from the determinant in the Gaussian probability density, is small when the noise is small ($C^{\epsilon\epsilon} \ll C^{\Theta\Theta}$) and we are far from the boundaries of the region of sky surveyed. This is because substituting the zero-noise limit for $\hat{\Theta}$ and $C^{\hat{\Theta}\hat{\Theta}}$ into the expectation value converts it into

$$\Lambda[\hat{\Phi}]\langle\{[(C^{\Theta\Theta})^{-1}\Theta]\nabla\Theta\}\rangle. \tag{26}$$

We next note that for a statistically isotropic unlensed Θ and an all-sky survey, the expectation value in Eq. (26) must vanish because it is a two-vector (i.e. a vector on S^2) and hence a nonzero value would pick out a preferred direction. Near a boundary of the surveyed region, this argument fails because the boundary breaks rotational symmetry. The expectation value in Eq. (23) thus acquires a nonzero value only in the presence of instrument noise and boundary effects. Conceptually, we understand this as a property of Eq. (13): noise adds the $C^{\epsilon\epsilon}$ term to $C^{\hat{\Theta}\hat{\Theta}}$, while boundary effects alter the unit determinant of Λ. Without these effects, $\det C^{\hat{\Theta}\hat{\Theta}} = \det C^{\Theta\Theta} = \text{const}$, and the expectation value in Eq. (23), which is merely a derivative of the log determinant, vanishes. We further note that modes with large noise ($C^{\epsilon\epsilon} \gg C^{\Theta\Theta}$) do not contribute to G because of the $C^{\hat{\Theta}\hat{\Theta}-1}$ which appears twice in Eq. (23). Since most CMB experiments have only a small range of l for which $C^{\epsilon\epsilon}$ and $C^{\Theta\Theta}$ are of the same order, and it is only in this regime and near boundaries that the expectation value in Eq. (23) is important, we will neglect the expectation value in the remainder of this paper. That is, we approximate

$$G[\Phi] \approx \nabla \cdot [(C^{\hat{\Theta}\hat{\Theta}}[\hat{\Phi}]^{-1}\hat{\Theta})\Lambda[\hat{\Phi}]\nabla C^{\Theta\Theta}$$
$$\times \Lambda[\hat{\Phi}]^{-1}(C^{\hat{\Theta}\hat{\Theta}}[\hat{\Phi}])^{-1}\hat{\Theta}]. \tag{27}$$

C. Linearized version of MLE

In order to connect Eq. (24) to previous work on quadratic estimators, we approximate the right hand side of the equation to linear order in Φ: $G \approx G_0 + F\Phi$, where the likelihood gradient G is computed from Eq. (27), F is the matrix of second derivatives $\delta^2\mathcal{L}/\delta\Phi\,\delta\Phi$ (independent of Φ in the Gaussian approximation), and G_0 is equal to G evaluated with no lensing:

$$G_0 \approx \nabla \cdot [\hat{\Theta}(C^{\Theta\Theta} + C^{\epsilon\epsilon})^{-1}\nabla C^{\Theta\Theta}(C^{\Theta\Theta} + C^{\epsilon\epsilon})^{-1}\hat{\Theta}]. \tag{28}$$

In order to obtain a quadratic (rather than merely a rational) estimator, the approximate curvature matrix F must be taken independent of $\hat{\Theta}$. We will therefore replace it by its expectation value $\langle F \rangle$ averaged over $\hat{\Theta}$. This expectation value is

$$F_{mn} \approx \langle F_{mn} \rangle|_\Phi = \int \mathcal{D}\hat{\Theta}\, P(\hat{\Theta}|\Phi)F_{mn}$$
$$= \int \mathcal{D}\hat{\Theta}\, e^{-\mathcal{L}(\Phi|\hat{\Theta})}\frac{\delta^2\mathcal{L}(\Phi|\hat{\Theta})}{\delta\Phi_m\delta\Phi_n}, \tag{29}$$

where m and n are the matrix indices of the matrix of second derivatives. We may compute the last integral by requiring that the probability distribution for $\hat{\Theta}$ be properly normalized:

$$1 = \int \mathcal{D}\hat{\Theta}\, P(\hat{\Theta}|\Phi) = \int \mathcal{D}\hat{\Theta}\, e^{-\mathcal{L}(\hat{\Theta}|\Phi)}. \tag{30}$$

Taking the second derivative of this equation with respect to Φ gives

$$0 = -\int \mathcal{D}\hat{\Theta}\, e^{-\mathcal{L}(\hat{\Theta}|\Phi)}\frac{\delta^2\mathcal{L}(\Phi|\hat{\Theta})}{\delta\Phi_m\delta\Phi_n}$$
$$+ \int \mathcal{D}\hat{\Theta}\, e^{-\mathcal{L}(\hat{\Theta}|\Phi)}\frac{\delta\mathcal{L}(\Phi|\hat{\Theta})}{\delta\Phi_m}\frac{\delta\mathcal{L}(\Phi|\hat{\Theta})}{\delta\Phi_n}, \tag{31}$$

which enables us to rewrite Eq. (29) as

$$F_{mn} \approx \int \mathcal{D}\hat{\Theta}\, e^{-\mathcal{L}(\hat{\Theta}|\Phi)}\frac{\delta\mathcal{L}(\Phi|\hat{\Theta})}{\delta\Phi_m}\frac{\delta\mathcal{L}(\Phi|\hat{\Theta})}{\delta\Phi_n} = \langle G_m G_n \rangle. \tag{32}$$

Thus the matrix of second derivatives is simply the covariance of the likelihood gradient. (Note: we will call the matrix of second derivatives the Fisher matrix even though the technical definition of the Fisher matrix differs from F for a non-Gaussian likelihood function.) We choose to evaluate F at $\Phi = 0$ (no lensing) for convenience, although within the Gaussian approximation F can be evaluated anywhere. At $\Phi = 0$, Eq. (27) for G simplifies dramatically and we have

$$F \approx \langle GG^T \rangle|_{\Phi=0} \approx \langle \nabla \cdot [(C^{\hat{\Theta}\hat{\Theta}-1}\hat{\Theta})\nabla C^{\Theta\Theta}C^{\hat{\Theta}\hat{\Theta}-1}\hat{\Theta}]$$
$$\times \{\nabla \cdot [(C^{\hat{\Theta}\hat{\Theta}-1}\hat{\Theta})\nabla C^{\Theta\Theta}C^{\hat{\Theta}\hat{\Theta}-1}\hat{\Theta}]\}^T \rangle, \tag{33}$$

which is recognizable as a four-point correlation function of $\hat{\Theta}$. If we switch to the flat sky approximation, and assume the noise is isotropic, $C^{\Theta\Theta}$ and $C^{\epsilon\epsilon}$ become diagonal in Fourier space. Then we can compute the four-point correlation function for Gaussian $\hat{\Theta}$ using Wick's theorem:

$$\langle \hat{\Theta}_{l_1}\hat{\Theta}_{l_2}\hat{\Theta}_{l_3}\hat{\Theta}_{l_4} \rangle = C^{\hat{\Theta}\hat{\Theta}}_{l_1}C^{\hat{\Theta}\hat{\Theta}}_{l_2}\delta_{l_1+l_3,0}\delta_{l_2+l_4,0}$$
$$+ C^{\hat{\Theta}\hat{\Theta}}_{l_1}C^{\hat{\Theta}\hat{\Theta}}_{l_2}\delta_{l_1+l_4,0}\delta_{l_2+l_3,0}$$
$$+ C^{\hat{\Theta}\hat{\Theta}}_{l_1}C^{\hat{\Theta}\hat{\Theta}}_{l_3}\delta_{l_2+l_3,0}\delta_{l_1+l_4,0}. \tag{34}$$

This gives the result for F:

$$F_{\mathbf{L}} = \frac{1}{8\pi^2} \int_{\mathbf{l}_1 + \mathbf{l}_2 = \mathbf{L}} d^2 \mathbf{l}_1 \frac{[\mathbf{L} \cdot (\mathbf{l}_1 C_{l_1}^{\Theta\Theta} + \mathbf{l}_2 C_{l_2}^{\Theta\Theta})]^2}{(C_{l_1}^{\Theta\Theta} + C_{l_1}^{\epsilon\epsilon})(C_{l_2}^{\Theta\Theta} + C_{l_2}^{\epsilon\epsilon})}, \quad (35)$$

Equation (35) is recognizable (apart from a factor of $4\pi/L^2$ due to normalization convention) as the noise variance and optimal weighting derived by Hu [2]. We use the flat-sky approximation only to compute the noise curves in Fig. 3; in our simulations we will evaluate F via a Monte Carlo technique (see Sec. V B).

We can then construct the maximum likelihood estimator for Φ under this approximation,

$$\hat{\Phi}_{\mathrm{MLE}} = -F^{-1} G_0 = -F^{-1} \nabla \cdot [\hat{\Theta} (C^{\Theta\Theta} + C^{\epsilon\epsilon})^{-1}$$
$$\times \nabla C^{\Theta\Theta} (C^{\Theta\Theta} + C^{\epsilon\epsilon})^{-1} \hat{\Theta}], \quad (36)$$

and the corresponding approximate mode of the posterior probability density:

$$\hat{\Phi} = -[(C^{\Phi\Phi})^{-1} + F]^{-1} G_0$$
$$= -[(C^{\Phi\Phi})^{-1} + F]^{-1}$$
$$\times \nabla \cdot [\hat{\Theta}(C^{\Theta\Theta} + C^{\epsilon\epsilon})^{-1} \nabla C^{\Theta\Theta} (C^{\Theta\Theta} + C^{\epsilon\epsilon})^{-1} \hat{\Theta}]. \quad (37)$$

Both of these are recognizable as quadratic estimators, i.e. they are second-order polynomials in $\hat{\Theta}$. By spherical symmetry, if $\hat{\Theta}$ is statistically isotropic then the vector quantity in brackets, and hence $\hat{\Phi}$, will have expectation value zero. Thus Eqs. (36) and (37) are measuring the deviation of $\hat{\Theta}$ from statistical isotropy that arises from lensing by a potential Φ. These deviations from statistical isotropy in position-space appear as correlations between different spherical harmonic modes in harmonic space; see Ref. [2] for the associated harmonic-space estimator. It can be shown [2] that within the linear approximation, Eq. (36) provides an unbiased estimate for the lensing potential Φ when averaged over an ensemble of primary CMB anisotropies Θ.

Having determined these approximations, we consider the conditions of their validity. The linearization of the right hand side of Eq. (20) clearly corresponds to a Gaussian approximation to the likelihood function, with the second-order Taylor expansion of $\mathcal{L}$ carried out around $\Phi = 0$. This can be expected to be valid when the maximum likelihood point is "near" $\Phi = 0$ in the sense that $\Phi \ll \mathcal{L}''/\mathcal{L}'''$ (where the ′ denotes a functional derivative with respect to Φ). Therefore it would be reasonable to expect that the estimators in Eqs. (36) and (37) break down when the lensing effects become large, i.e. when $C^{\Phi\Phi}$ becomes sufficiently large. We analyze this possibility analytically in Secs. III E and III F and numerically in Sec. V E.

D. Reconstructing the primary CMB

We next wish to reconstruct the primary (unlensed) CMB Θ from observations $\hat{\Theta}$ of the lensed temperature field. Be-

cause our instrument gives us one function on the sky, $\hat{\Theta}$, it is not in general possible to simultaneously reconstruct the primary CMB anisotropy, Θ, and the lensing potential Φ. If the CMB and lensing power spectra are given, however, we can use the power spectra as a prior and construct a Bayesian posterior probability distribution for Θ and Φ and maximize it. While this does not permit determination of the primary anisotropy to arbitrary accuracy, it is the best that one can hope for if only the lensed CMB temperature is available. The determination of the lensing potential and primary CMB power spectra is discussed in Secs. IV B and IV D.

