Reconstruction of lensing from the cosmic microwave background polarization

Christopher M. Hirata* and Uroš Seljak[†]

Department of Physics, Jadwin Hall, Princeton University, Princeton, New Jersey 08544, USA

Gravitational lensing of the cosmic microwave background (CMB) polarization field has been recognized as a potentially valuable probe of the cosmological density field. We apply likelihood-based techniques to the problem of lensing of CMB polarization and show that if the B-mode polarization is mapped, then likelihood-based techniques allow significantly better lensing reconstruction than is possible using the previous quadratic estimator approach. With this method the ultimate limit to lensing reconstruction is not set by the lensed CMB power spectrum. Second-order corrections are known to produce a curl component of the lensing deflection field that cannot be described by a potential; we show that this does not significantly affect the reconstruction at noise levels greater than 0.25 μK arcmin. The reduction of the mean squared error in the lensing reconstruction relative to the quadratic method can be as much as a factor of two at noise levels of 1.4 μK arcmin to a factor of ten at 0.25 μK arcmin, depending on the angular scale of interest.

I. INTRODUCTION

Over the past decade the cosmic microwave background (CMB) anisotropy has been established as a robust and powerful cosmological probe. While much attention has focused on the primary anisotropy generated in the early Universe, the CMB should also contain signatures of processes that occurred between the surface of last scatter and the present. One of these is weak gravitational lensing, which has been recognized as a probe of large scale structure (LSS) [1–5]. Aside from its use as a probe of the matter power spectrum at low redshift $z \ll 1100$, weak lensing of the CMB could be cross correlated against other tracers of the density field such as galaxy surveys [6] or weak lensing of galaxies [5]. Through cross-correlation with the CMB temperature, an improved measurement of the integrated Sachs-Wolfe effect over that possible using the CMB power spectrum alone is possible, yielding constraints on the late-time growth function and hence on the dark energy [7,8]. Lensing has also attracted attention recently as a cosmological source of B-mode polarization [9]; reconstruction and removal of lensing B modes will thus be an important part of a future search for B-mode polarization induced by primordial gravitational waves [10–12].

The lensing signal in the CMB is small, so it is important to construct optimal methods for estimating the lensing field from CMB data. The early investigations of lensing of the CMB temperature showed that while there is an effect of lensing on the CMB power spectrum [13,14], it is much more promising to estimate the lensing field using quadratic combinations of the CMB temperature, and to estimate the lensing power spectrum using the four-point correlation function (or its harmonic equivalent, the trispectrum) [1,3,15]. More recent work has identified the divergence of the temperature-weighted gradient as the optimal quadratic combination of the CMB temperature for use in lensing stud-

ies [4,16]. Analysis based on likelihood techniques [17] has since shown that the quadratic estimator technique is statistically optimal when the lensing effect on the CMB covariance matrix is small. This was shown to be a good approximation for lensing of temperature anisotropies in the range $l \lesssim 3500$. For the small scales $l \gtrsim 3500$, the primary CMB power spectrum is much smaller than the lensed power spectrum, hence this argument breaks down. In this case for sufficiently small instrument noise the reconstruction of projected mass density can be nearly perfect [18]. We will not discuss the reconstruction on these very small scales in this paper.

Our ability to reconstruct the lensing field using the CMB temperature is limited because the temperature fluctuations are stochastic and so we can only statistically determine the unlensed CMB temperature field. It is thus advantageous to consider lensing of the CMB polarization, since in the absence of primordial gravitational waves the unlensed CMB polarization is entirely in E rather than B modes. This implies that, in the terminology of galaxy lensing, there is no "shape noise" in the CMB polarization field. Several authors have developed algorithms that use the B modes induced by lensing to probe LSS [5,19]. The optimal quadratic estimator — the polarization analogue of the temperature-based quadratic estimator using the divergence of the temperature-weighted gradient — was constructed by Ref. [10]. There it was shown that for sufficiently small detector noise most of the lensing reconstruction information with this method is provided by the B mode.

Even with polarization information these quadratic estimators cannot improve the reconstruction beyond a certain level, set by the coherence length of the polarization. It has been argued that this provides a fundamental limit to our ability to separate the lensing induced B modes from the B modes induced by gravity waves [11,12]. However, it has not been determined whether quadratic estimation is optimal for the polarization-based lensing reconstruction, and indeed Refs. [11,12] comment that it might be possible to extract additional information in higher-order statistics. The argument for optimality of the quadratic estimator presented by

---

*Email address: chirata@princeton.edu

[†]Email address: useljak@princeton.edu

TABLE I. Parameters for CMB experiments.

Experiment	$\mathcal{N}_P/\mu K$ arcmin	θ_{FWHM}/arcmin	l_c
WMAP, 4 yrs (94 GHz)	700	13	620
Planck, 1 yr (143 GHz)	81	8	1010
Ref. expt. A	3.0	7	1160
Ref. expt. B	1.41	7	1160
Ref. expt. C	1.41	4	2020
Ref. expt. D	1.00	4	2020
Ref. expt. E	0.50	2	4050
Ref. expt. F	0.25	2	4050

Ref. [17] does not apply to polarization since the *B*-mode power is dramatically increased by lensing. Here we construct likelihood-based estimators for lensing using the CMB polarization and show that the likelihood-based estimator improves significantly on the quadratic estimator (although we do not present these as series of higher-order statistics). Indeed, as noise is decreased the accuracy of CMB lensing reconstruction continues to improve without bound. Conceptually this is because if the lensed polarization is measured with zero noise, then the equation $B_{\text{unlensed}} = 0$ can be solved (except possibly for a small number of degenerate modes) for the projected matter density with zero noise. The equation $B_{\text{unlensed}} = 0$ is ill-behaved in the presence of instrument noise; fortunately, the likelihood formalism easily incorporates noise and, as we show in this paper, regularizes the problem.

In practice, a perfect reconstruction of the lensing potential is impossible because as instrument noise is reduced, some contaminant to the lensing signal will eventually become more important than the instrument noise. One candidate for this limiting factor is lensing field rotation caused by the fact that the density perturbations causing the lensing are spread out along the line of sight (i.e. there is more than one "lens plane") and that the lensing is not perfectly weak (i.e. the first-order Born approximation to the lensing field is inexact). We will show that even for an experiment with noise 0.25 μK arcmin and 2 arcmin full-width half maximum beam, the field rotation does not substantially worsen the lensing reconstruction. It is however possible that foreground contamination will be a more serious problem.

Studies of lensing of CMB polarization will require that the polarization field be mapped with noise levels of the order of ~ 1 μK arcmin; this would be a substantial improvement in sensitivity beyond that of the current *Wilkinson Microwave Anisotropy Probe* (*WMAP*; http://map.gsfc.nasa.gov/) and the upcoming *Planck* (http://astro.estec.esa.nl/Planck/) experiments (see Table I). Nevertheless, ~ 1 μK arcmin may be achieved with a future polarization satellite. The noise levels of < 0.25 μK arcmin at which field rotation becomes important will probably remain unachievable for the foreseeable future.

This paper is organized as follows: in Sec. II, we define our notations and conventions. In Sec. III, we consider the properties of the likelihood function and its implications for likelihood and Bayesian analyses of CMB lensing recon-

struction. In Sec. IV, we investigate the breakdown of the Born approximation for CMB lensing and its implications for lensing reconstruction. Determination of the lensing power spectrum from CMB maps is discussed in Sec. V. We show numerical simulations of CMB polarization lensing and reconstruction using our estimators in Sec. VI, and conclude in Sec. VII.

The fiducial cosmology used in these simulations is a spatially flat cosmological constant-dominated universe with baryon fraction $\Omega_{b0} = 0.046$; cold dark matter fraction $\Omega_{c0} = 0.224$; cosmological constant fraction $\Omega_{\Lambda 0} = 0.73$; Hubble parameter $H_0 = 72$ km/s/Mpc; primordial helium abundance $Y_P = 0.24$; reionization optical depth $\tau_r = 0.17$; primordial scalar spectral index $n_s = 1$; and no primordial vector or tensor perturbations. We have used the CMBFAST numerical package [20] to compute all power spectra except in Sec. IV B. The experiments considered are as shown in Table I. The *WMAP* and the *Planck* will not be able to map the lensing field using polarization and are included in the table for comparison. The reference experiments A through F are successively lower-noise (or finer-beam) experiments that were analyzed to determine how the signal-to-noise ratio in the lensing reconstruction depends on experimental parameters. Note that experiment C is the reference experiment of Refs. [10,11].

II. FORMALISM

Here we describe our normalization conventions; note that for some quantities, there are many conventions in use in the literature, and appropriate conversion factors must be applied if one wishes to compare results.

A. CMB

We work in the normalized flat-sky approximation, i.e. the sky is taken to be a flat square of side length $\sqrt{4\pi}$ (i.e. total area 4π) with periodic boundary conditions. The CMB temperature and polarization fields can then be expressed as a sum over Fourier modes:

$$\begin{pmatrix} T(\hat{\mathbf{n}}) \\ Q(\hat{\mathbf{n}}) \\ U(\hat{\mathbf{n}}) \end{pmatrix} = \frac{1}{\sqrt{4\pi}} \sum_{\mathbf{l}} \begin{pmatrix} T_{\mathbf{l}} \\ Q_{\mathbf{l}} \\ U_{\mathbf{l}} \end{pmatrix} e^{i\mathbf{l}\cdot\hat{\mathbf{n}}}, \qquad (1)$$

where the **l** modes are distributed in the two-dimensional **l** space with number density $1/\pi$. Defining the angle of a mode by $\tan\phi_{\mathbf{l}} = l_y/l_x$, we have *E* and *B* polarization modes given by

$$\begin{pmatrix} E_{\mathbf{l}} \\ B_{\mathbf{l}} \end{pmatrix} = \begin{pmatrix} \cos(2\phi_{\mathbf{l}}) & \sin(2\phi_{\mathbf{l}}) \\ -\sin(2\phi_{\mathbf{l}}) & \cos(2\phi_{\mathbf{l}}) \end{pmatrix} \begin{pmatrix} Q_{\mathbf{l}} \\ U_{\mathbf{l}} \end{pmatrix}. \qquad (2)$$

(Technically the angle ϕ_0 of the $\mathbf{l} = 0$ mode is undefined, however this will not concern us since within the flat-sky approximation we will convert sums over **l** into integrals: $\Sigma_{\mathbf{l}} \rightarrow \int d^2\mathbf{l}/\pi$. If an integral is divergent at $\mathbf{l} = 0$, then it cannot be computed accurately within the flat-sky approximation.)
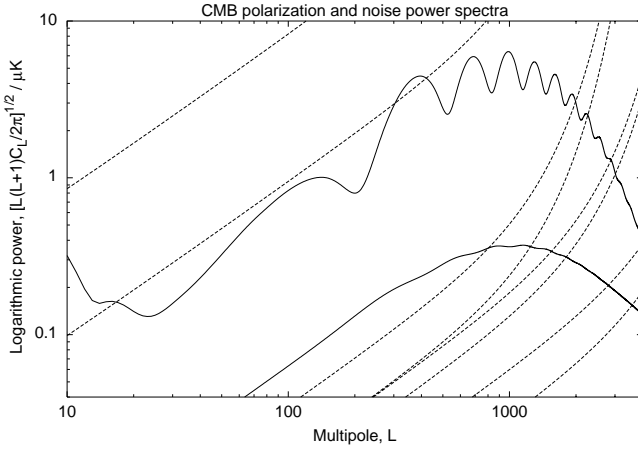
FIG. 1. CMB polarization power spectra for *E*-type and *B*-type polarization (upper and lower solid curves, respectively). The noise curves for the experiments of Table I are shown as dashed lines; from top to bottom: *WMAP*, *Planck*, and reference experiments A, B, C, D, E, and F.

We will use the following notations for CMB fields: $\{T, Q, U\}$ for the unlensed (primary) CMB anisotropies; $\{\tilde{T}, \tilde{Q}, \tilde{U}\}$ for the lensed CMB anisotropies; and $\{\hat{T}, \hat{Q}, \hat{U}\}$ for the measured anisotropies (including noise). These are measured in μK (blackbody temperature), and we will assume that the monopole (mean temperature) and special-relativistic effects (kinematic dipole or quadrupole and stellar aberration) have been removed. The instrument noise will be assumed to be statistically uncorrelated with any cosmological signal and will be denoted by η_X, where X is one of T, Q, or U (or T, E, and B depending on which basis is more convenient). The unlensed CMB will have a power spectrum given by

$$\langle X_{\mathbf{l}}^* X_{\mathbf{l}'}' \rangle = C_l^{XX'} \delta_{\mathbf{l},\mathbf{l}'} , \qquad (3)$$

where here X and X' are T, E, or B (here we desire rotational symmetry so we cannot use Q or U). We assume the universe is statistically parity-invariant so that $C_l^{TB} = C_l^{EB} = 0$; in some parts of this paper we will discuss universes with no tensor perturbations, in which case we also have $C_l^{BB} = 0$.

Throughout most of this paper we will take general noise covariance $\mathbf{N}$; when we wish to show expected performance for particular experiments, we will use the following noise power spectrum appropriate for a Gaussian beam profile:

$$N_l^{TT} = \mathcal{N}_T^2 e^{l(l+1)\theta_{FWHM}^2/8\ln 2} = \mathcal{N}_T^2 e^{l(l+1)/l_c(l_c+1)}, \qquad (4)$$

where θ_{FWHM} is the full width at half maximum (FWHM) of the beam. We take a similar form for $N_l^{EE} = N_l^{BB}$, except that $\mathcal{N}_T$ is replaced with $\mathcal{N}_P$. The quantities $\mathcal{N}_T$, $\mathcal{N}_P$, and θ_{FWHM} (combined with the fraction f_{sky} of the sky surveyed) thus parametrize the performance of the experiment. Noise curves compared to the CMB for the experiments shown in Table I are shown in Fig. 1. The l value at which the beam

transfer function drops to $1/\sqrt{e}$ is given approximately by $l_c \approx (8095 \text{ arcmin})/\theta_{FWHM}$.

We will introduce a vector $\mathbf{x}$ containing both temperature and polarization information: $\mathbf{x} = (T, Q, U)$. The lensed and measured temperature or polarization vectors will be denoted $\tilde{\mathbf{x}}$ and $\hat{\mathbf{x}}$, respectively. [Note: $\mathbf{x}$ is a "vector" in the sense of linear algebra, i.e. it is an element of a vector space, in this case a Hilbert space with the usual $L^2(S^2)$ inner product, on which matrix operations such as $\mathbf{C}$ can act. It is not a vector in the sense of differential geometry.] Since most of the fields we deal with, including CMB temperature and polarization, are real, their Fourier modes satisfy e.g. $T_{\mathbf{l}} = T_{-\mathbf{l}}^*$. Consequently, if we have N Fourier modes, the N-dimensional vector with components $\{T_{\mathbf{l}}\}$ only has $N/2$ independent complex components; the remainder contain redundant information. (Of course, there are still N independent real components.) The covariance matrix $\mathbf{C}$ is defined as $\mathbf{C}^{\hat{\mathbf{x}}\hat{\mathbf{x}}} = \langle \hat{\mathbf{x}}\hat{\mathbf{x}}^\dagger \rangle$; note that it is Hermitian by construction.

B. Lensing

The lensed temperature and polarization are given in terms of the unlensed temperature by means of the remapping function g: $\tilde{T}(\hat{\mathbf{n}}) = T[g(\hat{\mathbf{n}})]$, and similarly for Q and U [however, E and B do not transform this way, rather one must use Eq. (2)]. The remapping function can be decomposed into a longitudinal part generated by the lensing potential Φ, and a transverse part generated by the lensing cross-potential Ω:

$$g(\hat{\mathbf{n}}) = \hat{\mathbf{n}} + \nabla\Phi(\hat{\mathbf{n}}) + \star\nabla\Omega(\hat{\mathbf{n}}), \qquad (5)$$

where $\star$ is the ninety-degree rotation operator: $\star\hat{\mathbf{e}}_x = \hat{\mathbf{e}}_y$, $\star\hat{\mathbf{e}}_y = -\hat{\mathbf{e}}_x$. Past studies of CMB lensing reconstruction have ignored the cross-potential since (for scalar perturbations) it vanishes at first order in perturbation theory. In principle it could become important given the high precision enabled by lensing of CMB polarization. However, we will show in Sec. IV and Sec. VI A that the cross-potential is unimportant for most near-term experiments.

We restrict our attention to the weak lensing regime, i.e. we assume the magnification matrix:

$$
\mathbf{M} = \frac{\partial g(\hat{\mathbf{n}})}{\partial \hat{\mathbf{n}}} = \begin{pmatrix} \dfrac{\partial g_x}{\partial x} & \dfrac{\partial g_x}{\partial y} \\[2mm] \dfrac{\partial g_y}{\partial x} & \dfrac{\partial g_y}{\partial y} \end{pmatrix}
$$

$$
= \begin{pmatrix} 1 + \partial_x^2\Phi - \partial_x\partial_y\Omega & \partial_x\partial_y\Phi - \partial_y^2\Omega \\[2mm] \partial_x\partial_y\Phi + \partial_x^2\Omega & 1 + \partial_y^2\Phi + \partial_x\partial_y\Omega \end{pmatrix}
$$

$$
= \begin{pmatrix} 1 + \kappa + \gamma_Q & \gamma_U + \omega \\[2mm] \gamma_U - \omega & 1 + \kappa - \gamma_Q \end{pmatrix} \qquad (6)
$$

is everywhere invertible. This is a necessary and sufficient condition to disallow caustics and multiple images of the same portion of the surface of last scatter. Lensing by large-scale structure is too weak to create caustics on the surface of

last scatter, so the weak lensing assumption is violated only in the vicinity of astrophysical objects such as clusters. We classify clusters and other strong lenses as foreground contaminants and do not consider them further here. In the regime where the lensing distortion is small—i.e. $\mathbf{M}$ is close to the identity—we may interpret the four components $\kappa, \gamma_Q, \gamma_U, \omega$ in Eq. (6) as follows. The convergence κ magnifies a feature on the last-scattering surface of infinitesimal angular size $d\vartheta$ to size $(1+\kappa)d\vartheta$. The field rotation angle ω rotates the feature clockwise by ω radians. The Q-shear γ_Q produces a stretching along the x axis while compressing it along y: the apparent angular extents of a feature along the two axes are $(1+\gamma_Q)d\vartheta_x$ and $(1-\gamma_Q)d\vartheta_y$, respectively. The U-shear has a similar effect, stretching along the $y=x$ axis and compressing along the $y=-x$ axis. It should be noted that the four fields $\kappa, \gamma_Q, \gamma_U, \omega$ are *not* independent because they are all generated by differentiating the two fields Φ and Ω. In particular, if we *EB*-decompose the shear field into its positive-parity (ϵ) and negative-parity (β) components:

$$\epsilon_{\mathbf{l}} = [\gamma_Q]_{\mathbf{l}} \cos 2\phi_{\mathbf{l}} + [\gamma_U]_{\mathbf{l}} \sin 2\phi_{\mathbf{l}},$$

$$\beta_{\mathbf{l}} = -[\gamma_Q]_{\mathbf{l}} \sin 2\phi_{\mathbf{l}} + [\gamma_U]_{\mathbf{l}} \cos 2\phi_{\mathbf{l}}, \qquad (7)$$

we find that $\epsilon_{\mathbf{l}} = \kappa_{\mathbf{l}}$ and $\beta_{\mathbf{l}} = \omega_{\mathbf{l}}$. These are then related to the potentials via $\kappa_{\mathbf{l}} = (l^2/2)\Phi_{\mathbf{l}}$ and $\omega_{\mathbf{l}} = (l^2/2)\Omega_{\mathbf{l}}$. This immediately implies the power spectrum relations $C_l^{\kappa\kappa} = C_l^{\epsilon\epsilon} = \frac{1}{4}l^4 C_l^{\Phi\Phi}$ and $C_l^{\omega\omega} = C_l^{\beta\beta} = \frac{1}{4}l^4 C_l^{\Omega\Omega}$. It is of interest to note that the convergence and field rotation can be determined from the deflection angle $\mathbf{d}(\hat{\mathbf{n}}) = g(\hat{\mathbf{n}}) - \hat{\mathbf{n}}$ by

$$\kappa = -\frac{1}{2}\nabla \cdot \mathbf{d} \quad \text{and} \quad \omega = \frac{1}{2}\nabla \cdot \star \mathbf{d}. \qquad (8)$$

If $\kappa, \gamma_Q, \gamma_U, \omega$ are not small compared to 1, then the physical interpretation of these quantities is somewhat more complicated. We will continue to call κ the "convergence," (γ_Q, γ_U) the "shear," and ω the "field rotation angle" even in this case, although this convention is not universally followed in the literature. Note, however, that the relations $\epsilon_{\mathbf{l}} = \kappa_{\mathbf{l}}$ and $\beta_{\mathbf{l}} = \omega_{\mathbf{l}}$ continue to hold (in fact, they remain valid even for strong lenses!), which makes our definitions of convergence, shear, and field rotation angle particularly convenient.

Sometimes we will specify a lens remapping g by its convergence and field rotation, $g \equiv (\kappa, \omega)$. Most authors have performed the lensing analysis using Φ rather than κ as the field to be estimated, since the deflection angle is a local function of the former. In the present analysis, we take κ (and ω when it is important) to be the fundamental field. Of course, the two fields contain exactly the same information, being related by the differential relation $\kappa = -\frac{1}{2}\nabla^2\Phi$ in real space and by a multiplicative factor of $l^2/2$ in harmonic space.

It is convenient to introduce the lensing operator Λ_g defined by $\Lambda_g X(\hat{\mathbf{n}}) = X(g(\hat{\mathbf{n}}))$, where X is one of T, Q, or U. [In the $\{T, E, B\}$ basis, the action of Λ_g is not so simple and the transformation of Eq. (2) must be applied.] We define

the σ^κ differential operator (and analogously σ^ω) as the action of an infinitesimal lens configuration: $\sigma_{\mathbf{l}}^\kappa \equiv (\partial\Lambda_g/\partial\kappa_{\mathbf{l}}[g])|_{g=(0,0)}$. Then, noting that $\Lambda_{(0,0)}$ is the identity operator $\mathbf{1}$, we find that to first order in (κ, ω):

$$\Lambda_{(\kappa,\omega)} = \mathbf{1} + \sum_{\mathbf{l}} (\kappa_{\mathbf{l}}\sigma_{\mathbf{l}}^\kappa + \omega_{\mathbf{l}}\sigma_{\mathbf{l}}^\omega) + O(\kappa^2, \omega^2, \kappa\omega). \qquad (9)$$

The σ operator acts as follows on a field X in the $\{T, Q, U\}$ basis:

$$(\sigma_{\mathbf{l}}^\kappa X)_{\mathbf{l}'} = \left(\frac{2}{l^2}\right) \frac{\mathbf{l}\cdot(\mathbf{l}'-\mathbf{l})}{\sqrt{4\pi}} X_{\mathbf{l}'-\mathbf{l}},$$

$$(\sigma_{\mathbf{l}}^\omega X)_{\mathbf{l}'} = \left(\frac{2}{l^2}\right) \frac{\mathbf{l}\cdot\star(\mathbf{l}'-\mathbf{l})}{\sqrt{4\pi}} X_{\mathbf{l}'-\mathbf{l}}. \qquad (10)$$

In the $\{T, E, B\}$ basis, the σ-matrices mix E and B because these are nonlocal quantities. Specifically, they have components

$$[\sigma_{\mathbf{l}}^\kappa]_{\mathbf{l}_1, -\mathbf{l}_2} = -\frac{\delta_{\mathbf{l}_1, \mathbf{l}-\mathbf{l}_2}}{\sqrt{4\pi}} \left(\frac{2}{l^2}\right) (\mathbf{l}\cdot\mathbf{l}_2) \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos 2\alpha & -\sin 2\alpha \\ 0 & \sin 2\alpha & \cos 2\alpha \end{pmatrix}, \qquad (11)$$

where the rows correspond to the $\{T, E, B\}_{\mathbf{l}_1}$ and the columns to the $\{T, E, B\}_{-\mathbf{l}_2}$, and we have defined the angle $\alpha = \phi_{\mathbf{l}_1} - \phi_{\mathbf{l}_2}$. The matrix for $\sigma_{\mathbf{l}}^\omega$ differs by replacing the prefactor $\mathbf{l}\cdot\mathbf{l}_2$ with $\star\mathbf{l}\cdot\mathbf{l}_2$. The σ matrices satisfy $\sigma_{\mathbf{l}}^\kappa = (\sigma_{-\mathbf{l}}^\kappa)^\dagger$.

Lensing alters the CMB anisotropy covariance; the covariance matrix $\mathbf{C}^{\tilde{X}\tilde{X}'}$, where $X \in \{T, Q, U\}$ (or $X \in \{T, E, B\}$) of the lensed temperatures is dependent on the lens configuration g, and thus we will denote it by $\tilde{\mathbf{C}}_g^{XX'} \equiv \mathbf{C}_g^{\tilde{X}\tilde{X}'}$. Since the lensed CMB field is $\tilde{\mathbf{x}} = \Lambda_g\mathbf{x}$, we have $\tilde{\mathbf{C}}_g = \Lambda_g\mathbf{C}\Lambda_g^\dagger$. The lensed covariance averaged over the ensemble of LSS configurations will be denoted by $\langle\tilde{\mathbf{C}}^{XX'}\rangle_{LSS}$. Note, however, that whereas the primary CMB is expected to be nearly Gaussian, the lensed CMB is non-Gaussian and so $\langle\tilde{\mathbf{C}}^{XX'}\rangle_{LSS}$ does *not* specify completely the statistics of the lensed CMB field. Indeed, it is the non-Gaussianity of the lensed CMB that enables separation of the lensing and gravitational wave contributions to B. It also means that the standard Gaussian formula for the uncertainty in the power spectrum, $\sigma(C_l)/C_l = \sqrt{2/(2l+1)f_{sky}\Delta l}$, does not necessarily apply to B-mode polarization on small scales.

It is readily apparent from Eq. (11) that lensing can produce B modes even if these are not present in the primary CMB. We show in Appendix B that for "almost all" primary CMB realizations, there are only a small number of convergence modes that do not produce B-type polarization.

C. Chi-squared analysis of lensing

We illustrate our formalism with a simple lensing reconstruction via χ^2 minimization (the "least squares" method).