To estimate the primary CMB anisotropy Θ, we take a Gaussian prior for both the primary CMB and the lensing potential. This gives us a joint posterior probability distribution for Θ and Φ of $e^{-\mathcal{R}}$, where $\mathcal{R}$ is given (up to an additive constant) by

$$\mathcal{R}[\Theta,\Phi] = \frac{1}{2}\Theta^T (C^{\Theta\Theta})^{-1}\Theta + \frac{1}{2}\Phi^T (C^{\Phi\Phi})^{-1}\Phi$$
$$- \ln P(\hat{\Theta}|\Theta,\Phi), \quad (38)$$

where $P(\hat{\Theta}|\Theta,\Phi)$ is the conditional probability of observing temperature $\hat{\Theta}$ given a primary CMB temperature Θ and lensing potential Φ. It is readily noted that P is simply the instrument noise curve, which we take to be Gaussian:

$$-\ln P(\hat{\Theta}|\Theta,\Phi) = \frac{1}{2}(\hat{\Theta} - \Lambda[\Phi]\Theta)^T (C^{\epsilon\epsilon})^{-1}(\hat{\Theta} - \Lambda[\Phi]\Theta). \quad (39)$$

Equations (38) and (39) formally express the joint posterior probability distribution for Θ and Φ. In order to reconstruct the primary CMB, we integrate out the lensing potential to find the negative log posterior probability distribution $\bar{\mathcal{R}}$ for Θ:

$$e^{-\bar{\mathcal{R}}[\Theta]} = \int \mathcal{D}\Phi\, e^{-\mathcal{R}[\Theta,\Phi]}$$
$$= \int \mathcal{D}\Phi \exp\left[-\frac{1}{2}(\hat{\Theta} - \Lambda[\Phi]\Theta)^T \right.$$
$$\times C^{\epsilon\epsilon-1}(\hat{\Theta} - \Lambda[\Phi]\Theta)$$
$$\left. -\frac{1}{2}\Theta^T C^{\Theta\Theta-1}\Theta - \frac{1}{2}\Phi^T C^{\Phi\Phi-1}\Phi \right]. \quad (40)$$

This equation is difficult to evaluate. In the linear approximation, however, we may replace $\Lambda[\Phi]\Theta$ with $\Theta + \nabla\Phi\cdot\nabla\Theta$; this makes the integral Gaussian, so it can be evaluated analytically to give

$$\bar{\mathcal{R}}[\Theta] = -\frac{1}{2}\mathcal{G}[\Theta]^T (C^{\Phi\Phi-1} + \mathcal{F}[\Theta])^{-1}\mathcal{G}[\Theta]$$
$$+ \frac{1}{2}\Theta^T C^{\Theta\Theta-1}\Theta + \frac{1}{2}(\hat{\Theta} - \Theta)^T C^{\epsilon\epsilon-1}(\hat{\Theta} - \Theta)$$
$$+ \frac{1}{2}\ln\det(C^{\Phi\Phi-1} + \mathcal{F}[\Theta]), \quad (41)$$

where

$$\mathcal{G}[\Theta](\mathbf{x}) = \nabla \cdot [[C^{\epsilon\epsilon -1}(\hat{\Theta} - \Theta)]\nabla\Theta](\mathbf{x}) \qquad (42)$$

and

$$\mathcal{F}[\Theta](\mathbf{x},\mathbf{y}) = (\nabla_{\mathbf{x}} \cdot)(\nabla_{\mathbf{y}} \cdot)[\nabla\Theta(\mathbf{x})C^{\epsilon\epsilon -1}(\mathbf{x},\mathbf{y})\nabla\Theta(\mathbf{y})]. \qquad (43)$$

Note that we have used integration by parts to write $\mathcal{F}[\Theta]$ and $\mathcal{G}[\Theta]$. The matrix $\mathcal{F}[\Theta]$ is manifestly symmetric and can be seen to have all nonnegative eigenvalues as follows: if we take any real map $X(\mathbf{x})$, then

$$X^T\mathcal{F}[\Theta]X = (\nabla\Theta \cdot \nabla X)^T C^{\epsilon\epsilon -1}(\nabla\Theta \cdot \nabla X). \qquad (44)$$

Since the inverse noise matrix $C^{\epsilon\epsilon -1}$ is symmetric and positive-definite, this quantity must be nonnegative. Indeed, this can only be zero if $\nabla\Theta \cdot \nabla X = 0$ everywhere, that is, if X is constant on flows of $\nabla\Theta$. If we have all-sky coverage and Θ is well behaved, then all of the flow curves of $\nabla\Theta$ connect at the maxima, minima, and saddle points of Θ, consequently in this case $\mathcal{F}$ is positive-definite except for the constant $l=0$ mode. Consequently, the matrix $C^{\Phi\Phi -1} + \mathcal{F}[\Theta]$ must be positive-definite, which is required for Eq. (41) to make sense as a probability distribution (this was also implicitly assumed in doing the Gaussian integral). We next define the (not symmetric!) matrix $H[\Theta]$ by

$$H[\Theta]X = \nabla\Theta \cdot \nabla X \qquad (45)$$

so that $\mathcal{F}[\Theta] = H[\Theta]^T C^{\epsilon\epsilon -1}H[\Theta]$. We can see, using integration by parts, that $\mathcal{G}[\Theta] = H[\Theta]^T C^{\epsilon\epsilon -1}(\hat{\Theta} - \Theta)$.

To make further progress, we use Eq. (45) to rewrite the first term on the right-hand side of Eq. (41):

$$\mathcal{R}[\Theta] = -\frac{1}{2}(\hat{\Theta} - \Theta)^T(C^{\epsilon\epsilon -1}H[\Theta]^{T-1}C^{\Phi\Phi -1}$$
$$\times H[\Theta]^{-1}C^{\epsilon\epsilon} + C^{\epsilon\epsilon})^{-1}(\hat{\Theta} - \Theta)$$
$$+\frac{1}{2}\Theta^T(C^{\Theta\Theta})^{-1}\Theta + \frac{1}{2}(\hat{\Theta} - \Theta)^T C^{\epsilon\epsilon -1}(\hat{\Theta} - \Theta)$$
$$+\frac{1}{2}\ln\det(C^{\Phi\Phi -1} + \mathcal{F}[\Theta]). \qquad (46)$$

Even this equation is too complicated to be useful in this form, so we will make the replacements $H[\Theta] \to H[\hat{\Theta}]$ and $F[\Theta] \to F[\hat{\Theta}]$. This converts Eq. (41) into a Gaussian posterior probability distribution. The peak of the posterior probability distribution is

$$\Theta_{\mathrm{PEAK}} = [C^{\epsilon\epsilon -1} - (C^{\epsilon\epsilon}H[\hat{\Theta}]^{T-1}C^{\Phi\Phi -1}$$
$$\times H[\hat{\Theta}]^{-1}C^{\epsilon\epsilon} + C^{\epsilon\epsilon})^{-1} + C^{\Theta\Theta -1}]^{-1}$$
$$\times [C^{\epsilon\epsilon -1} - (C^{\epsilon\epsilon}H[\hat{\Theta}]^{T-1}C^{\Phi\Phi -1}$$
$$\times H[\hat{\Theta}]^{-1}C^{\epsilon\epsilon} + C^{\epsilon\epsilon})^{-1}]\hat{\Theta}, \qquad (47)$$

and its covariance (i.e. inverse curvature) is

$$\mathrm{Cov}[\Theta] = [C^{\epsilon\epsilon -1} - (C^{\epsilon\epsilon}H[\hat{\Theta}]^{T-1}C^{\Phi\Phi -1}H[\hat{\Theta}]^{-1}C^{\epsilon\epsilon}$$
$$+ C^{\epsilon\epsilon})^{-1} + C^{\Theta\Theta -1}]^{-1}. \qquad (48)$$

It is instructive to compare Eq. (47) to other means of estimating Θ. Note that determination of Θ is a nontrivial task since both lensing and instrument noise must be taken into account. In the limit that lensing is negligible ($C^{\Phi\Phi} \to 0$), we derive $\Theta_{\mathrm{PEAK}} \approx (C^{\epsilon\epsilon -1} + C^{\Theta\Theta -1})^{-1}C^{\epsilon\epsilon -1}\hat{\Theta}$, which is recognizable as a simple Wiener filter of $\hat{\Theta}$. In the opposite limit, where instrument noise is negligible compared to the effects of lensing (i.e. $C^{\epsilon\epsilon} \to 0$), we derive

$$C^{\epsilon\epsilon -1} - (C^{\epsilon\epsilon}H[\hat{\Theta}]^{T-1}C^{\Phi\Phi -1}H[\hat{\Theta}]^{-1}C^{\epsilon\epsilon} + C^{\epsilon\epsilon})^{-1}$$
$$\approx H[\hat{\Theta}]^{T-1}C^{\Phi\Phi -1}H[\hat{\Theta}]^{-1} \qquad (49)$$

via a first-order Taylor expansion in $C^{\Phi\Phi -1}$. Substituting this into Eq. (47) yields

$$\Theta_{\mathrm{PEAK}} = (H[\hat{\Theta}]^{T-1}C^{\Phi\Phi -1}H[\hat{\Theta}]^{-1} + C^{\Theta\Theta -1})^{-1}$$
$$\times H[\hat{\Theta}]^{T-1}C^{\Phi\Phi -1}H[\hat{\Theta}]^{-1}\hat{\Theta}, \qquad (50)$$

which is recognizable as a Wiener-filtered temperature map with $H[\hat{\Theta}]C^{\Phi\Phi}H[\hat{\Theta}]^T$ playing the role of the noise covariance. This is not surprising since $H[\hat{\Theta}]C^{\Phi\Phi}H[\hat{\Theta}]^T$ is the covariance of the temperature change due to lensing, $\tilde{\Theta} - \Theta$, and under our assumptions the correlation between $\tilde{\Theta} - \Theta$ and Θ vanishes. Further simplification is possible by noting that, for zero noise, the likelihood gradient G_0 of Sec. III C may be written as $G_0 = -H[\hat{\Theta}]^T C^{\Theta\Theta -1}\hat{\Theta}$. Then the Fisher matrix of Sec. III C can be approximated as

$$F = \langle G_0 G_0^T \rangle_{\Phi=0} = \langle H[\Theta]^T C^{\Theta\Theta -1}\Theta\Theta^T C^{\Theta\Theta -1}H[\Theta]\rangle$$
$$= \langle H[\Theta]^T C^{\Theta\Theta -1}\langle\Theta\Theta^T\rangle C^{\Theta\Theta -1}H[\Theta]\rangle$$
$$= \langle H[\Theta]^T C^{\Theta\Theta -1}H[\Theta]\rangle \approx H[\Theta]^T C^{\Theta\Theta -1}H[\Theta]. \quad (51)$$

(The last equality on the first line is justified as follows: since $H[\Theta]$ is a linear function of Θ, and Θ is a Gaussian random field, Θ and $H[\Theta]$ are jointly Gaussian. Thus the expectation value of the four-point function $H[\Theta]^T C^{\Theta\Theta -1}\Theta\Theta^T C^{\Theta\Theta -1}H[\Theta]$ can be expanded using Wick's theorem as a sum of three terms, each of which is a product of two-point functions. Since $\langle G_0 \rangle = 0$, the final expression on the first line of Eq. (51) is the only nonvanishing term.) If we further assume that the lensing effect can be treated as a perturbation on the background CMB—an assumption that we have made already through the linear approximation—we can approximate $H[\hat{\Theta}] \approx H[\Theta]$, which allows Eq. (50) to be rewritten as

$$\Theta_{\text{PEAK}} = H[\hat{\Theta}](C^{\Phi\Phi-1}+F)^{-1}C^{\Phi\Phi-1}H[\hat{\Theta}]^{-1}\hat{\Theta}$$

$$= \hat{\Theta} - H[\hat{\Theta}](C^{\Phi\Phi-1}+F)^{-1}FH[\hat{\Theta}]^{-1}\hat{\Theta}$$

$$= \hat{\Theta} - H[\hat{\Theta}](C^{\Phi\Phi-1}+F)^{-1}H[\hat{\Theta}]^{T}C^{\Theta\Theta-1}\hat{\Theta}.$$
(52)

Using Eqs. (28) and (45) and integration by parts we see that $H[\hat{\Theta}]C^{\Theta\Theta-1}\hat{\Theta} = -G_0$. Also, comparison of Eq. (45) to the lensing operator definition, Eq. (7), indicates that in the linear approximation, $\Lambda[\Phi_1]\Theta_1 = \Theta_1 + H[\Theta_1]\Phi_1$. This allows us to simplify Eq. (52) to

$$\Theta_{\text{PEAK}} = \hat{\Theta} + H[\hat{\Theta}](C^{\Phi\Phi-1}+F)^{-1}G_0$$

$$= \Lambda[(C^{\Phi\Phi-1}+F)^{-1}G_0]\hat{\Theta}.$$
(53)

This is the observed temperature map "corrected" for lensing using the Wiener-filtered potential map, Eq. (37). It is thus the temperature analogue of the approach used in Ref. [12] for reconstructing primary polarization.