We perform a full likelihood analysis in Sec. III, but the χ^2 analysis is sufficiently similar that it illustrates the basic concept. Define the functional $\chi^2(\kappa)$ of a convergence field given CMB data $\hat{\mathbf{x}}$:

$$\chi^2(\kappa) \equiv (\Lambda_{(\kappa,0)}^{-1}\hat{\mathbf{x}})^\dagger(\mathbf{C}+\mathbf{N})^{-1}\Lambda_{(\kappa,0)}^{-1}\hat{\mathbf{x}} + \kappa^\dagger\mathbf{C}^{\kappa\kappa-1}\kappa. \quad (12)$$

Here $\Lambda_{(\kappa,0)}^{-1}\hat{\mathbf{x}}$ is the delensed CMB, i.e. we have taken the measured CMB and projected it back onto the primary CMB assuming that the lens configuration is given by convergence κ with no rotation. To a first approximation, this should have covariance $\mathbf{C}+\mathbf{N}$ since it is the sum of primary CMB and instrument noise. (The matrix $\mathbf{C}+\mathbf{N}$ is equal to the measured CMB covariance in the absence of lensing and hence will frequently be denoted by $\hat{\mathbf{C}}_{(0,0)}$. Technically the noise covariance is not exactly $\mathbf{N}$ because the noise has been delensed; see Sec. III B.) We have thus chosen to define our χ^2 as the amount of power in this de-lensed CMB, with the various modes weighted according to their variance. The addition of the $\kappa^\dagger\mathbf{C}^{\kappa\kappa-1}\kappa$ term serves to regularize the problem by preventing the convergence from running off to ∞ in search of smaller primary CMB power.

If we take the first-order approximation to Λ^{-1} given by Eq. (9), Eq. (12) becomes

$$\chi^2(\kappa) = \chi^2(0) + 2\sum_{\mathbf{l}} m_{\mathbf{l}}^*\kappa_{\mathbf{l}} + \sum_{\mathbf{l},\mathbf{l}'} \kappa_{\mathbf{l}}^*(A_{\mathbf{l},\mathbf{l}'}$$
$$+ C_l^{\kappa\kappa-1}\delta_{\mathbf{l},\mathbf{l}'})\kappa_{\mathbf{l}'}, \quad (13)$$

where

$$\chi^2(0) = \hat{\mathbf{x}}^\dagger(\mathbf{C}+\mathbf{N})^{-1}\hat{\mathbf{x}},$$

$$m_{\mathbf{l}} = \hat{\mathbf{x}}^\dagger(\mathbf{C}+\mathbf{N})^{-1}\sigma_{-\mathbf{l}}^\kappa\hat{\mathbf{x}},$$

$$A_{\mathbf{l},\mathbf{l}'} = \hat{\mathbf{x}}^\dagger\sigma_{-\mathbf{l}}^\kappa(\mathbf{C}+\mathbf{N})^{-1}\sigma_{\mathbf{l}'}^\kappa\hat{\mathbf{x}}. \quad (14)$$

Note that $\mathbf{m}$ is a real vector and $\mathbf{A}$ is Hermitian. This is a quadratic function of κ and hence it has a minimum that can be determined via standard techniques. The minimum is at

$$\kappa_* = (\mathbf{A}+\mathbf{C}^{\kappa\kappa-1})^{-1}\mathbf{m}. \quad (15)$$

The error covariance $\mathbf{S}_{\chi^2}$ of κ is found by the usual method of setting $\chi^2(\kappa) = \chi^2(\kappa_*) + (\kappa-\kappa_*)^\dagger\mathbf{S}_{\chi^2}^{-1}(\kappa-\kappa_*)$; this yields

$$\mathbf{S}_{\chi^2} = (\mathbf{A}+\mathbf{C}^{\kappa\kappa-1})^{-1}. \quad (16)$$

The most important feature of this analysis is the reconstruction error, $\mathbf{S}_{\chi^2}$. Note that as the instrument noise goes to zero, the matrix $\mathbf{C}+\mathbf{N}$ develops null directions corresponding to the B modes. Therefore, $(\mathbf{C}+\mathbf{N})^{-1}$ has infinite eigenvalues in these directions, and if the number of convergence modes being reconstructed is less than or equal to the number of B modes measured, we have $\mathbf{A}\rightarrow\infty$ and $\mathbf{S}_{\chi^2}\rightarrow 0$. This leads us to the conclusion that the accuracy of convergence reconstruction is limited only by the sensitivity of the instru-

ment and the presence of foregrounds or other contaminants, not by statistics of the convergence or primary CMB field. One can note that for zero instrument noise, the χ^2, Eq. (12), is infinite unless the delensed CMB field $\Lambda_{(\kappa,0)}^{-1}\hat{\mathbf{x}}$ has vanishing B modes, i.e. in this case the χ^2 analysis is solving for $B_{\text{unlensed}}=0$ (except possibly for a few degenerate modes; see Appendix B). We extend this methodology to a full likelihood analysis in Sec. III, where we find that the general conclusions of this section remain valid.

III. LENSING RECONSTRUCTION: LIKELIHOOD ANALYSIS

In this section we explore the accuracy of reconstruction of lensing based on CMB temperature and polarization. We follow the analysis performed in Ref. [17] for the CMB temperature; most of the analysis extends easily to polarization, with one exception: the primary CMB has very little (if any) B-mode polarization. This means that the lensed CMB power spectrum $\langle\tilde{C}_l^{BB}\rangle_{LSS}$ cannot be expressed as a small perturbation on the unlensed power spectrum. We also include the effect of the field rotation in our discussion of the likelihood gradient and Fisher matrix, although we do not construct a "practical" estimator for it.

A. Likelihood function and gradient

For a given lens configuration with remapping function g, the covariance matrix $\hat{\mathbf{C}}$ of the measured CMB is computed from

$$\hat{\mathbf{C}}_g = \langle\hat{\mathbf{x}}\hat{\mathbf{x}}^\dagger\rangle = \langle(\tilde{\mathbf{x}}+\eta)(\tilde{\mathbf{x}}+\eta)^\dagger\rangle = \tilde{\mathbf{C}}_g + \mathbf{N} = \Lambda_g\mathbf{C}\Lambda_g^\dagger + \mathbf{N}, \quad (17)$$

where $\mathbf{N}=\langle\eta\eta^\dagger\rangle$ is the noise matrix. The measured CMB is Gaussian-distributed if we assume that the primary CMB $\mathbf{x}$ and instrument noise η are both Gaussian. (Note: the assumption of Gaussianity only applies *before* we average over LSS realizations.) The (negative log) likelihood function $\mathcal{L}$ for a lens configuration with remapping function g is then given (up to an irrelevant constant) by

$$\mathcal{L}(g) = \frac{1}{2}\ln\det\hat{\mathbf{C}}_g + \frac{1}{2}\hat{\mathbf{x}}^\dagger\hat{\mathbf{C}}_g^{-1}\hat{\mathbf{x}}. \quad (18)$$

Now we wish to determine the likelihood gradient with respect to the lens configuration $g=(\kappa,\omega)$. We will compute the gradient of the likelihood function, Eq. (18), using Eq. (17):

$$\frac{\partial\mathcal{L}}{\partial\kappa_{\mathbf{l}}} = \mathrm{Tr}\left(\hat{\mathbf{C}}_g^{-1}\frac{\partial\Lambda_g}{\partial\kappa_{\mathbf{l}}[g]}\mathbf{C}\Lambda_g^\dagger\right) - \hat{\mathbf{x}}^\dagger\hat{\mathbf{C}}_g^{-1}\frac{\partial\Lambda_g}{\partial\kappa_{\mathbf{l}}[g]}\mathbf{C}\Lambda_g^\dagger\hat{\mathbf{C}}_g^{-1}\hat{\mathbf{x}}. \quad (19)$$

The maximum-likelihood estimator is given by the relation $\partial\mathcal{L}/\partial\kappa_{\mathbf{l}}=0$. (We also require $\partial\mathcal{L}/\partial\omega_{\mathbf{l}}=0$ if we are estimating ω as well as κ.) However, maximum likelihood estimation of the lensing field is generally unstable because the lensing field has too many degrees of freedom. In order to regularize the problem, we introduce a Bayesian prior prob-

ability distribution $\propto e^{-\wp}$ for $g=(\kappa,\omega)$, i.e. we take prior probability $dP \propto e^{-\wp(g)}\Pi_{\mathbf{l}} d\kappa_{\mathbf{l}} d\omega_{\mathbf{l}}$. It is most convenient to take a Gaussian prior based on the power spectra of κ and (if applicable) ω:

$$\wp(\kappa,\omega)=\frac{1}{2}(\kappa^{\dagger}\mathbf{C}^{\kappa\kappa-1}\kappa+\ln\det\mathbf{C}^{\kappa\kappa})$$

$$+\frac{1}{2}(\omega^{\dagger}\mathbf{C}^{\omega\omega-1}\omega+\ln\det\mathbf{C}^{\omega\omega})$$

$$=\frac{1}{2}\sum_{\mathbf{l}}\left(\frac{|\kappa_{\mathbf{l}}|^2}{C_l^{\kappa\kappa}}+\ln C_l^{\kappa\kappa}\right)$$

$$+\frac{1}{2}\sum_{\mathbf{l}}\left(\frac{|\omega_{\mathbf{l}}|^2}{C_l^{\omega\omega}}+\ln C_l^{\omega\omega}\right), \quad (20)$$

where in the second equality we have assumed that the prior on κ and ω is statistically isotropic. (Note that this assumes the power spectra are known; we will consider the problem of estimating $C_l^{\kappa\kappa}$ from CMB data in Sec. V. The methods we present in Sec. V allow iterative determination of both the convergence field κ and the power spectrum $C_l^{\kappa\kappa}$.) If we are neglecting the field rotation then the terms involving ω should simply be removed. The mode of the posterior probability distribution is given by minimizing $\mathcal{L}+\wp$; we thus set $\partial\wp/\partial\kappa_{\mathbf{l}}=-\partial\mathcal{L}/\partial\kappa_{\mathbf{l}}$, or

$$[\mathbf{C}^{\kappa\kappa-1}\kappa]_{\mathbf{l}}^*=-\mathrm{Tr}\left(\hat{\mathbf{C}}_g^{-1}\frac{\partial\Lambda_g}{\partial\kappa_{\mathbf{l}}[g]}\mathbf{C}\Lambda_g^{\dagger}\right)$$

$$+\hat{\mathbf{x}}^{\dagger}\hat{\mathbf{C}}_g^{-1}\frac{\partial\Lambda_g}{\partial\kappa_{\mathbf{l}}[g]}\mathbf{C}\Lambda_g^{\dagger}\hat{\mathbf{C}}_g^{-1}\hat{\mathbf{x}}. \quad (21)$$

Because of the presence of the prior $\mathbf{C}^{\kappa\kappa-1}$, this estimator will filter out lensing modes that cannot be accurately reconstructed from the CMB data. It can thus be viewed as a sort of nonlinear generalization of the Wiener filter.

B. Practical estimator for the convergence

The likelihood gradient, Eq. (19), and hence the convergence estimator Eq. (21) based on it, are difficult to evaluate. We therefore investigate several approximations to the likelihood function. First, we consider only the convergence, κ; the rotation ω will be shown in Sec. IV to be unimportant unless instrument noise is very small. We note that Eq. (21) can be rewritten as

$$[\mathbf{C}^{\kappa\kappa,-1}\kappa]_{\mathbf{l}}^*=-\mathrm{Tr}\left(\Lambda_g^{\dagger-1}\mathbf{w}\Lambda_g^{-1}\frac{\partial\Lambda_g}{\partial\kappa_{\mathbf{l}}[g]}\mathbf{C}\mathbf{w}\Lambda_g^{-1}\hat{\mathbf{C}}_g\right)$$

$$+\hat{\mathbf{x}}^{\dagger}\Lambda_g^{\dagger-1}\mathbf{w}\Lambda_g^{-1}\frac{\partial\Lambda_g}{\partial\kappa_{\mathbf{l}}[g]}\mathbf{C}\mathbf{w}\Lambda_g^{-1}\hat{\mathbf{x}}, \quad (22)$$

where the weight matrix $\mathbf{w}$ is defined by

$$\mathbf{w}=\Lambda_g^{\dagger}\hat{\mathbf{C}}_g^{-1}\Lambda_g=(\mathbf{C}+\Lambda_g^{-1}\mathbf{N}\Lambda_g^{\dagger-1})^{-1}. \quad (23)$$

Here $\Lambda_g^{-1}\mathbf{N}\Lambda_g^{\dagger-1}$ is the delensed noise covariance matrix, which is equal to the noise covariance $\mathbf{N}$ for $g=0$ (no delensing). Under most circumstances, delensing has much less effect on the noise than on the CMB signal, because instrument noise is a relatively smooth function of l and contains both E and B modes with similar power. It is possible that $\Lambda_g^{-1}\mathbf{N}\Lambda_g^{\dagger-1}\ne\mathbf{N}$ if the noise power spectrum contains sharp features; in this case, the approximation $\mathbf{w}\approx\hat{\mathbf{C}}_{(0,0)}^{-1}$ used below may result in a nonoptimal, or (in extreme cases) unstable estimator.