E. Onset of nonlinearity

We examine the validity of the linear approximation leading to Eqs. (36) and (37) using the real-space Taylor expansion of the lensing formula, Eq (7):

$$\tilde{\Theta} = \Lambda\Theta = \Theta + \nabla\Phi \cdot \nabla\Theta + \frac{1}{2}\nabla\Phi\nabla\Phi : \nabla\nabla\Theta + O(\Phi^3).$$
(54)

The quadratic estimator was constructed based on the first-order (i.e. order Φ^1) effect of lensing on $C^{\hat{\Theta}\hat{\Theta}}$, which neglects the second-order and higher terms in Eq. (54), as well as the covariance of the first-order term. Thus we expect that this approximation will be good if the ratio of successive terms R in Eq. (54) is small. As a simple (and naive!) first approach to determining when the linear approximation is valid, we note that if L denotes the typical multipole of Φ, and l denotes the typical multipole of Θ, then $R = Ll\Phi$. Since the mean square value of Φ is roughly $L^2 C_L^{\Phi\Phi}$, we find that $R^2 \approx L^4 C_L^{\Phi\Phi}l^2$, so the linear-order approximation breaks down at $C^{\Phi\Phi} > L^{-4}l^{-2}$. Given that $L^4 C_L^{\Phi\Phi}$ has a maximum of approximately 10^{-6}, we would then conclude that nonlinear effects could become important at $l > 1000$, i.e. Planck ($l_{\text{max}} \approx 1600$) and higher-resolution experiments might be susceptible to these effects.

A more refined version of this analysis would examine the covariance $C^{\hat{\Theta}\hat{\Theta}}(\mathbf{x},\mathbf{y})$ of the observed temperature instead of simply the temperature fluctuation. This is because for the long-wavelength lensing (Φ) modes, the second-order (Φ^2) corrections to the covariance are significantly less than calculated by the naive method above. Conceptually, one can understand this by noting that, because the primary CMB is statistically isotropic, $C^{\hat{\Theta}\hat{\Theta}}$ is sensitive to the relative, not absolute, deflection of photon trajectories. In the flat-sky approximation we have

$$C^{\hat{\Theta}\hat{\Theta}}(\mathbf{x},\mathbf{y}) = C^{\epsilon\epsilon}(\mathbf{x},\mathbf{y}) + C^{\Theta\Theta}(\mathbf{x}-\mathbf{y})$$

$$+ [\nabla\Phi(\mathbf{x}) - \nabla\Phi(\mathbf{y})] \cdot \nabla C^{\Theta\Theta}(\mathbf{x}-\mathbf{y})$$

$$+ \frac{1}{2}[\nabla\Phi(\mathbf{x}) - \nabla\Phi(\mathbf{y})][\nabla\Phi(\mathbf{x})$$

$$- \nabla\Phi(\mathbf{y})] : \nabla\nabla C^{\Theta\Theta}(\mathbf{x}-\mathbf{y}) + O(\Phi^3).$$
(55)

We note that if $\nabla\Phi$ is slowly varying compared to the separation of $\mathbf{x}$ and $\mathbf{y}$ (that is, $L\gamma \ll 1$ where $\gamma = |\mathbf{x}-\mathbf{y}|$), a near cancellation occurs between the linear terms in Eq. (55). This cancellation reduces the squared expansion parameter R^2 from $L^4 C_L^{\Phi\Phi}l^2$ to $\gamma^2 L^6 C_L^{\Phi\Phi}l^2$. Since γ can take on a wide range of values from the scale of the lensing mode, L^{-1}, down to the limit of the instrument's resolution l_{max}^{-1}, it is not at all clear how to proceed analytically with this approach.

F. Bias of quadratic estimator

Another way to measure the importance of nonlinear terms is to compute the bias in the quadratic estimator [Eq. (36)] for Φ, over an ensemble of primary CMB anisotropies Θ and instrument noises ϵ with the same lensing potential Φ. This bias vanishes in the linear approximation [2]. It can be computed by noting that the expectation value of $\Phi_{\text{eq (36)}}$ is a linear combination of covariance matrix elements of $C^{\hat{\Theta}\hat{\Theta}}$. We first switch to working in Fourier modes on a flat sky; in Fourier space, the two-mode correlation function of the observed temperature is given by the Fourier transform of Eq. (55):

$$\langle \hat{\Theta}_{\mathbf{l}_1} \hat{\Theta}_{\mathbf{l}_2} \rangle = (C_{l_1}^{\epsilon\epsilon} + C_{l_1}^{\Theta\Theta})\delta_{\mathbf{l}_1+\mathbf{l}_2,0}$$

$$+ \frac{1}{\sqrt{4\pi}}(\mathbf{l}_1+\mathbf{l}_2)\cdot(\mathbf{l}_1 C_{l_1} + \mathbf{l}_2 C_{l_2})\Phi_{\mathbf{l}_1+\mathbf{l}_2}$$

$$+ \frac{1}{8\pi}\sum_{\mathbf{k}_1}\Phi_{\mathbf{k}_1}\Phi_{\mathbf{k}_2}[(\mathbf{k}_1\cdot\mathbf{l}_1)(\mathbf{k}_2\cdot\mathbf{l}_1)C_{l_1}^{\Theta\Theta}$$

$$+ (\mathbf{k}_1\cdot\mathbf{l}_2)(\mathbf{k}_2\cdot\mathbf{l}_2)C_{l_2}^{\Theta\Theta} - 2(\mathbf{k}_1\cdot\mathbf{J})(\mathbf{k}_2\cdot\mathbf{J})C_J^{\Theta\Theta}]$$
(56)

where we have set $\mathbf{k}_2 = \mathbf{l}_1+\mathbf{l}_2-\mathbf{k}_1$. Then we may use this two-mode correlation function to evaluate the expectation values of Eqs. (28) and hence the quadratic estimator (36)

$$(\langle \Phi_{\text{eq (36)}} \rangle_{\Theta,\epsilon})_{\mathbf{L}} = \Phi_{\mathbf{L}} + \frac{1}{32\pi^{3/2}F_{\mathbf{L}}}\sum_{\mathbf{k}_1}\Phi_{\mathbf{k}_1}\Phi_{\mathbf{k}_2}\cdot\sum_{\mathbf{l}_1}Y_{\mathbf{l}_1,\mathbf{l}_2}$$

$$\times[(\mathbf{k}_1\cdot\mathbf{l}_1)(\mathbf{k}_2\cdot\mathbf{l}_1)C_{l_1}^{\Theta\Theta} + (\mathbf{k}_1\cdot\mathbf{l}_2)$$

$$\times(\mathbf{k}_2\cdot\mathbf{l}_2)C_{l_2}^{\Theta\Theta} - 2(\mathbf{k}_1\cdot\mathbf{J})(\mathbf{k}_2\cdot\mathbf{J})C_J^{\Theta\Theta}],$$
(57)

where we have set $\mathbf{l}_2 = \mathbf{L}-\mathbf{l}_1$, $\mathbf{k}_2 = \mathbf{L}-\mathbf{k}_1$, and $\mathbf{J} = \mathbf{k}_1-\mathbf{l}_1$, and
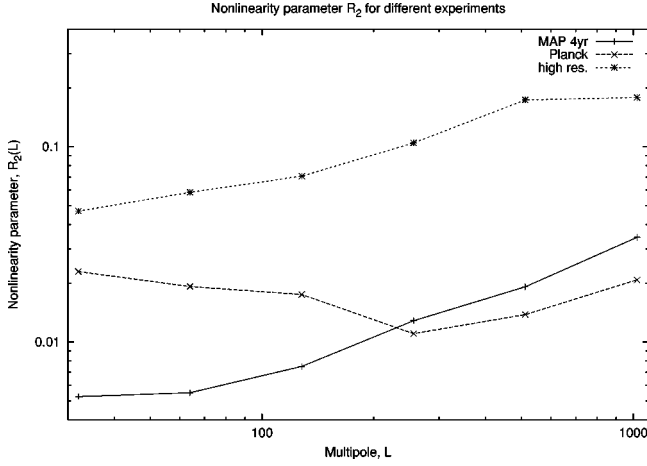
FIG. 1. The dimensionless nonlinearity parameter R_2, equal to the ratio of rms bias in the quadratic lensing potential estimator [Eq. (36)] to the rms value of the potential, is plotted here for several experiments as a function of multipole (wave number). Note that this quantity is less than unity for all of the experiments. See Table I for experiment parameters.

$$\Upsilon_{\mathbf{l}_1,\mathbf{l}_2} = \frac{\mathbf{L}\cdot(\mathbf{l}_1 C_{l_1}^{\Theta\Theta} + \mathbf{l}_2 C_{l_2}^{\Theta\Theta})}{(C_{l_1}^{\Theta\Theta} + C_{l_1}^{\epsilon\epsilon})(C_{l_2}^{\Theta\Theta} + C_{l_2}^{\epsilon\epsilon})}. \qquad (58)$$

We can then compute a mean-squared bias by squaring the bias and ensemble averaging over Φ. (Note that since different Fourier modes of Φ are uncorrelated, the terms in the sum over $\mathbf{k}_1$ will usually add incoherently. The exception to this rule is that terms related by switching $\mathbf{k}_1$ and $\mathbf{k}_2$ are equal, so the mean squared value of the sum is double the value obtained by summing the mean square of every term.) This mean-squared bias is then given by a quadrilateral integral:

$$\langle|\delta\Phi_{\mathbf{L}}|^2\rangle_\Phi = \langle|(\langle\Phi_{\text{eq (36)}}\rangle_{\Theta,\epsilon} - \Phi)|_{\mathbf{L}}^2\rangle_\Phi$$

$$\times \frac{1}{512\pi^6 F_L^2} \int d^2\mathbf{k}_1 C_{k_1}^{\Phi\Phi} C_{k_2}^{\Phi\Phi}$$

$$\times \left\{ \int d^2\mathbf{l}_1 \Upsilon_{\mathbf{l}_1,\mathbf{l}_2} [(\mathbf{k}_1\cdot\mathbf{l}_1)(\mathbf{k}_2\cdot\mathbf{l}_1)C_{l_1}^{\Theta\Theta} + (\mathbf{k}_1\cdot\mathbf{l}_2) \right.$$

$$\left. \times (\mathbf{k}_2\cdot\mathbf{l}_2)C_{l_2}^{\Theta\Theta} - 2(\mathbf{k}_1\cdot\mathbf{J})(\mathbf{k}_2\cdot\mathbf{J})C_J^{\Theta\Theta}] \right\}^2. \quad (59)$$

We can thus construct a nonlinearity parameter R_2 that is the ratio of the rms bias to the rms value of the lensing potential: $R_2^2 = \langle|\delta\Phi_{\mathbf{l}}|^2\rangle_\Phi / C_l^{\Phi\Phi}$. This nonlinearity parameter is plotted as a function of L for three experiments in Fig. 1; the parameters for the three experiments—MAP 4-year data, Planck, and a future high-resolution experiment—are shown in Table I. The R_2 nonlinearity parameter is small for all but the high-resolution experiment, indicating that the bias in the quadratic estimator [Eq. (36)] is small.

As a final means of testing the importance of the higher-order terms in the expansion of $C^{\hat{\Theta}\hat{\Theta}}$, we conduct a numerical "experiment" in Sec. V that compares nonlinear [Eq.

TABLE I. Reference parameters for CMB experiments.

Experiment	$w^{-1/2}/2.725$ K radian	$\sigma/$arcmin
MAP (4 yr)	5.6×10^{-8}	13
Planck	2.9×10^{-9}	6
High-res.	5.0×10^{-10}	1

(25), without the small and computationally difficult expectation value] and linear [Eq. (37)] lensing potential estimators. There we find only modest (10–20%) improvement in the rms error of the lensing potential reconstruction even for a high-resolution (1 arcminute beam) experiment.

IV. POWER SPECTRUM ESTIMATION

Having constructed an estimator for the lensing potential Φ, we next consider its power spectrum $C^{\Phi\Phi}$. Conceptually, the situation here is more complicated because once we average over an ensemble of lensing potentials derived from the same power spectrum, the lensed temperature field $\tilde{\Theta}$ is once again statistically isotropic with $\langle\tilde{\Theta}\tilde{\Theta}^T\rangle$ diagonal in harmonic space. (That is, the off-diagonal elements average to zero since $\langle\Phi\rangle=0$.) But we can still construct an estimator for $C^{\Phi\Phi}=\langle\Phi\Phi^T\rangle$ by taking the quadratic estimator for Φ and computing its "square." The resulting power spectrum estimator is thus constructed from the four-point correlation function of $\tilde{\Theta}$ or (in the presence of noise) $\hat{\Theta}$. It is thus measuring deviations of $\hat{\Theta}$ from Gaussianity. We will show that in the linear approximation, the maximum likelihood estimator reduces to the quadratic estimator.