We would like to approximate $\Lambda_g^{-1}(\partial\Lambda_g/\partial\kappa_{\mathbf{l}}[g])$ using the σ matrices; this can be done by expanding

$$\Lambda_g^{-1}\frac{\partial\Lambda_g}{\partial\kappa_{\mathbf{l}}[g]}=\left.\frac{\partial\kappa_{\mathbf{l}'}[g^{-1}g']}{\partial\kappa_{\mathbf{l}}[g']}\right|_{g'=g}\sigma_{\mathbf{l}}^{\kappa}$$

$$+\left.\frac{\partial\omega_{\mathbf{l}'}[g^{-1}g']}{\partial\kappa_{\mathbf{l}}[g']}\right|_{g'=g}\sigma_{\mathbf{l}}^{\omega}, \quad (24)$$

where the juxtaposition $g^{-1}g'$ indicates composition of the lensing operations: $(g^{-1}g')X=g^{-1}(g'(X))$. If the lensing is very weak we may take $(\partial\kappa_{\mathbf{l}'}[g^{-1}g']/\kappa_{\mathbf{l}}[g'])|_{g'=g}\approx[\Lambda_g^{-1}]_{\mathbf{l}',\mathbf{l}}$ and $(\partial\omega_{\mathbf{l}'}[g^{-1}g']/\kappa_{\mathbf{l}}[g'])|_{g'=g}\approx 0$, that is, the composition of lensing operations can be approximated by remapping the convergence field and neglecting rotation. In this regime, the statistical properties of the convergence field κ should not differ greatly from those of the "delensed" convergence field $\Lambda_g^{-1}\kappa$; mathematically, this means that we may take Λ_g and $\mathbf{C}^{\kappa\kappa}$ to commute. With these approximations, Eq. (22) becomes

$$C_l^{\kappa\kappa-1}(\Lambda_g\kappa)_{\mathbf{l}}^*=(\Lambda_g^{-1}\hat{\mathbf{x}})^{\dagger}\mathbf{w}\sigma_{\mathbf{l}}^{\kappa}\mathbf{C}\mathbf{w}\Lambda_g^{-1}\hat{\mathbf{x}}$$

$$-\mathrm{Tr}[\mathbf{w}\sigma_{\mathbf{l}}^{\kappa}\mathbf{C}\mathbf{w}\Lambda_g^{-1}\hat{\mathbf{C}}_g\Lambda_g^{\dagger-1}]. \quad (25)$$

The right-hand side of Eq. (25) is our approximation to the likelihood gradient, and the left-hand side is our (approximate)prior gradient. Note that the right-hand side evaluated at the correct lensing configuration g has expectation value zero, regardless of the choice of weight function; we will therefore choose the slightly suboptimal weight function $\mathbf{w}=\hat{\mathbf{C}}_{(0,0)}^{-1}$ in order to reduce computational difficulties. This leads us to the estimator

$$C_l^{\kappa\kappa-1}(\Lambda_g\kappa)_l^*=(\hat{\mathbf{C}}_{(0,0)}^{-1}\Lambda_g^{-1}\hat{\mathbf{x}})^{\dagger}\sigma_{\mathbf{l}}^{\kappa}\mathbf{C}\hat{\mathbf{C}}_{(0,0)}^{-1}\Lambda_g^{-1}\hat{\mathbf{x}}$$

$$-\mathrm{Tr}[\Lambda_g^{\dagger-1}\hat{\mathbf{C}}_{(0,0)}^{-1}\sigma_{\mathbf{l}}^{\kappa}\mathbf{C}\hat{\mathbf{C}}_{(0,0)}^{-1}\Lambda_g^{-1}\hat{\mathbf{C}}_g]. \quad (26)$$

[This choice leads to some difficulty for low-noise, wide-beam ($\theta_{FWHM}\gtrsim 10$ arcmin) experiments; see Sec. VI A for details.] By expanding $\hat{\mathbf{C}}_g$ using Eq. (17), and noting that in the harmonic-space basis, $\mathbf{C}$ and $\hat{\mathbf{C}}_{(0,0)}$ are diagonal whereas $\sigma_{\mathbf{l}}^{\Phi}$ has no nonzero diagonal elements, we convert this into

$$C_l^{\kappa\kappa-1}(\Lambda_g\kappa)_{\mathbf{l}}^* = (\hat{\mathbf{C}}_{(0,0)}^{-1}\Lambda_g^{-1}\hat{\mathbf{x}})^\dagger \sigma_{\mathbf{l}}^\kappa \mathbf{C}\hat{\mathbf{C}}_{(0,0)}^{-1}\Lambda_g^{-1}\hat{\mathbf{x}}$$
$$- \mathrm{Tr}[\Lambda_g^{\dagger-1}\hat{\mathbf{C}}_{(0,0)}^{-1}\sigma_{\mathbf{l}}^\kappa \mathbf{C}\hat{\mathbf{C}}_{(0,0)}^{-1}\Lambda_g^{-1}\mathbf{N}]. \tag{27}$$

C. Fisher matrix

The Fisher matrix is defined as the expectation value of the second derivative of the likelihood function:

$$F[\kappa_{\mathbf{l}},\kappa_{\mathbf{l}'}] \equiv \left\langle \frac{\partial^2\mathcal{L}}{\partial\kappa_{\mathbf{l}}^*\partial\kappa_{\mathbf{l}'}} \right\rangle \Bigg|_{(\kappa,\omega)} = \left\langle \frac{\partial\mathcal{L}}{\partial\kappa_{\mathbf{l}}^*}\frac{\partial\mathcal{L}}{\partial\kappa_{\mathbf{l}'}} \right\rangle \Bigg|_{(\kappa,\omega)}, \tag{28}$$

where the second equality follows from taking the second derivative $(\partial^2/\partial\kappa_{\mathbf{l}}^*\partial\kappa_{\mathbf{l}'})$ of the normalization condition $\int e^{-\mathcal{L}}\mathcal{D}\hat{\mathbf{x}}=1$ and noting that the expectation value of any statistic $\mathcal{S}$ is $\langle\mathcal{S}\rangle = \int \mathcal{S}e^{-\mathcal{L}}\mathcal{D}\hat{\mathbf{x}}$. (This also shows that $\mathbf{F}$ has all non-negative eigenvalues.) A similar relation holds for the field-rotation modes $\omega_{\mathbf{l}}$. We may thus compute the lensing

Fisher matrix as the covariance of the likelihood gradient; the easiest method of doing this is to apply Wick's theorem to compute the variance of Eq. (19). This yields

$$F[\kappa_{\mathbf{l}},\kappa_{\mathbf{l}'}] = \left\langle \frac{\partial\mathcal{L}}{\partial\kappa_{\mathbf{l}}^*}\frac{\partial\mathcal{L}}{\partial\kappa_{\mathbf{l}'}} \right\rangle = \mathrm{Tr}(\Lambda_g^\dagger\hat{\mathbf{C}}_g^{-1}\sigma_{-\mathbf{l}}^\kappa\mathbf{C}\Lambda_g^\dagger\hat{\mathbf{C}}_g^{-1}\sigma_{\mathbf{l}'}^\kappa\mathbf{C})$$
$$+ \mathrm{Tr}(\hat{\mathbf{C}}_g^{-1}\sigma_{-\mathbf{l}}^\kappa\mathbf{C}\Lambda_g^\dagger\hat{\mathbf{C}}_g^{-1}\Lambda_g\mathbf{C}\sigma_{\mathbf{l}'}^{\kappa\dagger}) \tag{29}$$

and similarly for the components of the Fisher matrix elements involving the field rotation. For simplicity, we compute the Fisher matrix at $g=(0,0)$, i.e. the $\kappa=\omega=0$ point, so that $\Lambda_{(0,0)}$ is the identity. The $\mathbf{C}$ and $\hat{\mathbf{C}}_{(0,0)}^{-1}$ matrices are diagonal in the $\{T,E,B\}$ basis:

$$\mathbf{C} = \begin{pmatrix} C_l^{TT} & C_l^{TE} & 0 \\ C_l^{TE} & C_l^{EE} & 0 \\ 0 & 0 & C_l^{BB} \end{pmatrix} \tag{30}$$

and

$$\hat{\mathbf{C}}_{(0,0)}^{-1} = \begin{pmatrix} (C_l^{EE}+N_l^{EE})/D_l & -C_l^{TE}/D_l & 0 \\ -C_l^{TE}/D_l & (C_l^{TT}+N_l^{TT})/D_l & 0 \\ 0 & 0 & 1/(C_l^{BB}+N_l^{BB}) \end{pmatrix}, \tag{31}$$

where $D_l = (C_l^{TT}+N_l^{TT})(C_l^{EE}+N_l^{EE}) - (C_l^{TE})^2$.

The overall Fisher matrix is then computed from Eq. (29):

$$F[\kappa_{\mathbf{l}},\kappa_{\mathbf{l}'}] \approx \frac{1}{2}\mathrm{Tr}[\hat{\mathbf{C}}_{(0,0)}^{-1}\mathbf{f}_{-\mathbf{l}}^\kappa\hat{\mathbf{C}}_{(0,0)}^{-1}\mathbf{f}_{\mathbf{l}'}^\kappa], \tag{32}$$

where

$$[\mathbf{f}_{\mathbf{l}}^\kappa]_{\mathbf{l}_1,-\mathbf{l}_2} = [\mathbf{C}\sigma_{-\mathbf{l}}^{\kappa\,\dagger} + \sigma_{\mathbf{l}}^\kappa\mathbf{C}]_{\mathbf{l}_1,\mathbf{l}_2}$$
$$= -\frac{\delta_{\mathbf{l}_1,\mathbf{l}-\mathbf{l}_2}}{\sqrt{4\pi}}\left(\frac{2}{l^2}\right)\mathbf{l}\cdot\begin{pmatrix} \mathbf{l}_1 C_{l_1}^{TT}+\mathbf{l}_2 C_{l_2}^{TT} & \mathbf{l}_1 C_{l_1}^{TE}\cos 2\alpha+\mathbf{l}_2 C_{l_2}^{TE} & -\mathbf{l}_1 C_{l_1}^{TE}\sin 2\alpha \\ \mathbf{l}_1 C_{l_1}^{TE}+\mathbf{l}_2 C_{l_2}^{TE}\cos 2\alpha & (\mathbf{l}_1 C_{l_1}^{EE}+\mathbf{l}_2 C_{l_2}^{EE})\cos 2\alpha & (\mathbf{l}_2 C_{l_2}^{BB}-\mathbf{l}_1 C_{l_1}^{EE})\sin 2\alpha \\ \mathbf{l}_2 C_{l_2}^{TE}\sin 2\alpha & (\mathbf{l}_2 C_{l_2}^{EE}-\mathbf{l}_1 C_{l_1}^{BB})\sin 2\alpha & (\mathbf{l}_1 C_{l_1}^{BB}+\mathbf{l}_2 C_{l_2}^{BB})\cos 2\alpha \end{pmatrix}. \tag{33}$$

(The matrix $\mathbf{f}_{\mathbf{l}}^\omega$ is identical except for the replacement $\mathbf{l}\cdot\rightarrow\star\mathbf{l}\cdot$.)Note that by hermiticity of $\mathbf{C}$, we have $\mathbf{f}_{\mathbf{l}}^\kappa = \mathbf{f}_{-\mathbf{l}}^{\kappa\,\dagger}$; for the individual 3×3 blocks in the harmonic-space basis, $[\mathbf{f}_{\mathbf{l}}^\kappa]_{\mathbf{l}_1,-\mathbf{l}_2} = [\mathbf{f}_{-\mathbf{l}}^{\kappa\,\dagger}]_{-\mathbf{l}_1,\mathbf{l}_2}$. Also our construction guarantees that $\mathbf{f}_{\mathbf{l}}^\kappa = \partial\hat{\mathbf{C}}/\partial\kappa_{\mathbf{l}}$ where the derivative is evaluated at $\kappa=\omega=0$.

It can be verified by explicit matrix multiplication that the computation for $F[\kappa,\kappa]$ here yields the uncertainty in the minimum-variance quadratic estimator of Ref. [10], with one exception: we have computed the Fisher matrix at $g=0$, hence the denominator of Eq. (32) contains the *unlensed* CMB power spectrum plus the instrument noise, whereas the equivalent calculation in Ref. [10] contains the *lensed* CMB

power spectrum plus the instrument noise. In the case of quadratic estimation, it is clear that the lensed power spectrum should be used in order to minimize the variance of the estimator. Conceptually, this is because the lensing B modes can be iteratively cleaned from the map, thereby reducing the post-cleaning B-mode power spectrum and reducing the noise in the lensing estimator. Our ability to clean the map is bounded, of course, by the sum of the unlensed CMB and noise contributions to C_l^{BB}.

D. Uncertainty in lens reconstruction

The usual method of estimating the uncertainty in lens reconstruction would be to invert the Fisher matrix. This

approach is motivated by the Cramer-Rao inequality, which states that an unbiased estimator of the lensing configuration must have covariance at least equal to $\mathbf{F}^{-1}$. Unfortunately, the Cramer-Rao inequality is only an inequality, and there is no guarantee that the bound $\mathbf{F}^{-1}$ can actually be reached; indeed this bound is only achieved in the case where the likelihood function is Gaussian with curvature $\mathbf{F}$. The traditional justification for assuming Gaussianity of the likelihood function is the central limit theorem. This works for studies of lensing of the CMB temperature field, in which the typical lensing mode being reconstructed is at $l \sim 100$ whereas the temperature fluctuations that are being lensed have wave number $l \sim 1000$; thus there are roughly $(1000/100)^2 = 100$ patches of primary CMB behind each lensing mode. (Most of the information comes from nonlocal correlations in the T field, so this argument technically requires more justification; nevertheless the calculations in Ref. [17] seem to indicate that it gives the correct answer.) This argument does not apply to lensing of the CMB polarization because the wave numbers of the primary E polarization modes and of the lensing field modes (κ_l) are both at wave numbers of order $l \sim 1000$. We should therefore be careful of possible problems with the Fisher matrix estimate, Eq. (32) of the uncertainty in the lensing field. In this section, we outline two such problems that occur in lensing reconstruction: first, a complete breakdown of the Fisher matrix approach when the field rotation ω becomes important; and second, fluctuations in the curvature matrix resulting from the statistical nature of the primary E field.