We begin this section by formally writing out the likelihood function for the lensing power spectrum $C^{\Phi\Phi}$ as an integral, and then approximating this integral as Gaussian in Sec. IV A. In Sec. IV B, we approximate the curvature (inverse-covariance) matrix of this Gaussian in order to obtain a maximum likelihood estimator that is computationally tractable. We show in Sec. IV C that within the linear approximation, the MLE and quadratic estimator are equivalent. Computation of the primary CMB power spectrum $C^{\Theta\Theta}$ is considered in Sec. IV D, and cross-correlations, e.g. $C^{\Theta\Phi}$, are considered in Sec. IV E.

A. Likelihood function and Gaussian approximation

In principle, we could estimate the power spectrum $C_l^{\Phi\Phi}$ by constructing a grand likelihood function $\bar{\mathcal{L}}$ given (up to an additive constant to $\bar{\mathcal{L}}$, or equivalently a multiplicative constant to $e^{-\bar{\mathcal{L}}}$) by

$$\bar{\mathcal{L}} = -\ln\int \mathcal{D}\Phi P(\Phi|C^{\Phi\Phi})e^{-\mathcal{L}} = -\ln\int \mathcal{D}\Phi$$

$$\times \exp\left(-\mathcal{L}[\Phi] - \frac{1}{2}\Phi^T C^{\Phi\Phi-1}\Phi\right)$$

$$= -\ln\int \mathcal{D}\Phi e^{-\mathcal{P}}, \qquad (60)$$

where the integral over $\mathcal{D}\Phi$ is a functional integral with the usual measure on $\mathbf{R}^N$ where N is the number of pixels in the map. Note that $\bar{\mathcal{L}}$ is a function of the covariance $C^{\Phi\Phi}$; however, we cannot simply maximize the grand likelihood function, Eq. (60), and thus obtain an estimate of $C^{\Phi\Phi}$, because our map provides us with N real observations whereas $C^{\Phi\Phi}$ has $N(N+1)/2$ independent parameters. In order to obtain a meaningful result for the power spectrum, we must restrict the form of $C^{\Phi\Phi}$. Fortunately, the spherical $SO(3)$ symmetry of the sky provides us with just such a restriction—it forces $C^{\Phi\Phi}$ to be diagonal in l space. We will thus assume in this section that $C^{\Phi\Phi}$ can be written as a linear combination,

$$C^{\Phi\Phi}(\mathbf{x},\mathbf{y})=\sum_{\alpha} C_{\alpha}^{\Phi\Phi}\mathcal{C}_{\alpha}(\mathbf{x},\mathbf{y}), \qquad (61)$$

where the α's are indices labeling the basis covariance functions and we wish to evaluate the coefficients $C_{\alpha}^{\Phi\Phi}$. There are two interesting choices of basis function $\mathcal{C}_{\alpha}$. The first is the Legendre polynomials, which span the space of $C^{\Phi\Phi}$ that are consistent with symmetry requirements. These basis functions are given by

$$\mathcal{C}_{l}(\mathbf{x},\mathbf{y})=\frac{2l+1}{4\pi}P_{l}(\mathbf{x}\cdot\mathbf{y}). \qquad (62)$$

This results in coefficients $C_{l}^{\Phi\Phi}$ that are the power spectrum of Φ. The other choice, useful in the low singal-to-noise ratio (SNR) case, is to add several functions of the Eq. (62) type together to boost the overall SNR, i.e. to estimate the lensing power spectrum in a band rather than for each individual l. In this case the coefficient $C_{\alpha}^{\Phi\Phi}$ is a weighted average of the power spectrum over the range of l values covered by the basis function $\mathcal{C}_{\alpha}$.

We have now set up the maximum likelihood estimation problem for $C^{\Phi\Phi}$. Before proceeding to compute the maximum-likelihood point, we warn the reader that there is no guarantee that the likelihood function is devoid of local maxima. Most of the methods described here cannot avoid local maxima, nor can they be readily adapted to detect local maxima. The exception is the Markov chain method, although the number of iterations required to escape from a local maximum may be prohibitively large.

Since N is a large number (typically 10^6–10^7), brute-force integration of Eq. (60), does not appear feasible. There are at least two conceivable approaches to this problem: a Markov chain (MC) integration, or a Taylor expansion of the integrand. While the MC approach is dramatically faster than a brute-force integration, it is apparent from the high dimensionality (one dimension for each map pixel) of the problem that many iterations in the sequence will be necessary for convergence. We have not found a computationally feasible implementation of MC for this problem. The alternate approach is to Taylor expand $\mathcal{P}$ to quadratic order in Φ around its minimum $\Phi_{\min}$, i.e. to approximate the posterior probability distribution for Φ as a Gaussian. This gives

$$\bar{\mathcal{L}}[C_{\alpha}^{\Phi\Phi}]=-\ln\int\mathcal{D}\Phi\, e^{-\mathcal{P}[\Phi,C_{\alpha}^{\Phi\Phi}]}$$

$$=-\ln\int\mathcal{D}\Phi\,\exp\left(-\mathcal{L}[\Phi]-\frac{1}{2}\Phi^T C^{\Phi\Phi-1}\Phi\right)$$

$$\approx-\ln\int\mathcal{D}\Phi\,\exp\left(-\mathcal{P}[\Phi_{\min},C_{\alpha}^{\Phi\Phi}]\right.$$

$$\left.-\frac{1}{2}(\Phi-\Phi_{\min})^T\frac{\delta^2\mathcal{P}[\Phi_{\min},C_{\alpha}^{\Phi\Phi}]}{\delta\Phi\,\delta\Phi}(\Phi-\Phi_{\min})\right)$$

$$\approx\mathcal{P}[\Phi_{\min},C_{\alpha}^{\Phi\Phi}]+\frac{1}{2}\ln\det\frac{\delta^2\mathcal{P}[\Phi_{\min},C_{\alpha}^{\Phi\Phi}]}{\delta\Phi\,\delta\Phi}$$

$$\equiv\mathcal{P}[\Phi_{\min},C_{\alpha}^{\Phi\Phi}]+\frac{1}{2}\ln\det K, \qquad (63)$$

where the curvature matrix K (the matrix of second derivatives of $\mathcal{P}$ with respect to Φ, evaluated at $\Phi_{\min}$) has been introduced and an irrelevant additive constant has been dropped. Using Eq. (63), we seek to minimize the grand likelihood function $\bar{\mathcal{L}}$. To do this, we differentiate the final result of Eq. (63), yielding

$$0=\frac{\partial\bar{\mathcal{L}}}{\partial C_{\alpha}^{\Phi\Phi}}=\frac{\partial}{\partial C_{\alpha}^{\Phi\Phi}}\left(\mathcal{P}[\Phi_{\min},C_{\alpha}^{\Phi\Phi}]+\frac{1}{2}\ln\det K\right)$$

$$=\frac{\partial\mathcal{P}}{\partial C_{\alpha}^{\Phi\Phi}}+\frac{1}{2}\frac{\partial}{\partial C_{\alpha}^{\Phi\Phi}}\ln\det K|_{\Phi_{\min}}$$

$$+\frac{1}{2}\frac{\partial\Phi_{\min}^T}{\partial C_{\alpha}^{\Phi\Phi}}\frac{\delta\ln\det K}{\delta\Phi}\bigg|_{\Phi_{\min}}, \qquad (64)$$

where the final (chain-rule) term reflects the shifting position $\Phi_{\min}$ of the minimum as we change $C^{\Phi\Phi}$. There is no corresponding chain-rule term for $\mathcal{P}$ because at the minimum, $\delta\mathcal{P}/\delta\Phi$ vanishes. We may now evaluate the derivative of $\mathcal{P}$, noting that only the prior term in Eq. (17) has a dependence on $C^{\Phi\Phi}$. Combining the log-determinant of $C^{\Phi\Phi}$ from the prior with the log-determinant of K in Eq. (64) transforms Eq. (64) into

$$0=-\frac{1}{2}\Phi_{\min}^T(C^{\Phi\Phi})^{-1}\mathcal{C}_{\alpha}(C^{\Phi\Phi})^{-1}\Phi_{\min}$$

$$+\frac{1}{2}\frac{\partial}{\partial C_{\alpha}^{\Phi\Phi}}\ln\det(C^{\Phi\Phi}K)|_{\Phi_{\min}}$$

$$+\frac{1}{2}\frac{\partial\Phi_{\min}^T}{\partial C_{\alpha}^{\Phi\Phi}}\frac{\delta\ln\det K}{\delta\Phi}\bigg|_{\Phi_{\min}}. \qquad (65)$$

At this point we are confronted with the difficulty of computing the curvature matrix K. Unfortunately, brute force computation of K requires $O(N^2)$ computations of $\mathcal{P}$, each of which must require at least $O(N)$ elementary operations

since it accesses N data points; in practice, a computation of $\mathcal{P}$ involves spherical harmonic transforms consisting of $O(N^{3/2})$ operations.

B. Approximating the curvature matrix

In order to compute the curvature matrix in Eq. (65), we split it into two parts: the curvature of the likelihood function (F, which is the Fisher matrix in the Gaussian approximation) and the curvature of the prior, which is always $(C^{\Phi\Phi})^{-1}$: $K = F + (C^{\Phi\Phi})^{-1}$. This provides us with the identity

$$C^{\Phi\Phi}K = C^{\Phi\Phi}F + 1_{N \times N}, \qquad (66)$$

where $1_{N \times N}$ denotes the $N \times N$ identity matrix. We use the value of F from the Gaussian approximation: $F = \langle GG^T \rangle$. If we are far from boundaries or regions of nonuniform noise, F is diagonal in harmonic space and we may approximate it in bins accordingly:

$$F^{-1}(\mathbf{x},\mathbf{y}) \approx \sum_\alpha [F^{-1}]_\alpha \mathcal{C}_\alpha(\mathbf{x},\mathbf{y}) \equiv \sum_\alpha \frac{1}{F_\alpha} \mathcal{C}_\alpha(\mathbf{x},\mathbf{y}), \quad (67)$$

where the last equality defines the binned Fisher matrix F_α. In this approximation, $F = \langle GG^T \rangle$ reduces to

$$F_\alpha = \frac{1}{d_\alpha} \langle G^T \mathcal{C}_\alpha G \rangle, \qquad (68)$$

where the expectation value can be computed by a Monte Carlo analysis, and d_α is the number of lensing modes in the band covered by $\mathcal{C}_\alpha$. Technically it is best to compute the Fisher matrix at the value of $\Phi_{\min}$, however for purposes of computational tractability we only compute it once at $C^{\Phi\Phi} = 0$, $\Phi_{\min} = 0$ [see Eq. (96)]. In this approximation $\delta K / \delta \Phi$ vanishes so we will drop the final term in Eq. (65). We can then differentiate the log determinant of $C^{\Phi\Phi}K$ with respect to a power spectrum coefficient:

$$\frac{\partial}{\partial C_\alpha^{\Phi\Phi}} \ln \det(C^{\Phi\Phi}K) = \text{Tr}[(C^{\Phi\Phi}F + 1_{N \times N})^{-1} \mathcal{C}_\alpha F]$$

$$= \text{Tr}[(C^{\Phi\Phi} + F^{-1})^{-1} \mathcal{C}_\alpha]. \qquad (69)$$

With this approximation, Eqs. (65) and (69) give

$$\Phi_{\min}^T (C^{\Phi\Phi})^{-1} \mathcal{C}_\alpha (C^{\Phi\Phi})^{-1} \Phi_{\min} = \frac{d_\alpha}{C_\alpha^{\Phi\Phi} + F_\alpha^{-1}}, \qquad (70)$$

where the denominator in the second term uses the $C_l^{\Phi\Phi}$ value appropriate for the range of multipoles covered by the α basis function.

C. Linearization

If the lensing is sufficiently weak, i.e. if we are in the linear regime (see Sec. III E), and we are only using the Φ's

far from our boundary, we can solve Eq. (70) directly. To do this, begin by examining Eq. (70) in l space (assuming diagonality):

$$\sum_{lm} \frac{|\Phi_{lm}|^2}{(C_\alpha^{\Phi\Phi})^2} = \frac{1}{C_\alpha^{\Phi\Phi} + F_\alpha^{-1}}, \qquad (71)$$

where the sum is over the d_α lensing modes grouped into the band α. If we now take the multipole moments $\Phi_{lm} = (C_\alpha^{\Phi\Phi-1} + F_\alpha)^{-1}G_{0lm}$ given by Eq. (37), we derive

$$C_\alpha^{\Phi\Phi} = \frac{1}{d_\alpha F_\alpha^2} \sum_{lm} |G_{0lm}|^2 - F_\alpha^{-1}, \qquad (72)$$

which is the same result derived by Hu [2] in the flat-sky approximation. (It is valid in the all-sky approximation if we re-interpret F and G_0 as all-sky variables.)

D. Primary CMB power spectrum

The grand likelihood function $\bar{\mathcal{L}}$ defined in Eq. (60) contains the complete dependence of the probability density of $\hat{\Theta}$ on the primary CMB power spectrum, $C^{\Theta\Theta}$. Thus it can be simultaneously maximized over $C^{\Phi\Phi}$ and $C^{\Theta\Theta}$. We first parametrize the primary temperature power spectrum $C^{\Theta\Theta}$ in analogy to Eq. (61):

$$C^{\Theta\Theta}(\mathbf{x},\mathbf{y}) = \sum_\alpha C_\alpha^{\Theta\Theta} \mathcal{W}_\alpha(\mathbf{x},\mathbf{y}), \qquad (73)$$

where the $\mathcal{W}_\alpha$'s are the basis functions. Then we differentiate the Gaussian approximation, Eq. (64), with respect to the coefficients $C_\alpha^{\Theta\Theta}$ to determine the condition for maximization of the likelihood $e^{-\bar{\mathcal{L}}}$:

$$0 = \frac{\partial \bar{\mathcal{L}}}{\partial C_\alpha^{\Theta\Theta}} \approx \frac{\partial}{\partial C_\alpha^{\Theta\Theta}} \left(\mathcal{P}[\Phi_{\min}, C_\alpha^{\Phi\Phi}] + \frac{1}{2} \ln \det K \right)$$

$$= \frac{\partial \mathcal{P}}{\partial C_\alpha^{\Theta\Theta}} + \frac{1}{2} \frac{\partial}{\partial C_\alpha^{\Theta\Theta}} \ln \det K \Big|_{\Phi_{\min}}$$

$$+ \frac{1}{2} \frac{\partial \Phi_{\min}^T}{\partial C_\alpha^{\Theta\Theta}} \frac{\delta \ln \det K}{\delta \Phi} \Big|_{\Phi_{\min}}. \qquad (74)$$

We proceed in analogy to our analysis of the lensing potential power spectrum in Sec. IV B. We neglect the change of $\det K$ with Φ, thus eliminating the last term in Eq. (74). We can simplify the first term by noting that $\mathcal{P}$ consists of a prior and the unmarginalized likelihood $\mathcal{L}$; the prior has no dependence on $C_\alpha^{\Theta\Theta}$, while the unmarginalized likelihood [given by Eq. (15)] has derivative

$$\frac{\partial \mathcal{L}}{\partial C_\alpha^{\Theta\Theta}} = -\frac{1}{2}\hat{\Theta}^T C^{\hat{\Theta}\hat{\Theta}-1}\Lambda[\Phi]\mathcal{W}_\alpha\Lambda[\Phi]^T C^{\hat{\Theta}\hat{\Theta}-1}\hat{\Theta}$$

$$+\frac{1}{2}\mathrm{Tr}(C^{\hat{\Theta}\hat{\Theta}-1}\Lambda[\Phi]\mathcal{W}_\alpha\Lambda[\Phi]^T). \tag{75}$$

Combining this with Eq. (74), and using Eq. (66) to eliminate K in favor of F in the last term, gives us

$$0 = -\frac{1}{2}\hat{\Theta}^T C^{\hat{\Theta}\hat{\Theta}-1}\Lambda[\Phi]\mathcal{W}_\alpha\Lambda[\Phi]^T C^{\hat{\Theta}\hat{\Theta}-1}\hat{\Theta}$$

$$+\frac{1}{2}\mathrm{Tr}(C^{\hat{\Theta}\hat{\Theta}-1}\Lambda[\Phi]\mathcal{W}_\alpha\Lambda[\Phi]^T)$$

$$+\frac{1}{2}\,\mathrm{Tr}\left[(F+C^{\Phi\Phi-1})^{-1}\frac{\partial F}{\partial C_\alpha^{\Theta\Theta}}\right]. \tag{76}$$

One can readily see that in the absence of lensing, the final term in this equation vanishes, the $\Lambda[\Phi]$ matrices become the identity, and this equation reduces to the standard maximum likelihood result for CMB power spectrum estimation:

$$0 = -\frac{1}{2}\,\hat{\Theta}^T C^{\hat{\Theta}\hat{\Theta}-1}\mathcal{W}_\alpha C^{\hat{\Theta}\hat{\Theta}-1}\hat{\Theta} + \frac{1}{2}\,\mathrm{Tr}(C^{\hat{\Theta}\hat{\Theta}-1}\mathcal{W}_\alpha). \tag{77}$$

E. Correlation of lensing with other observables

We may want to compute the correlation of the lensing potential Φ with some other quantity. Examples could include the CMB temperature $\tilde{\Theta}$, Sunyaev-Zel'dovich or x-ray observations of hot gases, or galaxy maps. Since the focus of this paper is on likelihood methods, and approximations to them, we will restrict our attention here to the case of determining $C_l^{Z\Phi}$ where Z is an observable which has a jointly Gaussian distribution with Φ. This situation is expected to be a very good approximation for the CMB-lensing correlation $C_l^{\tilde{\Theta}\Phi}$ introduced by the ISW effect, since ISW is expected to be apparent primarily on large scales which are still in the linear regime; some non-Gaussianity in the potential-induced $\tilde{\Theta}$ fluctuations may be expected from nonlinear growth at $l > 100$ [13], but this should have negligible effect on the expected signal to noise. For the other observables, the situation is complicated by nonlinear evolution and the method described here should be used with caution.

We will neglect any error in the determination of Z. This is not as restrictive an assumption as it might seem; if we wish to cross-correlate Φ with an observable that has Gaussian error bars, we may write

$$Z = \check{Z} + \zeta, \tag{78}$$

where Z is the measured value of the observable, $\check{Z}$ is the actual value, and ζ is the error. If ζ is Gaussian and independent of $\check{Z}$ or Φ, and $\check{Z}$ is jointly Gaussian distributed with the lensing potential Φ, we infer the relations

$$C_l^{ZZ} = C_l^{\check{Z}\check{Z}} + C_l^{\zeta\zeta} \quad \text{and} \quad C_l^{Z\Phi} = C_l^{\check{Z}\Phi}, \tag{79}$$

so that estimating the cross-correlation of Z and Φ becomes equivalent to measuring the desired correlation $C_l^{\check{Z}\Phi}$. We can then construct the likelihood function

$$\bar{\mathcal{L}}[C_\alpha^{\Phi\Phi}, C_\alpha^{Z\Phi}] = -\ln\int \mathcal{D}\Phi \exp\left[-\mathcal{L}[\Phi] - \frac{1}{2}\binom{Z}{\Phi}^T\right.$$

$$\left. \times \begin{pmatrix} C^{ZZ} & C^{Z\Phi} \\ C^{\Phi Z} & C^{\Phi\Phi} \end{pmatrix}^{-1}\binom{Z}{\Phi}\right]. \tag{80}$$

The estimators of Secs. III B and IV B need only minor modification in order to do a joint maximum-likelihood analysis of $C^{\Phi\Phi}$ and $C^{Z\Phi}$. To see this, note that for a joint distribution with specified covariance, the expected value of Φ given Z is

$$E[\Phi|Z] \equiv \langle\Phi\rangle|_Z = C^{\Phi Z}C^{ZZ-1}Z \equiv AZ, \tag{81}$$

where we have defined the slope matrix $A = C^{\Phi Z}C^{ZZ-1}$. The variance given Z is

$$C^{\Phi\Phi}|_Z \equiv \langle(\Phi - \langle\Phi\rangle|_Z)^2\rangle = C^{\Phi\Phi} - C^{\Phi Z}C^{ZZ-1}C^{Z\Phi}, \tag{82}$$

where we have used $C^{Z\Phi}(\mathbf{x},\mathbf{y}) \equiv \langle Z(\mathbf{x})\Phi(\mathbf{y})\rangle$. (Note that $C^{Z\Phi}$ is the matrix transpose of $C^{\check{Z}\Phi}$.) Equations (81) and (82) are general for any joint Gaussian distribution, hence they are valid here even considering the existence of boundaries. Using them, we can re-write the likelihood function [Eq. (80)] as

$$\bar{\mathcal{L}}[C_\alpha^{\Phi\Phi}, C_\alpha^{Z\Phi}] = \frac{1}{2}Z^T C^{ZZ-1}Z - \ln\int \mathcal{D}\Phi$$

$$\times \exp\left\{-\mathcal{L}[\Phi] - \frac{1}{2}(\Phi - E[\Phi|Z])^T\right.$$

$$\left. \times [C^{\Phi\Phi}|_Z]^{-1}(\Phi - E[\Phi|Z])\right\}, \tag{83}$$

which is of the same form as the first (exact!) line of Eq. (60). The additive constant $\frac{1}{2}Z^T C^{ZZ-1}Z$ has no effect since we take Z and C^Z to be constant, so the estimators developed earlier in this paper to compute Φ and $C^{\Phi\Phi}$ can be re-written to compute $\Phi - E[\Phi|Z]$ and $C^{\Phi\Phi}|_Z$, respectively. We next construct these estimators before turning our attention to the problem of estimating the slope matrix A that relates Z to $E[\Phi|Z]$.

If we are sufficiently far from a boundary, we can diagonalize in harmonic space to yield

$$E\Phi|Z| = \left[\sqrt{\frac{C_l^{\Phi\Phi}}{C_l^{ZZ}}}\rho_l^{Z\Phi}\right]Z = A_l Z$$

and

$$C^{\Phi\Phi}{}_Z = C^{\Phi\Phi}(1 - \rho_l^{Z\Phi 2}), \tag{84}$$

where $\rho_l^{Z\Phi}$ is the correlation coefficient of the lth-order multipoles. We note at this point that (assuming C^{ZZ} is known or has been separately measured) that the sets of variables $(C_l^{\Phi\Phi}, C_l^{Z\Phi})$, $(C_l^{\Phi\Phi}, \rho_l^{Z\Phi})$, and $(C_l^{\Phi\Phi}|_Z, A_l)$ are merely different parametrizations of the same space of models. We can estimate any of these pairs; $(C_l^{\Phi\Phi}|_Z, A_l)$ is introduced here precisely because it is the easiest to estimate directly. We can now immediately convert Eq. (25), without the trace, to yield

$$\Phi = E[\Phi|Z] + C^{\Phi\Phi}|_Z G \tag{85}$$

as the mode of the posterior probability distribution [where G is the likelihood gradient as specified in Eq. (28)]. The power spectrum estimator result, Eq. (70), becomes

$$1 = \frac{C_\alpha^{\Phi\Phi}|_Z + F_\alpha^{-1}}{d_\alpha} G^T \mathcal{C}_\alpha G, \tag{86}$$

where G in Eq. (86) is evaluated at the solution to Eq. (85). These equations specify the conditions for the likelihood $\bar{\mathcal{L}}(C^{\Phi\Phi}, \rho^{Z\Phi})$ to be stationary with respect to first-order variations in $C^{\Phi\Phi}|_Z$ with A_l constant. In order to complete the analysis, we must also identify the condition for $\bar{\mathcal{L}}$ to be stationary with respect to first-order variations in the A_l (slope) coefficients in Eq. (84) with $C^{\Phi\Phi}|_Z$ constant. For these variations, if we again approximate the Fisher matrix as $F = \langle GG^T \rangle|_{\Phi=0}$, we derive a constant curvature matrix K. Then, parametrizing the A_l in bands in analogy to Eq. (61) gives

$$A(\mathbf{x}, \mathbf{y}) = \sum_\alpha A_\alpha \mathcal{B}_\alpha(\mathbf{x}, \mathbf{y}), \tag{87}$$

where it is assumed that $\mathcal{B}_\alpha$ and hence A are symmetric. We obtain a maximum likelihood condition on A_α by differentiating $\bar{\mathcal{L}}$:

$$\frac{\partial \bar{\mathcal{L}}}{\partial A_\alpha} = \frac{\partial \mathcal{P}}{\partial A_\alpha} = \frac{\partial \mathcal{L}}{\partial A_\alpha} = \int_\Omega d^2\mathbf{x} \frac{\delta \mathcal{L}}{\delta \Phi(\mathbf{x})} \frac{\partial \Phi(\mathbf{x})}{\partial A_\alpha}$$

$$= \int_\Omega d^2\mathbf{x} \frac{\delta \mathcal{L}}{\delta \Phi(\mathbf{x})} \frac{\partial E[\Phi|Z](\mathbf{x})}{\partial A_\alpha} = Z^T \mathcal{B}_\alpha G. \tag{88}$$