Consider first the problem of simultaneous reconstruction of both κ and ω. (We will see in Sec. IV that the noise levels required for this are not achievable in the near term, however, this extreme example serves to illustrate the problem.) One can see that if there are no primary B modes, then as instrument noise goes to zero, the uncertainty in κ and ω obtained by inverting Eq. (32) goes to zero. But this cannot be true because the *one* equation $B_{\text{unlensed}} = 0$ cannot be used to solve for the *two* fields κ and ω simultaneously. Therefore inverting the Fisher matrix yields a qualitatively absurd conclusion. What went wrong? The observation that one equation (the vanishing of the unlensed B-mode field) cannot be solved for two variables (κ and ω) yields a clue. Consider the case where instrument noise is negligible; then we know that the measured B mode is purely caused by lensing:

$$\hat{B}_{\mathbf{l}} = \sum_{\mathbf{l}'} \frac{1}{\sqrt{4\pi}} \left(\frac{2}{l'^2} \right) [\mathbf{l}' \cdot (\mathbf{l} - \mathbf{l}') \kappa_{\mathbf{l}'} + \star \mathbf{l}' \cdot (\mathbf{l} - \mathbf{l}') \omega_{\mathbf{l}'}] E_{\mathbf{l} - \mathbf{l}'} \sin 2\alpha, \tag{34}$$

where $\alpha = \phi_{\mathbf{l}} - \phi_{\mathbf{l} - \mathbf{l}'}$. To the extent that the E mode is dominated by the primary (not lensing) contribution, Eq. (34) is a linear system containing $2N$ unknown variables (the amplitudes of the κ and ω modes) but only N equations (the knowledge of the lensed B modes), thus there are degeneracy directions in lens configuration space which are unconstrained by the vanishing of the B modes. These directions must be constrained by a combination of the statistical prop-

erties of primary temperature and E-type polarization, and prior knowledge about the lensing field.

We can now understand why the lensing Fisher matrix $\mathbf{F}$ is inadequate for determining the uncertainty in the lensing fields κ and ω. The curvature matrix:

$$\mathcal{F}[\kappa_{\mathbf{l}}, \kappa_{\mathbf{l}'}](g) = \frac{\partial^2 \mathcal{L}}{\partial \kappa_{\mathbf{l}}^* \partial \kappa_{\mathbf{l}'}} \tag{35}$$

has very small eigenvalues in the directions of degeneracy of Eq. (34) and very large eigenvalues (approaching ∞ as $\mathbf{N} \to 0$) in the orthogonal directions. But as one can see from Eq. (34), the direction of degeneracy depends on E and hence on the specific realization of the CMB. If we average over CMB realizations to obtain a Fisher matrix $\mathbf{F}$, then we derive $\mathbf{F} = \infty$, which does not accurately reflect the nonzero errors in the degenerate directions of Eq. (34). Mathematically, the Fisher matrix methodology does not work because the error bars on (κ, ω) are extremely non-Gaussian. The lesson is that we should be careful about interpreting the inverse of the Fisher matrix as an uncertainty in parameters when the central limit theorem does not come to our aid.

A similar but less spectacular problem occurs in attempting reconstruction of small-scale lensing modes even when there is sufficient instrument noise that Ω is irrelevant. This is the regime of interest to a near-future high-resolution polarization experiment. The statistical uncertainty in the lensing reconstruction is given by the inverse of the curvature matrix $\mathcal{F}$. When doing a lens reconstruction, this curvature matrix is augmented by the curvature of the prior, $\mathbf{C}^{\kappa\kappa-1}$, so that the posterior error covariance matrix of the lensing reconstruction is approximately $(\mathcal{F} + \mathbf{C}^{\kappa\kappa-1})^{-1}$. We wish to compute the mean squared error in the reconstructed convergence $\hat{\kappa}$, which is obtained by computing the ensemble average of this covariance matrix over all realizations of CMB, noise, and LSS:

$$\mathbf{S}^{\kappa\kappa} = \langle (\hat{\kappa} - \kappa)(\hat{\kappa} - \kappa)^\dagger \rangle_{LSS} \approx \langle (\mathcal{F} + \mathbf{C}^{\kappa\kappa-1})^{-1} \rangle_{LSS}. \tag{36}$$

The Fisher matrix $\mathbf{F}$ is defined to be the expectation value of the curvature: $\mathbf{F} = \langle \mathcal{F} \rangle$ (with no LSS average). If the curvature matrix were always equal to $\mathbf{F}$, then it would be permissible to approximate $\mathbf{S}^{\kappa\kappa} \approx (\mathbf{F} + \mathbf{C}^{\kappa\kappa-1})^{-1}$. It can be shown (see Appendix C) that the statistical fluctuations of $\mathcal{F}$ always increase the uncertainty, Eq. (36); this increase we call the "curvature correction."

Conceptually, the naive calculation that the mean squared error is approximately $\mathbf{S}_0 = (\mathbf{F} + \mathbf{C}^{\kappa\kappa-1})^{-1}$ suffers problems for the same reason that the Fisher matrix calculation for simultaneously estimating κ and ω failed: the different realizations of the primary CMB introduce fluctuations in $\mathcal{F}$, and when we average over CMB realizations we generate a non-Gaussian error distribution for the estimated convergence.

The actual computation of the curvature corrections is not pursued here; some of the relevant issues are discussed in Appendix C, where we show that the "first-order noise contribution" of Ref. [21] arises as one part of the second-order curvature correction.

E. Relation to quadratic estimators

It is of interest to compare the estimator we have derived, Eq. (27), to the quadratic estimation method of Ref. [10]. The performance of the estimators is compared numerically in Sec. III B. Here we display the quadratic estimator and note the major differences between the quadratic and iterative estimators. The Wiener-filtered quadratic estimator is

$$\hat{\kappa}_{\mathbf{l}}^* = [C_l^{\kappa\kappa-1}\delta_{\mathbf{l},\mathbf{l}'} + F_{\mathbf{l},\mathbf{l}'}^{(\text{quad})}]^{-1}\hat{\mathbf{x}}^\dagger \langle\hat{\mathbf{C}}\rangle_{LSS}^{-1}\sigma_{\mathbf{l}'}^\kappa \mathbf{C}\langle\hat{\mathbf{C}}\rangle_{LSS}^{-1}\hat{\mathbf{x}}, \tag{37}$$

where the quadratic Fisher matrix is determined as

$$F_{\mathbf{l},\mathbf{l}'}^{(\text{quad})} = \frac{1}{2}\text{Tr}[\langle\hat{\mathbf{C}}\rangle_{LSS}^{-1}\mathbf{f}_{-\mathbf{l}}^\kappa\langle\hat{\mathbf{C}}\rangle_{LSS}^{-1}\mathbf{f}_{\mathbf{l}'}^\kappa]. \tag{38}$$

The "unbiased" (to first order in Φ), non-Wiener-filtered temperature is given by Eq. (37) with the "prior term"

$C_l^{\kappa\kappa-1}\delta_{0,\mathbf{l}+\mathbf{l}'}$ omitted. (The quadratic Fisher matrix is not technically a Fisher matrix, but its inverse does give the covariance of the unbiased quadratic estimator.) We prove in Appendix A that Eq. (37) and its unbiased equivalent are identical to the minimum-variance quadratic estimator that arises from the optimal weighting scheme of Ref. [10]. The mean squared error in the reconstructed convergence according to Eq. (37) is

$$\mathbf{S}^{\kappa\kappa(\text{quad})} = (\mathbf{C}^{\kappa\kappa-1} + \mathbf{F}^{(\text{quad})})^{-1}. \tag{39}$$

Several features of Eqs. (37) through (39) are readily apparent. First, the estimator Eq. (37) is a quadratic function of the CMB temperature or polarization field **x**. Secondly, we note that the uncertainty in the quadratic estimator is determined by the quadratic Fisher matrix, which contains the inverse of $\langle\hat{\mathbf{C}}\rangle_{LSS}$. For statistically isotropic noise, this inverse is given by

$$\langle\hat{\mathbf{C}}\rangle_{LSS}^{-1} = \begin{pmatrix} (\tilde{C}_l^{EE}+N_l^{EE})/\tilde{D}_l & -\tilde{C}_l^{TE}/\tilde{D}_l & 0 \\ -\tilde{C}_l^{TE}/\tilde{D}_l & (\tilde{C}_l^{TT}+N_l^{TT})/\tilde{D}_l & 0 \\ 0 & 0 & 1/(\tilde{C}_l^{BB}+N_l^{BB}) \end{pmatrix}, \tag{40}$$

where $\tilde{D}_l = \tilde{C}_l^{TT}\tilde{C}_l^{EE} - \tilde{C}_l^{TE\,2}$, and $\tilde{C}_l^{XX'}$ is the lensed CMB power spectrum (or cross-spectrum): $\tilde{C}_l^{XX'} = \langle\tilde{X}_\mathbf{l}\widetilde{X'}_\mathbf{l}^*\rangle_{LSS}$. Comparison of the quadratic Fisher matrix [Eq. (38)] to the full Fisher matrix [Eq. (32)] shows that the two are identical except for replacement of Eq. (40) by Eq. (31). This results in a qualitative difference between the two estimators: as instrument noise is reduced toward zero, the full Fisher matrix improves without bound ($\mathbf{F}\to\infty$), so (aside from foregrounds, field rotation, primary B modes, and the statistical concerns outlined in Sec. III D) the iterative estimator should be able to reconstruct the convergence with arbitrary accuracy. This is not so for the quadratic estimator, whose reconstruction accuracy is limited by the nonzero value of $\tilde{C}_l^{BB}$ and the resulting upper bounds on $\langle\hat{\mathbf{C}}\rangle_{LSS}^{-1}$ and $\mathbf{F}^{(\text{quad})}$. At high noise levels where the B mode cannot be mapped, however, $C_l^{BB}+N_l^{BB} \approx \tilde{C}_l^{BB}+N_l^{BB}$ since both sides of the equation are noise-dominated, and in this regime the performance of the two estimators should be nearly identical.

IV. STATISTICS OF FIELD ROTATION

Here we investigate the statistics of weak lensing fields with the objective of understanding the importance of the field rotation ω (or equivalently the cross-potential Ω) in CMB weak lensing. Field rotation is a cosmological contaminant in the sense that even with noiseless CMB data and no foregrounds, we cannot hope to recover two fields κ and ω from the single equation $B_{\text{unlensed}} = 0$. Therefore a nonzero power spectrum $C_l^{\omega\omega}$ translates into an uncertainty in the lens

reconstruction. We compute the power spectrum $C_l^{\omega\omega}$ by considering deflection angles; this has the advantage of providing a unified treatment of the higher-order Born approximation and "lens-lens coupling" effects considered by Ref. [22]. We work in the longitudinal gauge because in this gauge the perturbations to the metric remain small (of order 10^{-5} except in the very small portion of the universe near neutron stars and black holes) and so perturbation theory techniques are valid. We then consider the implications for lensing reconstruction; for near-term experiments, the effect is seen to be negligible.

A. Lensing power spectra

In the flat-sky approximation, we treat the photons as propagating in roughly the $-\hat{\mathbf{e}}_z$ direction so that the CMB experiment looks in the $\hat{\mathbf{e}}_z$ direction; the "sky" is in the xy plane. The spacetime metric observed by the photon is (so long as it does not stray far from the z axis)

$$ds^2 = a^2(\tau)[-(1+2\Psi)d\tau^2 + (1-2\Psi)$$
$$\times(d\chi^2 + \sin_K^2\chi\,(d\hat{n}_x^2 + d\hat{n}_y^2))], \tag{41}$$

where the Newtonian potential Ψ is generated by the non-relativistic matter inhomogeneities, and $\sin_K\chi = K^{-1/2}\sin(K^{1/2}\chi)$ where $K = -\Omega_K H_0^2$ is the curvature of the universe. The null geodesic equation in this metric is

$$\frac{d}{dr}\left(\frac{d\hat{\mathbf{n}}}{dr}\sin_K\chi\right) = -2\frac{\partial\Psi}{\partial\hat{\mathbf{n}}}\sin_K\chi. \tag{42}$$

The initial conditions are $\hat{\mathbf{n}}(\chi=0)=\hat{\mathbf{n}}_0$ and $\partial_\chi\hat{\mathbf{n}}(\chi=0)=0$.