Note that we have taken $\Phi - E[\Phi|Z]$ to be constant here; this was merely a convenient choice. [Since we have maximized $\mathcal{P}$ with respect to Φ, we can choose any first-order variation in Φ without affecting the derivative in Eq. (88).] It follows that the joint maximum-likelihood estimator for $(C^{\Phi\Phi}, A)$ satisfies

$$Z^T \mathcal{B}_\alpha G = 0. \tag{89}$$

We can then reconstruct the full lensing power spectrum and cross-correlation using the relations

$$C_l^{\check{Z}\Phi} = C_l^{ZZ} A_l \quad \text{and} \quad C_l^{\Phi\Phi} = A_l^2 C_l^{ZZ} + C_l^{\Phi\Phi}|_Z. \tag{90}$$

If we take the linear approximation of Sec. III E, that $G \approx G_0 + F\Phi$, and approximate diagonality in harmonic space as in Sec. IV C, Eq. (89) becomes

$$A_\alpha = -\frac{Z^T \mathcal{B}_\alpha G_0}{Z^T \mathcal{B}_\alpha F \mathcal{B}_\alpha Z}. \tag{91}$$

If we note that $C^{Z\Phi} = C^{ZZ} A_l$, we note that this is the same result as obtained by correlating the linearized maximum likelihood estimator, Eq. (36) with Z. In the case of the $\tilde{\Theta}\Phi$ correlation ($Z = \hat{\Theta}$), which is of interest for investigating the ISW effect, the numerator is cubic in $\hat{\Theta}$, i.e. the maximum likelihood estimator for A_α is the same as that computed from the bispectrum.

As a final point, we note that for the $\tilde{\Theta}\Phi$ correlation, the error $\zeta = \epsilon$ in $Z = \hat{\Theta}$ is not entirely independent of the estimation procedure for Φ, since we are after all determining Φ from the CMB temperature measurements. Since we assumed ζ to be uncorrelated with Φ and its determination, this is a potential flaw in our calculations as applied to the $\tilde{\Theta}\Phi$ correlation. We expect the error induced by this effect to be small, since the ISW effect is most important on the large scales where the instrument noise is small: $C_l^{\epsilon\epsilon} \ll C_l^{\Theta\Theta}$. We additionally note that the determination of Φ primarily uses information from much higher l.

V. IMPLEMENTATION AND RESULTS

In order to demonstrate the feasibility of computing the estimators above in a realistic situation, and to assess their performance, we ran several simulations in which a data set was generated and analyzed. The data sets are generated on a full sphere assuming isotropic Gaussian temperature fluctuations, lensing potential, and instrument noise. For Planck and the high-resolution reference experiment ($l_{\max} \approx 3500$, beam full width at half maximum$=1$ arcminute), we reconstruct the lensing potential and compare the reconstruction and original map. The lensing power spectrum was estimated for the Planck-type experiment, but computer time constraints prevented a similar analysis for the high-resolution experiment.

A. Utilities

A lensing simulation requires the capability to work with maps on the unit sphere, or some subset thereof, particularly the capability to perform the elementary algebraic and calculus operations and to perform convolutions and both forward and reverse spherical harmonic transforms (SHT). We therefore require the use of a map projection or grid. In order to perform SHT in a reasonable amount of time, we must use an isolatitude projection, i.e. one in which horizontal lines are parallels of the same latitude. Furthermore, we found conformality to be convenient for differentiation and useful for reducing gridding errors (see Sec. V B). The only projection with these properties is the Mercator projection, in which the

map coordinates x and y are related to the longitude ϕ and colatitude θ by the formulas $\phi = \tau(x - x_0)$ and $\cos \theta = \tanh(\tau y)$.

The conformal magnification Γ defined by $ds^2 = \Gamma^2(dx^2 + dy^2)$ satisfies $\Gamma = \tau \sin \theta$. A map of some quantity A is stored as a two-dimensional array $A(x,y)$ of values at the points of integer x and y. Spherical harmonic transforms are performed by transforming first the longitude (x or ϕ) direction to produce the Fourier coefficients of A at constant latitude and then the latitude (y or θ) direction. On a grid with N points, this is an $O(N^{3/2})$ process. Convolutions are performed with two successive SHT's.

B. Estimator for map of lensing potential

Our implementation presently approximates the estimator (25) as follows. The expectation value is ignored since it is expected to be small; see the discussion following Eq. (25). We must also approximate the vector

$$\mathbf{V} = -[(C^{\hat{\Theta}\hat{\Theta}}[\Phi]^{-1}\hat{\Theta})\Lambda[\Phi]\nabla C^{\Theta\Theta}\Lambda[\Phi]^{-1}$$
$$\times (C^{\hat{\Theta}\hat{\Theta}}[\Phi])^{-1}\hat{\Theta}], \quad (92)$$

and use it to determine the likelihood gradient $G = \nabla \cdot \mathbf{V}$. (Note that $G = \delta\mathcal{L}/\delta\Phi$ is a scalar function on Ω.) Because of difficulties computing $C^{-1}\hat{\Theta}$ in a reasonable amount of time, we chose to approximate $C^{-1}\hat{\Theta}$ by a sequence of (i) filtering of $\Lambda^{-1}\hat{\Theta}$ using the harmonic-space kernel $C^{\Theta\Theta}/(C^{\Theta\Theta} + C^{\epsilon\epsilon})$ and (ii) convolution with the $(C_l)^{-1}$ kernel. Of these steps, both break down near the boundaries and (ii) breaks down when the lensing is strong enough so that the noise $C^{\epsilon\epsilon}$ in $\hat{\Theta}$ is no longer a good approximation to the noise in $\Lambda^{-1}\hat{\Theta}$. We note that if $C^{\epsilon\epsilon}$ were flat (i.e. $l^2 C^{\epsilon\epsilon} \propto l^2$), this approximation would become exact far from the boundaries. ("Real" instrument errors show some increase in $C^{\epsilon\epsilon}$ at high l due to finite beam size [14].) In order to reduce gridding errors, the $(C_l)^{-1}$ operation is performed by convolving with the kernel $[l(l+1)C_l]^{-1}$ and then taking the Laplacian. In order to avoid boundary effects, $\mathbf{V}$ is multiplied by a function q that is equal to one inside Ω far from the boundary, but falls off smoothly to zero at the boundary.

After computing $\mathbf{V}$, we take its divergence $G = \nabla \cdot \mathbf{V}$; then we must determine $C^{\Phi\Phi}G$. In order to reduce errors due to the gridding (pixelization), we perform the convolution in two steps. First, we apply an inverse Laplacian operator ∇^{-2}, and then we apply the remainder of the convolution, $l(l+1)C_l^{\Phi\Phi}$. Because we use a conformal coordinate system, the inverse Laplacian can be done in the plane where gridding errors vanish (the forward and reverse Fourier transforms are exact inverses of each other, even on a discrete grid, which does not occur for SHT). This is important since the low-l modes of G, which correspond to the lensing modes that can be recovered at moderate signal-to-noise ratio, are buried in high-l noise due to the power spectrum $l^2 C_l^{GG} \propto \approx l^5$ in the range of interest $50 \leq l \leq 1000$. The inverse Laplacian operation does not add significant time to the computation because it utilizes a fast Fourier transform re-

quiring $O(N \log N)$ operations, whereas the computation time is dominated by SHT's requiring $O(N^{3/2})$ time. With some attention paid to gridding issues, this two-step process may turn out to be unnecessary.

An iterative procedure is needed for solving Eq. (25). The obvious iterative procedure, $\Phi_{n+1} = C^{\Phi\Phi}G[\Phi_n]$, is (in the linear approximation) unstable for any lensing mode with $\text{SNR}^2 = C^{\Phi\Phi}F > 2$, and hence is not a good choice. We therefore use the underrelaxed version,

$$\Phi_{n+1} = (1-f)\Phi_n + fC^{\Phi\Phi}G[\Phi_n], \quad (93)$$

where f is a convergence parameter. In the linear approximation, convergence would require $0 < f < 2/(1 + \text{SNR}^2)$, however, a smaller value of f is necessary in practice to avoid instabilities resulting from boundary effects and nonlinear lensing effects.

C. Power spectrum

We use Eq. (70) to estimate the power spectrum $C^{\Phi\Phi}$. The basis functions of choice have constant $l^2(l+1)^2 C_l^{\Phi\Phi}$ within some band $l_{\min} \leq l \leq l_{\max}$. The number of modes covered by the basis function $\mathcal{C}_\alpha$ can be estimated as

$$d_\alpha = A(\Omega) \sum_{l=l_{\min}}^{l_{\max}} \frac{2l+1}{4\pi} = \frac{A(\Omega)}{4\pi}[(l_{\max}+1)^2 - l_{\min}^2]. \quad (94)$$

The estimator, Eq. (70), then can be written in the iterative form:

$$C_\alpha^{\Phi\Phi} = C_\alpha^{\Phi\Phi}\left[\frac{C_\alpha^{\Phi\Phi} + F_\alpha^{-1}}{d_\alpha}(\nabla^{-2}G)^T P_\alpha \nabla^{-2}G\right]^\beta, \quad (95)$$

where P_α is the projector onto the band $l_{\alpha,\min} \leq l \leq l_{\alpha,\max}$, i.e. the operation that filters out all multipoles not included in this band. We use $\nabla^{-2}G$ here because it and its spherical harmonic transform are already being computed for the estimation of the map of Φ. The parameter β is an adjustable convergence parameter. The Fisher matrix F_α is computed by the Monte Carlo procedure:

$$F_\alpha = \frac{1}{d_\alpha}\langle(\nabla^{-2}G_0)^T P_\alpha \nabla^{-2}G_0\rangle, \quad (96)$$

where the average is taken over an unlensed temperature field (including noise).

Note that Eq. (95) exhibits a difference from the quadratic estimator, Eq. (72): while Eq. (72) can, in principle, be negative, P_α is positive definite and hence Eq. (95) can never yield any result less than zero. It is straightforward to show that in this case, assuming the linearized approximation of Secs. III C and IV C, and assuming a positive initial guess is used for the power spectrum to start the iteration, that Eq. (95) tends to zero (estimates no power). Because negative results are replaced by zeroes, Eq. (95) technically converges to a biased estimator, with expectation value

$$\frac{\langle C_\alpha^{\Phi\Phi}(\text{est.})\rangle}{C_\alpha^{\Phi\Phi}(\text{actual})} = 1 - \frac{1}{2}\,\text{erfc}\frac{x}{\sqrt{2}} + \frac{1}{\sqrt{2\pi}}\,x^{-1}e^{-x^2/2}$$

$$= 1 + \pi^{-1/2}e^{-x^2/2}$$

$$\times (x^{-3} - 3x^{-5} + 5!!\,x^{-7} - \cdots), \quad (97)$$

where $x = d_\alpha C_\alpha^{\Phi\Phi} F_\alpha/2$ is the squared signal-to-noise ratio (SNR^2) in the power spectrum determination. If the signal-to-noise ratio is large ($x \gg 1$) then this bias is irrelevant. Note that in the context of maximum likelihood estimation, a negative power spectrum estimate does not make sense, because the corresponding probability distribution for Φ and hence the likelihood integral, Eq. (60), are ill defined. In particular, the avoidance of negative power spectra is not an artifact of any approximation we have made.

Obtaining convergence from the coupled iterative estimators, Eqs. (93) and (95) requires some care. Convergence depends not only on the values of the parameters f and β, but also on the pattern of how many times the map is updated using Eq. (93) each time the power spectrum is updated using Eq. (95). As an extreme example, we note that if f and β are taken to be very small, and we alternate between updating the map and power spectrum, convergence can be expected only for negative β; whereas if we iterate the map many times between iterations of the power spectrum, convergence requires positive β. After some experimentation, we found that iterating the map many ($M \gg 1/f$) times between iterations of the power spectrum and taking $\beta = 1$ resulted in convergence.

D. Improvements

While the implementation described here is sufficient for evaluating the importance of nonlinear effects, much work remains before it could be used to analyze real data. First, real data have boundaries (if for no other reason than the presence of a galactic plane cut) and usually have inhomogeneous noise. Thus, the C^{-1} operation used here will need to be performed by actual matrix inversion rather than by convolution. The latter also becomes necessary in the event of nonuniform noise. Also, the iteration of equation Eq. (95) converges slowly and for long-wavelength modes ($1/l$ comparable to the size of the gridded region) may fail to converge entirely.