The usual method here is to apply the first-order Born approximation to Eqs. (42), i.e. we perform the integration over the unperturbed photon trajectory. If we integrate forward, we find that

$$\hat{\mathbf{n}}(\chi)=\mathbf{n}_0-2\int_0^\chi W(\chi',\chi)\frac{\partial\Psi(\chi',\hat{\mathbf{n}}(\chi'))}{\partial\hat{\mathbf{n}}}d\chi', \quad (43)$$

where $W(\chi',\chi)=\cot_K\chi'-\cot_K\chi$. We may now apply the second-order Born approximation, in which we integrate not over the unperturbed photon trajectory but rather over the photon trajectory given by the first-order Born approximation, Eq. (42). Taylor-expanding the result to second order in Ψ yields

$$\hat{\mathbf{n}}(\chi)=\hat{\mathbf{n}}_0-2\int_0^\chi W(\chi',\chi)\partial_{\hat{\mathbf{n}}}\Psi(\chi',\hat{\mathbf{n}}_0)d\chi'$$
$$+4\int_0^\chi\int_0^{\chi'}W(\chi'',\chi')W(\chi',\chi)\partial_{\hat{\mathbf{n}}}^2$$
$$\times\Psi(\chi',\hat{\mathbf{n}}_0)\cdot\partial_{\hat{\mathbf{n}}}\Psi(\chi'',\hat{\mathbf{n}}_0)d\chi''d\chi'. \quad (44)$$

The convergence and field rotation at radial coordinate χ are most easily derived by taking the angular Fourier transform of this result. If we compute the deflection angle and perform the (κ,ω) decomposition of Eq. (8), we derive

$$\kappa_{\mathbf{l}}=-l^2\int_0^\chi W(\chi',\chi)\Psi_{\mathbf{l}}(\chi')d\chi'-2\sum_{\mathbf{l}'}\ (\mathbf{l}'\cdot\mathbf{l})[\mathbf{l}'\cdot(\mathbf{l}-\mathbf{l}')]$$
$$\times\int_0^\chi\int_0^{\chi'}W(\chi'',\chi')W(\chi',\chi)$$
$$\times\Psi_{\mathbf{l}'}(\chi')\Psi_{\mathbf{l}-\mathbf{l}'}(\chi'')d\chi''d\chi' \quad (45)$$

and

$$\omega_{\mathbf{l}}=-2\sum_{\mathbf{l}'}\ (\mathbf{l}'\cdot\star\mathbf{l})[\mathbf{l}'\cdot(\mathbf{l}-\mathbf{l}')]$$
$$\times\int_0^\chi\int_0^{\chi'}W(\chi'',\chi')W(\chi',\chi)$$
$$\times\Psi_{\mathbf{l}'}(\chi')\Psi_{\mathbf{l}-\mathbf{l}'}(\chi'')d\chi''d\chi'. \quad (46)$$

We now turn our attention to the statistics of Eqs. (45) and (46). We assume that Ψ can be described as a Gaussian random field because even in the nonlinear regime, our line of sight passes through many regions of independent density fluctuation and hence non-Gaussianity is suppressed by the central limit theorem. The power spectrum is

$$\langle\Psi_{\mathbf{l}_1}(\chi_1)\Psi_{\mathbf{l}_2}(\chi_2)\rangle=\delta_{\mathbf{l}_1+\mathbf{l}_2,0}\mathcal{C}_{l_1}^{\Psi\Psi}(\chi_1)\delta(\chi_1-\chi_2). \quad (47)$$

Here the projected potential power spectrum is determined using the Limber equation:

$$\mathcal{C}_l^{\Psi\Psi}(\chi)=\frac{1}{\sin_K^2\chi}P_\Psi\left(k=\frac{l}{\sin_K\chi},\chi\right)=\frac{9\sin_K^2\chi}{4l^4}\Omega_{m0}^2H_0^4(1$$
$$+z)^2P_\delta\left(k=\frac{l}{\sin_K\chi},\chi\right). \quad (48)$$

Here we have used the 3D power spectra of the Newtonian potential Ψ and fractional density perturbation $\delta=\rho/\bar{\rho}-1$; these are normalized in accordance with (here $q\in\{\Psi,\delta\}$)

$$P_q(k)=\int\ \langle q(0)q(\mathbf{x})\rangle e^{i\mathbf{k}\cdot\mathbf{x}}d^3\mathbf{x}, \quad (49)$$

so that the logarithmic band power is given by $\Delta_q^2(k)=(k^3/2\pi^2)P(k)$. The lowest-order contribution to the convergence power spectrum is given by the familiar result (here χ_0 is the comoving radial distance to the surface of last scatter)

$$C_l^{\kappa\kappa}=l^4\int_0^{\chi_0}W(\chi,\chi_0)^2\mathcal{C}_l^{\Psi\Psi}(\chi)d\chi. \quad (50)$$

(There are higher-order corrections to $C_l^{\kappa\kappa}$, but we do not consider them here since the purpose of this paper is to investigate lensing reconstruction, not to provide a precision theoretical computation of the lensing power spectra. Clearly if a sufficiently high-precision measurement of $C_l^{\kappa\kappa}$ is made, higher-order Born corrections should be considered in the theoretical interpretation of the power spectrum.) The field rotation power spectrum is given to lowest order by

$$C_l^{\omega\omega}=4\sum_{\mathbf{l}'}\ (\mathbf{l}'\cdot\star\mathbf{l})^2[\mathbf{l}'\cdot(\mathbf{l}-\mathbf{l}')]^2$$
$$\times\int_0^{\chi_0}d\chi\int_0^\chi d\chi'W(\chi,\chi_0)^2$$
$$\times W(\chi',\chi)^2\mathcal{C}_{l'}^{\Psi\Psi}(\chi)\mathcal{C}_{l-l'}^{\Psi\Psi}(\chi'). \quad (51)$$

Note that the lowest-order (in the Born expansion) contribution to $C_l^{\omega\omega}$ comes from the trispectrum of the density field. If the density field is non-Gaussian and this non-Gaussianity is insufficiently suppressed by the central limit theorem, then Eq. (51) will also contain a term from the connected trispectrum $\langle\Psi\Psi\Psi\Psi\rangle_{connected}$. However, because the factor $W(\chi'',\chi')$ in Eq. (46) vanishes as $\chi''\to\chi'$, it follows that the trispectrum components contributing to $C_l^{\omega\omega}$ involve correlations between points at widely spaced radial coordinates, which are suppressed. (Conceptually, this is because a single-screen lens only produces convergence and not field rotation, regardless of its Gaussianity or lack thereof. Thus if structures at different radial distances are independent, as assumed in the Limber approximation, then there is no connected contribution to $C_l^{\omega\omega}$.)

B. Effect on lensing estimation

We have computed the field rotation power spectrum, Eq. (51) for our fiducial cosmology using an analytic approxima-
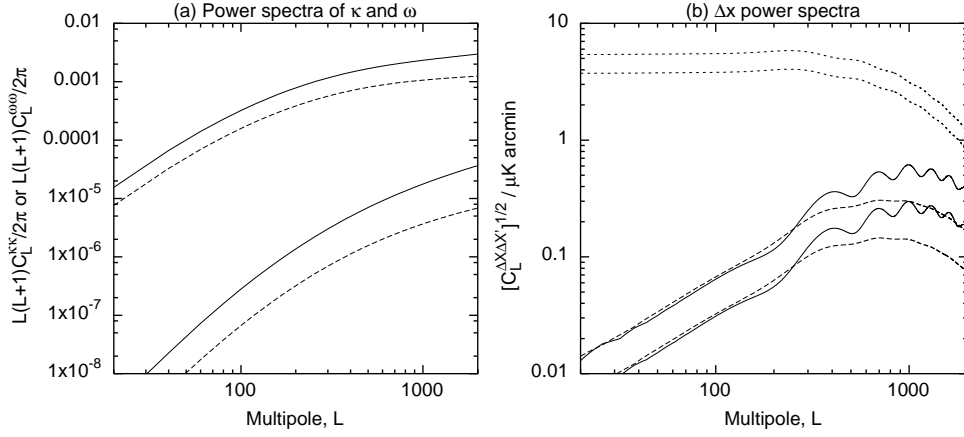
FIG. 2. (a) The convergence (upper curves) and field rotation (lower curves) power spectra in the fiducial cosmology. These are normalized to $\sigma_8^{\text{linear}} = 1.0$ (solid curves) and $\sigma_8^{\text{linear}} = 0.7$ (dashed curves). (b) The power spectra of ΔE (solid curves) and ΔB (long dashed) in "noise units" (μK arcmin). The short-dashed curves are the total B-mode power introduced by the convergence component. The upper curves are calculated for $\sigma_8^{\text{linear}} = 1.0$, the lower curves for 0.7.

tion to the growth factor [23] and a nonlinear mapping of the power spectrum [24]. The results are plotted in Fig. 2(a).

The effect of the field rotation on the lens reconstruction is to add an additional term to the CMB given by $\Delta x = \nabla x \cdot \star \nabla \Omega$. The power spectrum of Δx is given by

$$C_l^{\Delta T \Delta T} = \frac{1}{\pi} \sum_{\mathbf{l'}} \frac{1}{l'^4} [\star \mathbf{l'} \cdot (\mathbf{l} - \mathbf{l'})]^2 C_{l'}^{\omega\omega} C_{\mathbf{l}-\mathbf{l'}}^{TT},$$

$$C_l^{\Delta T \Delta E} = \frac{1}{\pi} \sum_{\mathbf{l'}} \frac{1}{l'^4} [\star \mathbf{l'} \cdot (\mathbf{l} - \mathbf{l'})]^2 C_{l'}^{\omega\omega} C_{\mathbf{l}-\mathbf{l'}}^{TE} \cos 2\alpha,$$

$$C_l^{\Delta E \Delta E} = \frac{1}{\pi} \sum_{\mathbf{l'}} \frac{1}{l'^4} [\star \mathbf{l'} \cdot (\mathbf{l} - \mathbf{l'})]^2 C_{l'}^{\omega\omega} (C_{\mathbf{l}-\mathbf{l'}}^{EE} \cos^2 2\alpha$$

$$+ C_{\mathbf{l}-\mathbf{l'}}^{BB} \sin^2 2\alpha), \qquad (52)$$

$$C_l^{\Delta B \Delta B} = \frac{1}{\pi} \sum_{\mathbf{l'}} \frac{1}{l'^4} [\star \mathbf{l'} \cdot (\mathbf{l} - \mathbf{l'})]^2 C_{l'}^{\omega\omega} (C_{\mathbf{l}-\mathbf{l'}}^{EE} \sin^2 2\alpha$$

$$+ C_{\mathbf{l}-\mathbf{l'}}^{BB} \cos^2 2\alpha),$$

where $\alpha = \phi_{\mathbf{l'}} - \phi_{\mathbf{l}}$. The field rotation is forbidden to have first-order correlations with the primary CMB and the convergence ($C_l^{T\omega} = C_l^{E\omega} = C_l^{\kappa\omega} = 0$) by parity; higher-order correlations with the primary CMB will be highly suppressed because ω is determined by small-scale fluctuations in density along the line of sight with window function that vanishes at the last-scattering surface. There are nonvanishing higher-order correlations between κ and ω, but we do not investigate these here. [But note that by reducing the conditional covariance $\langle \omega^2 \rangle|_\Phi - (\langle \omega \rangle|_\Phi)^2$, these correlations may enable us to reduce the "noise" due to field rotation further.]

The $\Delta \mathbf{x}$ power spectrum [Fig. 2(b)] shows that the ω-induced modifications to the CMB B modes are of the same order as instrument noise when the latter is reduced to $(N_l^{BB})^{1/2} \approx \mathcal{N}_P e^{l^2/2l_c^2} \approx 0.2 \, \mu$K arcmin. [Since we are trying

to set $B_{\text{unlensed}} = 0$, contamination in the B modes is more serious for lensing than contamination in the E modes; this is made mathematically explicit by multiplication by $\hat{\mathbf{C}}_{(0,0)}^{-1}$ in Eq. (31).] Since $\Delta \mathbf{x}$ has vanishing first-order correlation with $\mathbf{x}$, one might conjecture that the field rotation begins to interfere with lensing when the noise $\mathcal{N}_P$ is reduced to $\sim 0.2 \, \mu$K arcmin; however, $\Delta \mathbf{x}$ is highly non-Gaussian and exhibits many higher-order correlations with $\mathbf{x}$, so we should be cautious of trusting this conjecture. In the simulations (Sec. VI A), we find that even for our reference experiment F with 0.25 μK arcmin the field rotation does not significantly contaminate the reconstruction of the convergence field — it increases the mean squared error of the reconstruction by only $\sim 15\%$. We conclude that (at least at the level of the experiments considered here) the field rotation is not a problem for lens reconstruction.

V. ESTIMATING THE CONVERGENCE POWER SPECTRUM

Having investigated the reconstruction of the lensing field, we turn our attention to the convergence power spectrum, or equivalently the potential power spectrum, since the two are related by $C_l^{\kappa\kappa} = \frac{1}{4} l^4 C_l^{\Phi\Phi}$. In this section, we will ignore any complications associated with the field rotation as these are likely to be small for near-term experiments. In Sec. V A, we integrate the likelihood function for the convergence to yield the "grand likelihood function" for the lensing power spectrum; since this results in a functional integral over lens realizations, we simplify the problem by introducing a Gaussian approximation. We make further approximations in Sec. V B to yield an estimator that is suitable for actual computation.