Additionally, it would be desirable to use a better approximation to Eq. (60) than the Gaussian approximation, Eq. (63), but we were unable to identify a computationally tractable method of doing this.

E. Results

Here we investigate the effects of nonlinearity on future CMB experiments. We use the form for the instrument noise [14]:

$$C_l^{\epsilon\epsilon} = w^{-1} e^{l(l+1)\sigma^2/8\ln 2}, \quad (98)$$

where the weight w and beam full-width at half maximum σ are parameters, and the beam spot is assumed to be Gaussian.
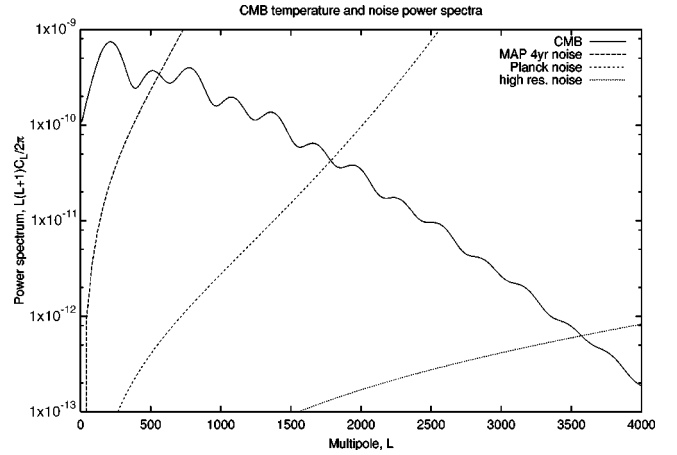

FIG. 2. The solid line illustrates the model primary CMB temperature power spectrum $l(l+1)C_l^{\Theta\Theta}/2\pi$. The noise curves $l(l+1)C_l^{\epsilon\epsilon}/2\pi$ are shown for MAP 4-year data (top, long-dashed), Planck (center, short-dashed), and the high resolution reference experiment (bottom, dotted).

We use a primary CMB power spectrum $C_l^{\Theta\Theta}$ generated by CMBFAST, assuming a flat universe with a cosmological constant and parameters $H_0 = 72$ km/s/Mpc, $T_{\text{CMB}} = 2.725$ K, $Y_{\text{He}} = 0.24$, $N_\nu = 3$ (massless), $\Omega_b = 0.04$, $\Omega_{cdm} = 0.30$. The primary CMB model is shown in Fig. 2. The lensing power spectrum, shown in Fig. 3, is computed normalized to $\sigma_8 = 1$.

We compare the linearized estimator to the "full" nonlinear estimator (as implemented here) for two experiments: the upcoming Planck satellite mission, and the proposed Atacama Cosmology Telescope (ACT) as an example of upcoming high resolution, low noise experiments. The parameters for the Planck and the high-resolution reference experiments are shown in Table I. (The MAP 4-year experiment is also shown for comparison.) For purposes of computational tractability, we have restricted ourselves to a small portion of the
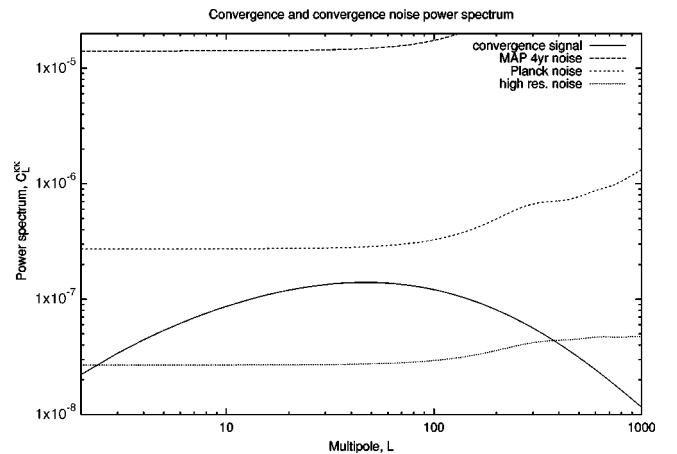

FIG. 3. The solid line illustrates the model convergence power spectrum $C_l^{\kappa\kappa} = l^2(l+1)^2 C^{\Phi\Phi}/4$. The noise curves $l^2(l+1)^2(F_l^{\Phi\Phi})^{-1}/4$ are shown for (top to bottom) MAP 4-year data, Planck, and the high-resolution reference experiment, using curvature matrix elements F_l computed from Eq. (35).
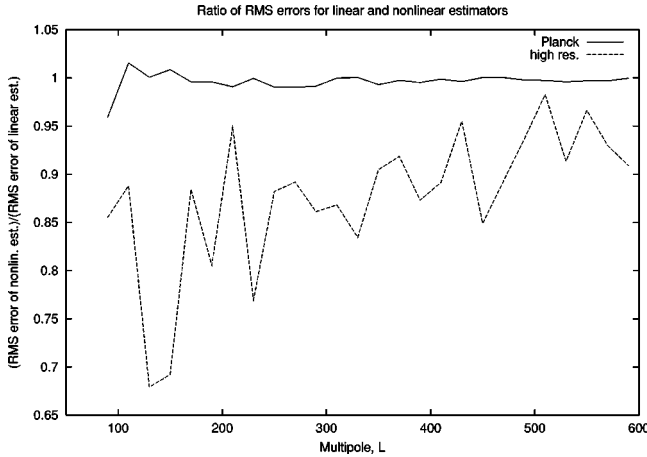
FIG. 4. The ratio of the root mean squared error ($\sqrt{\text{MSE}}$) for the nonlinear estimator, Eq. (93), to that of the linear estimator, Eq. (37), in bins of $\Delta l = 20$. Results are obtained from a Monte Carlo simulation, which is responsible for the bumpiness of the graph. The solid line is for Planck parameters, and the dotted line is for the high-resolution experiment (see Table I).

FIG. 5. The true convergence power spectrum, $C_l^{\kappa\kappa}$, is shown by the solid line. The points (with error bars) indicate estimated convergence power spectra from the linear ($+$ points) and nonlinear ($\times$ points) estimators [Eqs. (72) and (95) respectively]. To prevent the error bars from overlapping and causing confusion, we have displaced the data point for the linear estimator slightly to the left and the data point for the nonlinear estimator slightly to the right. The error bars are the 1σ Monte Carlo error bars on the expectation value of the estimator. The estimated power spectra plotted are averages over 10 trials of 0.14 sr solid angle each using Planck parameters, and thus shows the error bar on the power spectrum using data from a region of area $A(\Omega) = 1.4$ sr.

sky: the Planck simulation was run on a 751×751 grid with spacing at equator $\tau = 5 \times 10^{-4}$ radians (even though Planck is an all-sky experiment), and the high-resolution simulation was run on a 1251×1251 grid with $\tau = 1.5 \times 10^{-4}$ radians. (Note that the ACT survey region is a long, rectangular stripe on the sky as opposed to a more compact patch. Because of our implementation's susceptibility to boundary effects, we cannot do our simulations on a stripe.) The solid angles covered by the simulations are 0.14 sr ($\sim 2\%$ of the Planck survey area) for the Planck-type experiment and 0.035 sr for the high-resolution experiment.

The results of these simulations are shown in Fig. 4. The convergence map errors (i.e. $\kappa_{\text{est}} - \kappa$) for both the nonlinear estimator, Eq. (95) and the linear estimator, Eq. (37) were computed. The convergence map errors were then Fourier-transformed (since we are working on a small patch of sky), yielding the error amplitude $\kappa_{\mathbf{l}}$ for each Fourier mode. The modes were then sorted into bins of $\Delta l = 20$ according to their l value, and an RMS amplitude $\sqrt{\bar{\kappa}_{\mathbf{l}}^* \kappa_{\mathbf{l}}}$ was computed for each bin. The ratios of these RMS amplitudes are plotted in Fig. 4. Note that for the Planck experiment, there is only a slight advantage in using the nonlinear estimator, whereas for the high-resolution experiment, the accuracy of the reconstruction is improved by using the full nonlinear estimator, Eq. (37).

Both the comparison via simulation of the linear and non-linear estimators (Fig. 4) and the semianalytic bias calculation (Fig. 1) are methods of assessing the validity of the linear approximation. Both of them suggest that nonlinear effects are more important for the higher-resolution experiment than for Planck, but (at least for the experiment considered here) are not dominant. Note, however, that for the high-resolution experiment the semianalytic calculation found nonlinear effects to be more important at higher l, whereas the simulation found a greater improvement in switching to the nonlinear estimator at lower l. Note, how-
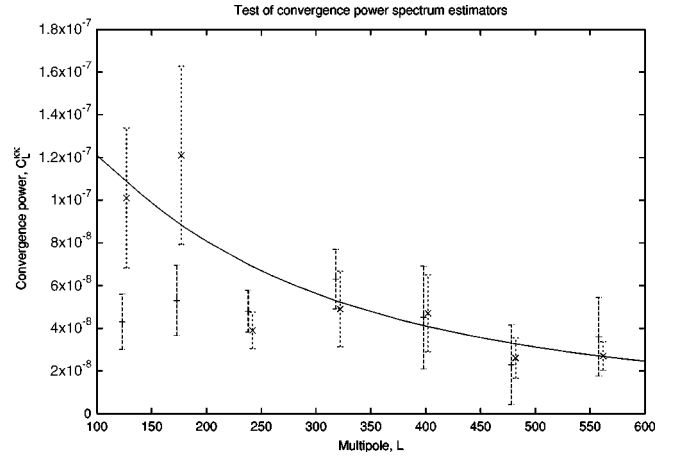
ever, that the semianalytical calculations of Sec. III E and the simulation of this section are not measuring the same quantity: in Sec. III E we were examining the bias of Eq. (36), whereas here we are considering the mean squared error of the optimally filtered version of that estimator.

We also simulated the performance of the linear [Eq. (72)] and nonlinear [Eq. (95), 16 iterations] convergence power spectrum estimators for Planck parameters. These were performed on the aforementioned 0.14 sr patch of sky, for seven l bins: 100–150, 150–200, 200–280, 280–360, 360–440, 440–520, and 520–600. The results are shown in Fig. 5. The (Monte Carlo) mean of each estimator, computed from $n = 10$ trials of area 0.14 sr each, are shown. Note the similar performance of the estimators except in the low-l bands. Note that in the full Planck experiment (≈ 8 sr), the error bars would be smaller by a factor of $\approx \sqrt{1.4/8}$.

We may test both the linear and nonlinear estimators for bias using the t test. The t statistic for band α is given by

$$t[\alpha] = \frac{\text{SampleMean}(C_\alpha^{\Phi\Phi}) - C_\alpha^{\Phi\Phi}}{\sqrt{s^2/n}}, \tag{99}$$

where s^2 is the sample variance of the $C_\alpha^{\Phi\Phi}$. The t-test results are shown in Table II. A positive t statistic indicates that we are overestimating the power spectrum, a negative t statistic indicates that we are underestimating it. The t statistics here are designated t_9 in the table because they have 9 degrees of freedom. Also shown in the table is the two-tailed

TABLE II. Convergence power spectrum estimators: t test.

l range	d_α	linear t_9	linear p	nonlin. t_9	nonlin. p
100–150	127	-5.08	0.0007	-0.23	0.8200
150–200	178	-2.23	0.0527	0.79	0.4523
200–280	392	-2.16	0.0592	-3.66	0.0052
280–360	522	0.73	0.4822	-0.19	0.8558
360–440	653	0.17	0.8654	0.31	0.7670
440–520	783	-0.51	0.6238	-0.71	0.4976
520–600	914	0.49	0.6352	0.03	0.9763

p-value, i.e. the probability of a perfect t random variable (with 9 degrees of freedom) having absolute value exceeding t_9:

$$p = 2 \int_{|t_9|}^{\infty} \frac{\Gamma(n/2+1/2)}{\sqrt{n\,\pi}\,\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} dx. \quad (100)$$

If the power spectrum estimator is unbiased and normally distributed, the p value for each l bin is uniformly distributed between 0 and 1. (Warning: because, for each estimator, we derived the p values for all the l bins from the same 10 simulations, there is no reason to believe that the p values are independent.) We note that, for the high-l bins ($l > 200$), the linear and nonlinear estimators give similar results; both are consistent with being unbiased, although the nonlinear estimator shows a lower sample variance. (This is partially the result of negative power-spectrum estimates being set to zero by the nonlinear estimator.) In the two lowest-l bins, the sample variance of the nonlinear estimator is enormous. We note that in some of our simulations, the nonlinear estimator assigned anomalously large (in one case $> 4 \times 10^{-7}$) values of $C^{\kappa\kappa}$ to these bins; this suggests a problem with the estimator. This may be due to smearing of the bins by the finite width of the scanned region (width ≈ 0.38 sr) or may represent a problem with the iterative procedure (e.g. convergence to a local maximum of the likelihood).