A. Likelihood function and Gaussian approximation

Our basic approach, modeled after Ref. [17], is to compute the grand likelihood function $\bar{\mathcal{L}}$, which is a function of the lensing power spectrum:

$$\bar{\mathcal{L}}[C_l^{\kappa\kappa}] = -\ln \int \mathcal{D}\kappa \ e^{-\wp(\kappa)-\mathcal{L}(\kappa)} = \frac{1}{2}\ln\det\mathbf{C}^{\kappa\kappa}$$

$$-\ln \int \mathcal{D}\kappa \ \exp\left[-\frac{1}{2}\kappa^\dagger\mathbf{C}^{\kappa\kappa-1}\kappa - \mathcal{L}(\kappa)\right].$$

$$(53)$$

The objective of this section is to develop formalism to compute the minimum of $\bar{\mathcal{L}}$. A "practical" version suitable for numerical computation will be given in Sec. V B.

The integral in Eq. (53) has one dimension for each lensing mode and hence cannot be performed by any brute-force technique. In this situation the preferred solution is usually to use a Markov chain; unfortunately, the integral has of order 10^6 dimensions, and the integrand is expensive to compute,

hence Eq. (53) does not appear to be solvable by Markov chains either. We therefore choose to approximate Eq. (53) as a Gaussian, in which case the functional integral can be computed exactly:

$$\bar{\mathcal{L}}[C_l^{\kappa\kappa}] \approx \frac{1}{2}\ln\det\mathbf{C}^{\kappa\kappa} + \frac{1}{2}\ln\det[\mathcal{F}(\kappa_*)+\mathbf{C}^{\kappa\kappa-1}]$$

$$+ \frac{1}{2}\kappa_*^\dagger\mathbf{C}^{\kappa\kappa-1}\kappa_* + \mathcal{L}(\kappa_*), \quad (54)$$

where κ_* is the point where $\mathcal{L}+\frac{1}{2}\kappa^\dagger\mathbf{C}^{\kappa\kappa-1}\kappa$ is minimized, and $\mathcal{F}(\kappa_*)$ is the curvature matrix, Eq. (35), evaluated at the lens configuration $g=(\kappa_*,0)$.

A grand likelihood gradient Γ_l can then be defined:

$$\Gamma_l = \frac{\partial\bar{\mathcal{L}}}{\partial C_l^{\kappa\kappa}} \approx \frac{1}{2}\text{Tr}\left[\mathbf{C}^{\kappa\kappa-1}\frac{\partial\mathbf{C}^{\kappa\kappa}}{\partial C_l^{\kappa\kappa}}\right] + \frac{1}{2}\text{Tr}\left[(\mathcal{F}(\kappa_*)+\mathbf{C}^{\kappa\kappa-1})^{-1}\left(\frac{\partial\mathcal{F}(\kappa_*)}{\partial C_l^{\kappa\kappa}} - \mathbf{C}^{\kappa\kappa-1}\frac{\partial\mathbf{C}^{\kappa\kappa}}{\partial C_l^{\kappa\kappa}}\mathbf{C}^{\kappa\kappa-1}\right)\right]$$

$$-\frac{1}{2}\kappa_*^\dagger\mathbf{C}^{\kappa\kappa-1}\frac{\partial\mathbf{C}^{\kappa\kappa}}{\partial C_l^{\kappa\kappa}}\mathbf{C}^{\kappa\kappa-1}\kappa_* + \frac{\partial\mathcal{L}(\kappa_*)}{\partial C_l^{\kappa\kappa}} + \frac{\partial\kappa_*}{\partial C_l^{\kappa}}\left\{\text{Tr}\left[(\mathcal{F}(\kappa)+\mathbf{C}^{\kappa\kappa-1})^{-1}\frac{\partial\mathcal{F}(\kappa)}{\partial\kappa}\right] + \frac{\partial}{\partial\kappa}\left(\frac{1}{2}\kappa^\dagger\mathbf{C}^{\kappa\kappa-1}\kappa + \mathcal{L}(\kappa)\right)\right\}\Bigg|_{\kappa=\kappa_*}.$$

$$(55)$$

Note that $\mathcal{L}(\kappa_*)$, and hence $\mathcal{F}(\kappa_*)$, do not depend on $C_l^{\kappa\kappa}$ except implicitly through κ_*. Also, if we use that κ_* is the maximum of $\frac{1}{2}\kappa^\dagger\mathbf{C}^{\kappa\kappa-1}\kappa + \mathcal{L}$ with respect to κ, we find that the final derivative with respect to κ in this equation vanishes. We further note that $\partial\mathbf{C}^{\kappa\kappa}/\partial C_l^{\kappa\kappa}$ is simply the projection operator onto the l representation of $SO(3)$, i.e. in harmonic space it has 1's as diagonal elements with multipole l and 0's everywhere else. Defining d_l to be the number of modes of multipole l (note that on the sphere, $d_l=2l+1$), this allows us to simplify Eq. (55) to

$$\Gamma_l \approx \frac{d_l}{2C_l^{\kappa\kappa}} - \sum_{\mathbf{l}:\ |\mathbf{l}|=l}\frac{[(\mathcal{F}(\kappa_*)+\mathbf{C}^{\kappa\kappa-1})^{-1}]_{\mathbf{l},\mathbf{l}}}{2C_l^{\kappa\kappa\,2}} - \sum_{\mathbf{l}:\ |\mathbf{l}|=l}\frac{|\kappa_{*\mathbf{l}}|^2}{2C_l^{\kappa\kappa\,2}} + \frac{\partial\kappa_*}{\partial C_l^{\kappa\kappa}}\text{Tr}\left[(\mathcal{F}(\kappa_*)+\mathbf{C}^{\kappa\kappa-1})^{-1}\frac{\partial\mathcal{F}(\kappa)}{\partial\kappa}\Bigg|_{\kappa_*}\right]. \quad (56)$$

Here the sums are over all modes $\mathbf{l}$ of multipole l.

It sometimes occurs that we wish to estimate the lensing power spectrum not by estimating the individual $C_l^{\kappa\kappa}$, but rather by "binning" the power spectrum. This is useful if, e.g., the $(S/N)^2$ per multipole is low or if the partial-sky nature of a survey causes confusion between power in neighboring multipoles. In this case, we introduce "basis functions" $\{\mathcal{M}^\mu\}$ for the lensing power spectrum:

$$C_l^{\kappa\kappa} = \sum_\mu c_\mu\mathcal{M}_l^\mu. \quad (57)$$

The coefficients c_μ are now to be estimated. The maximum-likelihood estimator is then the choice of c_μ that satisfies:

$$\sum_l \mathcal{M}_l^\mu\Gamma_l=0 \ \ \forall\mu. \quad (58)$$

B. Practical estimator and uncertainty

Ideally, we would like to implement the maximum-likelihood estimator for the coefficients c_μ, i.e. Eq. (58). Unfortunately, this involves setting to zero some linear combination of the Γ_l's given by Eq. (56), which is a highly nontrivial task. We therefore take the approximation that the curvature matrix $\mathcal{F}$ does not depend on κ, then Γ_l is seen to depend only on the quantities $C_l^{\kappa\kappa}$ and

$$\dot{v}_l \equiv \frac{1}{2C_l^{\kappa\kappa\,2}}\sum_{\mathbf{l}:\ |\mathbf{l}|=l}|\kappa_{*\mathbf{l}}|^2, \quad (59)$$

which explicitly depends on $C_l^{\kappa\kappa}$ but is also implicitly a function of $C_l^{\kappa\kappa}$ through its dependence on κ_*. Note that the functional form of Γ_l is $\Gamma_l(C_l^{\kappa\kappa},\dot{v}_l)=\Gamma_l^{(0)}(C_l^{\kappa\kappa})-\dot{v}_l$. Equation (58) then reads

$$0 = \sum_l \mathcal{M}_l^\mu [\Gamma_l^{(0)}(C_l^{\kappa\kappa}) - \dot{v}_l(c_\nu)] = \Gamma_\mu^{(0)}(c_\nu) - v_\mu(c_\nu),$$

$$(60)$$

where we have defined $v_\mu(c_\nu) = \Sigma_l \mathcal{M}_l^\mu \dot{v}_l(c_\nu)$. We are thus attempting to solve $v_\mu(c_\nu) = \Gamma_\mu^{(0)}$, but $\Gamma_\mu^{(0)}$ is some complicated function of the convergence power spectrum coefficients $\{c_\nu\}$. We solve this problem by approximating $\Gamma_\mu^{(0)}(c_\nu) \approx \langle v_\mu(c_\nu) \rangle_{LSS[c_\nu]}$, i.e. the expected value of $v_\mu(c_\nu)$ where the LSS realizations are drawn from a lensing convergence power spectrum $C_l^{\kappa\kappa} = \Sigma_\nu c_\nu \mathcal{M}_l^\nu$. We therefore use the estimator

$$v_\mu(c_\nu) = \langle v_\mu(c_\nu) \rangle_{LSS[c_\nu]} \ \forall \mu. \tag{61}$$

Equation (61) is somewhat abstract, so we clarify its meaning here. The statistic $v_\mu(c_\nu)$ is proportional to the power spectrum of the iterative convergence estimator obtained by solving Eq. (27); this depends on the prior power spectrum $C_l^{\kappa\kappa} = \Sigma_\nu c_\nu \mathcal{M}_l^\nu$ as well as on the data. The solution $\{c_\nu\}$ to Eq. (61) is the set of power spectrum coefficients for which v_μ equals its expected value (which is most easily determined via Monte Carlo simulation). This approach has the advantage of "calibrating out" the noise biases discussed by Ref. [21]. (Note that some convergence modes—those corresponding to large eigenvalues of the curvature $\mathcal{F}$—are reconstructed better than others. What is especially useful about v_μ, or equivalently the power spectrum of the iterative estimator, is that the iterative estimator filters out the poorly reconstructed modes. Thus the convergence modes that are reconstructed more accurately are weighted more heavily in determining v_μ and hence in determining the convergence power spectrum.)

Finally, we wish to determine the uncertainty on the solution $\{c_\nu\}$ to Eq. (61). If we average over many convergence modes, then this uncertainty should be given by the inverse of the grand Fisher matrix $^{(G)}\mathbf{F}$ for power spectrum determination:

$$^{(G)}F_{\mu,\mu'} = \langle \Gamma_\mu \Gamma_{\mu'} \rangle_{LSS} \approx \langle \delta v_\mu \delta v_{\mu'} \rangle_{LSS}$$
$$= \langle v_\mu v_{\mu'} \rangle_{LSS} - \langle v_\mu \rangle_{LSS} \langle v_{\mu'} \rangle_{LSS}, \tag{62}$$

i.e. $^{(G)}\mathbf{F}$ is the covariance matrix of v_μ. If the *reconstructed* convergence κ_* can be approximated as a Gaussian random field (which is true in the case where the reconstruction has high signal-to-noise ratio since in this case the $\kappa_* \approx \kappa$, which is Gaussian because κ is produced by many LSS fluctuations along the line of sight), then we can take the Gaussian approximation to Eq. (62). This is obtained by considering the covariance of $\dot{v}_l$ according to Eq. (59) and using Wick's theorem; this yields

$$^{(G)}F_{\mu,\mu'} = \sum_l \frac{d_l \mathcal{M}_l^\mu \mathcal{M}_l^{\mu'}}{2} \left(\frac{C_l^{\kappa_* \kappa_*}}{C_l^{\kappa\kappa\ 2}} \right)^2. \tag{63}$$

We remind the reader once again that the approximation Eq. (63) to the power spectrum estimation uncertainty is only valid if the reconstructed convergence field κ_* is approximately Gaussian. If κ_* has a significant trispectrum when averaged over LSS+CMB+noise realizations, then Eq. (62) must be used instead. This is only a problem in the low signal-to-noise (high l) regime in which lensing modes cannot be reconstructed individually and their power is only statistically detected.

VI. NUMERICAL SIMULATIONS

Throughout our derivation of lensing and tensor power estimators, we have made various approximations that should be tested. The most robust way to do this is to conduct a numerical simulation of the CMB and lensing field, and then construct lensing estimators, comparing the error to the theoretical estimates of Eqs. (32) and (38). In all cases, we have used a flat sky with toroidal boundary conditions. We will only simulate the CMB polarization here; formally, the polarization-only estimators are obtained by setting $\mathcal{N}_T = \infty$ in the relevant equations.

A. Reconstructing the convergence

The simplest simulations involve reconstruction of the convergence κ. We generate simulated CMB T, Q, U, and κ fields on a $34°08'$ square patch of sky with resolution 1 arcmin per pixel (2048×2048 pixels); lens the simulated CMB; and add appropriate noise.

We wish to compare the quadratic estimator, Eq. (37) with our new estimator, Eq. (27). The former is relatively straightforward to compute; the latter requires that we apply the methods of Sec. III B. We simulate Gaussian random realizations of the Q, U, and κ fields, perform the lensing remapping, and add appropriate noise. We then compute the estimators of Eqs. (37) and (27). There are two tricks that are very useful in numerical computation of these estimators: first, simultaneous computation of inner products $\mathbf{t}^\dagger \sigma_\mathbf{l}^\kappa \mathbf{u}$ for all $\mathbf{l}$; and second, stochastic trace computation. We discuss each of these here.