Due to the excessive computation time requirements, we were unable to run a similar simulation of power spectrum estimators for the high-resolution experiment. Such a simulation would be interesting because the nonlinear lensing potential estimator showed improvements at the $\geqslant 10\%$ level over the quadratic estimator for this experiment.

VI. CONCLUSIONS

Weak lensing of CMB temperature maps has been recognized for some time as a potential probe for mapping the mass distribution of the universe (in projection), and determining quantities derivable from such a map: its power spectrum and cross-correlation with the CMB or other maps. In the past several years, methods for carrying out this statistical analysis have been proposed [1,7] and dramatically improved [2]. We have shown, by comparison to likelihood-based approaches, that quadratic estimator [2] for the lensing potential [equivalent to our Eq. (36)] is very close to optimal for the Planck experiment. That is, for this experiment, there

is no hope of further reduction in the statistical noise of the lensing potential. Similarly, simulations of the fully nonlinear power spectrum estimator do not show much improvement over the linear version.

If the lensing potential Φ can be treated as a Gaussian random field, and for experiments for which the linearized approximation suffices, then our maximum-likelihood analysis indicates that the CMB temperature bispectrum and trispectrum are optimal estimators for the temperature-lensing potential cross-correlation ($C^{\tilde\Theta\Phi}$) and lensing power spectrum, respectively. There is, in this case, no additional information in the fifth-order and higher statistics of the CMB. These higher-order statistics may be useful if the lensing potential is non-Gaussian; for example, the matter density bispectrum will result in a similar bispectrum in the quadratic estimator, Eq. (36), i.e. it its effect will be seen in the CMB six-point correlation function.

For the higher-resolution experiments, our results indicate that the residual error in the lensing potential maps can be reduced by switching from the quadratic estimator to a "full" likelihood-based estimator. In order to make this approach practical, further work will be needed to develop a version of the algorithm that works near the survey boundaries and to improve the stability of the algorithm. Other approaches to the functional integral in Eq. (60) besides the Gaussian approximation used here, such as Markov chains, could be used. It is even possible that, due to use of a better approximation to the integral, the error can be reduced further. However, given that the high-resolution (1 arcminute beam) simulation only showed improvement in RMS error at the 10–20 % level, it may be preferable to simply use the quadratic estimator for these experiments, accept this minor loss in signal-to-noise ratio, and avoid the difficulties associated with the nonlinear estimator.

One problem for both the quadratic estimator approach [2] and our likelihood-based approach are extremely sensitive to errors in the primary CMB power spectrum, $C_l^{\Theta\Theta}$. In the quadratic approach, this can be seen by noting that the power spectrum $C^{\Phi\Phi}$ is obtained by differencing two quantities which may be very close to each other in Eq. (72); the inverse Fisher matrix F_α^{-1} may be tens of times greater than $C_\alpha^{\Phi\Phi}$ for Planck (see Fig. 3) and consequently the quantities being subtracted in Eq. (72) may differ by only several percent. In the likelihood approach of Sec. V C this problem is masked by the formalism of the iterative scheme, but it is still there. Note that this problem is more serious for the lower-resolution experiments. It is apparent that a lensing power spectrum analysis must be accompanied by extremely accurate determination of $C_l^{\Theta\Theta}$, or an estimation scheme must be introduced that is robust against small errors in $C_l^{\Theta\Theta}$ must be introduced, or both.

We have performed no analysis here of lensing estimators using the CMB polarization. In Ref. [10] a quadratic estimator analysis for the CMB polarization has been performed. Because the polarization is a Gaussian random field whose covariance depends on the lensing potential (i.e. there are observed covariances $C^{\hat\Theta\hat E}$, $C^{\hat E\hat B}$, etc. analogous to the $C^{\hat\Theta\hat\Theta}$ used here), an analysis analogous to that of Secs. III and IV

establishes that the quadratic estimator approach is optimal if the lensing is sufficiently weak. Because our simulation code cannot handle polarization, we cannot determine whether realistic lensing is sufficiently weak or whether a nonlinear maximum-likelihood analysis is required to make full use of polarization data sets. Any high resolution polarization experiment will have one of its main goals gravity wave detection in B mode. Since weak lensing creates B modes out of E modes [15] it is important to remove this contamination as well as possible by using the weak lensing reconstruction [12,16]. Given that polarization and its E and B decomposition is sensitive to the direction of polarization in addition to its amplitude, it may be more susceptible to the errors induced by the linearization procedure. In this case the nonlinear analysis will be essential to exploit fully the potential of any future high resolution CMB polarization experiment.

APPENDIX: POTENTIALS, CONVERGENCE, AND PROJECTED DENSITY

Here we sketch a derivation of Eq. (6), and use this to relate the lensing potential Φ to quantities of more direct physical interest. We use a Robertson-Walker metric with a Newtonian perturbation (i.e. a weak perturbation $|\Psi| \ll 1$ induced by nonrelativistic matter):

$$ds^2 = a^2[-(1-2\Psi)d\tau^2 + (1+2\Psi)(dr^2 + S(r)^2 d\omega^2)]$$
(A1)

where a is a function of the conformal time τ, and Ψ is the gravitational potential, generally a function of all the coordinates. The comoving distance is r, and $\omega \in S^2$ is a direction on the unit sphere with the usual line element $d\omega^2$. We have used the sinelike function $S(r) = k^{-1/2}\sin(k^{1/2}r)$, and will use its derivative, the cosinelike function $C(r) = \cos(k^{1/2}r)$, and their ratio $T(r) = S(r)/C(r)$, where k is the spatial curvature. We use as an initial condition $\tau = 0$ at present, and normalize $a(\tau=0) = 1$. The simplest way to find the photon deflection is to consider the conformal metric (which must have exactly the same null geodesics):

$$d\tilde{s}^2 = -(1-4\Psi)d\tau^2 + dr^2 + S(r)^2 d\omega^2.$$
(A2)

In this metric we compute for the null geodesics (to linear order in Ψ and assuming that the geodesic is nearly radial):

$$\frac{d^2\omega}{dr^2} = \frac{C(r)}{S(r)}\frac{d\omega}{dr} - 2\frac{\partial\Psi}{\partial\omega}.$$
(A3)

In the Born approximation, where the gradient over ω is evaluated along the unperturbed line of sight, this is an inhomogeneous linear equation in **n** which can be solved by the Green's function method to find ω at the last scattering surface. The result is that the null geodesic arriving at "us" ($r = \tau = 0$) from direction **n** is found to have originated in direction $\mathbf{n} + \nabla\Phi(\mathbf{n})$, where

$$\Phi(\mathbf{n}) = -2\int_0^{r_{ls}} dr \Psi(r\mathbf{n}, -r)\left(\frac{1}{T(r)} - \frac{1}{T(r_{ls})}\right).$$
(A4)

Through the use of the trigonometric identities and their hyperbolic counterparts we can show that:

$$\frac{1}{T(r)} - \frac{1}{T(r_{ls})} = \frac{S(r_{ls}-r)}{S(r)S(r_{ls})}$$
(A5)

with which Eq. (A4) can be shown to be equivalent to the forms provided by, e.g. Refs. [2,3].

Since the lensing potential Φ is represented here as a projected gravitational potential, it would make sense that its second derivative $\kappa = -\frac{1}{2}\nabla^2\Phi$ would represent a projected density perturbation. This is indeed the case, although there are other contributions to κ. If we define Δ to be the comoving three-dimensional Laplacian [i.e. on $dr^2 + S(r)^2 d\omega^2$], as distinguished from the two-dimensional Laplacian ∇^2 on the unit sphere, we have the usual relation for Δ:

$$\Delta = \frac{1}{S(r)^2}\left[\nabla^2 + \frac{\partial}{\partial r}\left(S(r)^2\frac{\partial}{\partial r}\right)\right].$$
(A6)

If we solve this relation for ∇^2, we can split $\kappa = -\frac{1}{2}\nabla^2\Phi$ into two terms: one involving the Δ operator and one involving the radial operator:

$$\kappa = \int_0^{r_{ls}} dr\left(\frac{1}{T(r)} - \frac{1}{T(r_{ls})}\right)\left[S(r)^2\Delta\Psi - \frac{\partial}{\partial r}\left(S(r)^2\frac{\partial\Psi}{\partial r}\right)\right].$$
(A7)

The first term can be replaced with a density using Poisson's equation, thus generating a projected density. To study the second term, we replace the partial derivative over r (at constant τ) with a total derivative along the line of sight and a time derivative. The time derivative can be neglected here if the matter is nonrelativistic. [Indeed we have already made this assumption implicitly when we write Eq. (A1).] Next integrate by parts so that the d/dr acts on $1/T(r) - 1/T(r_{ls})$. (Since $S(0) = 0$, the surface terms generated by the integration by parts will vanish.) Then we use the identity $(d/dr)[1/T(r)] = -1/S(r)^2$ to convert Eq. (A7) into

$$\kappa(\mathbf{n}) = 4\pi G_N \int_0^{r_{ls}} dr\left(\frac{1}{T(r)} - \frac{1}{T(r_{ls})}\right)$$
$$\times S(r)^2 a(-r)^2 \delta\rho(r\mathbf{n}, -r) - \Psi(r_{ls}\mathbf{n}, -r_{ls}) + \Psi(0)$$
(A8)

where G_N is the universal gravitation constant and $\delta\rho$ is the density perturbation. Note that convergence can be broken into two components: a component due to the density fluctuations $\delta\rho$ along the line of sight, and a component due to the potential difference between the source and observer. The second component is due to tidal forces acting to separate the trajectories of CMB photons; conceptually, it has the same origin as the compression of the sky into a small solid angle near the zenith as seen by an observer near a black hole despite the absence of any mass energy along the line of sight.

Finally, we attempt to determine the magnitude of the potential-difference contribution to the convergence. Here $\Psi(0)$ is a constant (isotropic) and can be removed by a gauge transformation, so we do not consider it further. The term $\Psi(r_{ls})$ can be estimated based on the CMB temperature fluctuation using the Sachs-Wolfe relation $\Psi(r_{ls}) = 3\Theta$ [17]; thus the contribution of the $\Psi(r_{ls})$ term to κ is -3Θ. Since $C_l^{-3\Theta,-3\Theta} = 9C_l^{\Theta\Theta}$ is always at least a factor of l^2 less than the overall power spectrum $C_l^{\kappa\kappa}$, its effect on the convergence power spectrum is subdominant with respect to cosmic variance.

[1] U. Seljak and M. Zaldarriaga, Phys. Rev. Lett. **82**, 2636 (1999).

[2] W. Hu, Astrophys. J. Lett. **557**, L79 (2001).

[3] W. Hu, Phys. Rev. D **62**, 043007 (2000).

[4] M. Zaldarriaga and U. Seljak, Phys. Rev. D **59**, 123507 (1999).

[5] K. Benabed, F. Bernardeau, and L. van Waerbeke, Phys. Rev. D **63**, 043501 (2001).

[6] D.M. Goldberg and D.N. Spergel, Phys. Rev. D **59**, 103002 (1999).

[7] U. Seljak and M. Zaldarriaga, Phys. Rev. D **60**, 043504 (1999).

[8] W. Hu, Phys. Rev. D **64**, 083005 (2001).

[9] M. Zaldarriaga, Phys. Rev. D **62**, 063510 (2000).

[10] W. Hu and T. Okamoto, Astrophys. J. **574**, 566 (2002).

[11] S.P. Oh, D.N. Spergel, and G. Hinshaw, Astrophys. J. **510**, 551 (1999).

[12] L. Knox and Y. Song, Phys. Rev. Lett. **89**, 011303 (2002).

[13] L. Verde and D.N. Spergel, Phys. Rev. D **65**, 043007 (2002).

[14] L. Knox, Phys. Rev. D **52**, 4307 (1995).

[15] M. Zaldarriaga and U. Seljak, Phys. Rev. D **58**, 023003 (1998).

[16] M. Kesden, A. Cooray, and M. Kamionkowski, Phys. Rev. Lett. **89**, 011304 (2002).

[17] R.K. Sachs and A.M. Wolfe, Astrophys. J. **147**, 73 (1967).