The simultaneous computation of inner products was introduced by Refs. [4,10] in order to compute quadratic estimators. A general version of this is (on a flat sky; see Ref. [25] for an all-sky version)

$$\sum_\mathbf{l} \left(\frac{l^2}{2} \right) (\mathbf{t}^\dagger \sigma_{-\mathbf{l}}^\kappa \mathbf{u}) e^{i\mathbf{l} \cdot \hat{\mathbf{n}}} = \sum_{X \in \{T,Q,U\}} \nabla_{\hat{\mathbf{n}}} \cdot [t_X^*(\hat{\mathbf{n}}) \nabla_{\hat{\mathbf{n}}} u_X(\hat{\mathbf{n}})].$$

$$(64)$$

(Note that this equation requires that $\mathbf{t}$ and $\mathbf{u}$ be written in the $\{T, Q, U\}$ basis since E and B have different transformation properties under lensing. Also the asterisk on t_X is of course unnecessary if $\mathbf{t}$ is a real field.) If $\mathbf{t}$ and $\mathbf{u}$ are expressed in real-space, then the right-hand side is easily evaluated. The quantities $\frac{1}{2} l^2 \mathbf{t}^\dagger \sigma_\mathbf{l}^\kappa \mathbf{u}$ are then obtained via a fast Fourier transform; division then trivially removes the $l^2/2$. There is a zero-wavenumber mode corresponding to $l = 0$ which presents a problem for division. Here we simply set this convergence mode to zero; in the complete all-sky treatment this would be justified by noting that the convergence is $-\frac{1}{2}$ times the divergence of the deflection angle vector, hence

$\int \kappa \, d^2 \hat{\mathbf{n}} = 0$ and so $\kappa_{l=0} = 0$. [There is a corresponding trick for the field rotation: Eq. (64) remains valid if we make the replacements $\sigma_{\mathbf{l}}^{\kappa} \to \sigma_{\mathbf{l}}^{\omega}$ and $\nabla_{\hat{\mathbf{n}}} u_X(\hat{\mathbf{n}}) \to \star \nabla_{\hat{\mathbf{n}}} u_X(\hat{\mathbf{n}})$.]

The trace in Eq. (27) is most easily evaluated stochastically: if we generate a random noise vector η with covariance $\mathbf{N}$, then the trace is equal to the expectation value:

$$\mathrm{Tr}[\Lambda_g^{\dagger \, -1} \mathbf{C}_{(0,0)}^{-1} \sigma_{\mathbf{l}}^{\kappa} \mathbf{C} \hat{\mathbf{C}}_{(0,0)}^{-1} \Lambda_g^{-1} \mathbf{N}]$$
$$= \langle (\hat{\mathbf{C}}_{(0,0)}^{-1} \Lambda_g^{-1} \eta)^{\dagger} \sigma_{\mathbf{l}}^{\kappa} \mathbf{C} \hat{\mathbf{C}}_{(0,0)}^{-1} \Lambda_g^{-1} \eta \rangle. \tag{65}$$

If this Monte Carlo method is used to compute the trace, then the Monte Carlo error in its computation for one realization of η is less than or equal to the instrument noise contribution to the uncertainty on the right-hand side of Eq. (27). [This is because the right-hand side of Eq. (27) is a quadratic function of $\hat{\mathbf{x}}$, with covariance $\hat{\mathbf{C}}$, which is greater than the noise covariance $\mathbf{N}$ along all directions.] Since the Monte Carlo error variance scales as the reciprocal of the number of realizations of η used, it follows that of order a few realizations of η are sufficient in evaluating Eq. (65). In fact for the reference experiments described here, we find that there is little gain in taking more than one realization of η.

We solve Eq. (27) using the iterative procedure:

$$\kappa_{n+1,\mathbf{l}}^{*} = \zeta_l [\Lambda_g^{-1}]_{-\mathbf{l},-\mathbf{l}'} C_{l'}^{\kappa\kappa} \{ (\hat{\mathbf{C}}_{(0,0)}^{-1} \Lambda_g^{-1} \hat{\mathbf{x}})^{\dagger} \sigma_{\mathbf{l}'}^{\kappa} \mathbf{C} \hat{\mathbf{C}}_{(0,0)}^{-1} \Lambda_g^{-1} \hat{\mathbf{x}}$$
$$- \mathrm{Tr}[\Lambda_g^{\dagger \, -1} \hat{\mathbf{C}}_{(0,0)}^{-1} \sigma_{\mathbf{l}'}^{\kappa} \mathbf{C} \hat{\mathbf{C}}_{(0,0)}^{-1} \Lambda_g^{-1} \mathbf{N}] \} + (1 - \zeta_l) \kappa_{n,\mathbf{l}}^{*}. \tag{66}$$

Here g is the lens configuration with convergence κ and no rotation: $g = (\kappa, 0)$, and the ζ_l are convergence parameters; we choose them to be

$$\zeta_l = \frac{\zeta_{(c)}}{1 + C_l^{\kappa\kappa} F_l^{\kappa\kappa}}, \tag{67}$$

where $\zeta_{(c)}$ is a constant satisfying $0 < \zeta_{(c)} < 2$. It is found that the Wiener-filtered quadratic estimator, Eq. (37), is a good choice for initializing this iteration. The choice of $\zeta_{(c)}$ is an intricate issue: if it is set too small, the rate of convergence of the iteration becomes unacceptably slow; if it is set too high, the iteration can fail to converge entirely. The convergence can be understood by approximating Eq. (66) as linear in κ_n:

$$\kappa_{n+1,\mathbf{l}}^{*} \approx \zeta_l C_l^{\kappa\kappa} [\mathcal{F}(\kappa_n - \hat{\kappa})]_{\mathbf{l}} + (1 - \zeta_l) \kappa_{n,\mathbf{l}}. \tag{68}$$

Here we have approximated the response matrix of the likelihood gradient using the curvature matrix. Then the requirement for convergence is that all of the (possibly complex) eigenvalues of the matrix

$$R_{\mathbf{l},\mathbf{l}'} = \delta_{\mathbf{l},\mathbf{l}'} - \zeta_{(c)} \frac{C_{\mathbf{l},\mathbf{l}'}^{\kappa\kappa-1} + \mathcal{F}_{\mathbf{l},\mathbf{l}'}}{C_l^{\kappa\kappa-1} + F_l^{\kappa\kappa}} \tag{69}$$

lie within the unit circle. Note that, averaged over CMB +noise realizations, $\langle \mathcal{F} \rangle = \mathbf{F}$, and hence $\langle \mathbf{R} \rangle = (1 - \zeta_{(c)}) \mathbf{1}$;

hence we conclude that the iterative procedure should be convergent for $0 < \zeta_{(c)} < 2$ in the absence of curvature corrections. In reality, very small values of $\zeta_{(c)}$ may be necessary for convergence, especially in cases where curvature corrections are large. Since $\mathcal{F} + \mathbf{C}^{\kappa\kappa-1}$ is positive definite at the maximum posterior probability point, there is always a positive value of $\zeta_{(c)}$ that results in convergence. The cases in which the small values of $\zeta_{(c)}$ are required are those in which curvature corrections are large; we have found from our simulations that these are the low-noise experiments. Convergence is generally found to be faster for the high l convergence modes.

One problem we have encountered is that for experiments with low noise and wide beam ($\theta_{FWHM} \gtrsim 10$ arcmin), the iterative estimator given by Eq. (66) is unstable. This instability arises because the noise N_l is strongly blue; hence the delensing operation Λ_g^{-1} in Eq. (66) mixes high-multipole noise down to lower multipoles where it disrupts the lensing estimation. This problem is in principle solvable by using the correct $(\mathbf{C} + \Lambda_g^{-1} \mathbf{N} \Lambda_g^{\dagger \, -1})^{-1}$ weight function in place of $\mathbf{C}_{(0,0)}^{-1}$ in Eq. (23). However, since this occurs in a regime where the iterative approach does not improve upon the quadratic estimator approach anyway, we recommend simply using the quadratic estimator for wide-beam experiments.

We illustrate by considering the reconstruction of lensing using reference experiments A–F. The residual error in the reconstructed convergence $\hat{\kappa}$, as measured by computing the power spectrum of the difference $\hat{\kappa} - \kappa$ between input and reconstructed convergence maps, is shown in Fig. 3 for both quadratic and iterative estimators. For the iterative estimator applied to Ref. expt. C, we set $\zeta_{(c)} = 0.12$ in Eq. (66), used three realizations of the η field in Eq. (65), and performed 64 iterations. Reference expt. F has a lower noise level and so it is necessary to use the smaller convergence parameter $\zeta_{(c)} = 0.04$; the convergence is thus slower and we used 256 iterations. Reference expt. A has a higher noise level and so we can use $\zeta_{(c)} = 0.2$ and 24 iterations. Maps of the input and reconstructed convergence fields for the Ref. expt. C reconstruction are shown in Fig. 4. The dependence of the iterative estimator reconstruction accuracy on noise $\mathcal{N}_P$ and beam size (FWHM) θ_{FWHM} is shown in Fig. 5. We have also displayed in Fig. 5 the (theoretical) reconstruction error curves for the quadratic estimator in the absence of instrument noise. These curves represent the fundamental limit to the reconstruction accuracy possible with quadratic estimators; it is readily seen that the iterative estimator can do better if noise is low ($\mathcal{N}_P < 0.5$–1 μK arcmin, depending on the range of l considered). [Note that we display $C_l^{\kappa\kappa}$ in these plots, whereas some authors have displayed instead $l(l+1)C^{\mathbf{dd}}/2\pi$, where $\mathbf{d} = \nabla\Phi$ is the deflection angle. The two are related by $l(l+1)C^{\mathbf{dd}}/2\pi = (2/\pi)C_l^{\kappa\kappa}$.]

The accuracy of reconstruction can also be represented by the correlation coefficient $\rho_l = C_l^{\kappa\hat{\kappa}} / \sqrt{C_l^{\kappa\kappa} C_l^{\hat{\kappa}\hat{\kappa}}}$. The correlation coefficient is the figure of merit if the objective is to cross-correlate the convergence from CMB lensing with another tracer of the density (e.g. weak lensing of galaxies), since the signal-to-noise ratio of the cross-correlation is de-
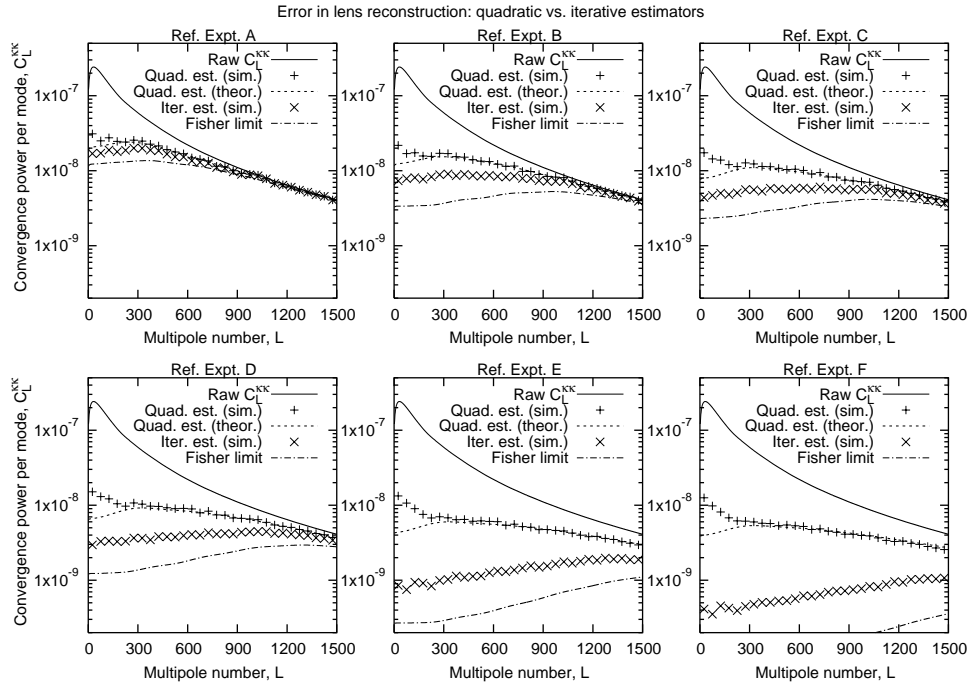
FIG. 3. The power spectrum of the error in the convergence reconstruction for reference expts. A–F. The top curve in each panel shows the overall convergence power spectrum $C_l^{\kappa\kappa}$. The middle curve shows the theoretical, i.e. from Eq. (38) power spectrum of the convergence error $\hat{\kappa} - \kappa$ in the Wiener-filtered quadratic estimator Eq. (37); the "+" data points indicate the power spectrum of this error as recovered from simulations. The error power spectrum for the iterative estimator Eq. (66), again as recovered from simulations, is shown with the "×" data points. The bottom curve shows the theoretical best performance if the Fisher matrix limit Eq. (32) can be achieved, i.e. if we had a truly optimal estimator and no curvature corrections. Note the more dramatic improvement provided by the iterative estimator when the noise is small. Field rotation was neglected in the calculations for this figure.

termined by ρ_l. We have plotted the correlation coefficient in Fig. 6 for the various reference experiments. The iterative estimator offers improved reconstruction, especially for the lower-noise experiments (C–F).

Up until this point we have neglected the field rotation ω; we should verify that this is justified. We do this by introducing field rotation with power spectrum given by Eq. (51) as computed in Sec. IV B with normalization $\sigma_8^{\text{linear}} = 0.84$. We then compare the performance of the iterative estimator, Eq. (66), with and without the field rotation. The comparison

is shown in Fig. 7; it is seen that the field rotation increases the mean squared error of the reconstruction by only $\sim 5\%$ for Ref. expt. E (0.5 μK arcmin noise, 2 arcmin beam) and $\sim 15\%$ for Ref. expt. F (0.25 μK arcmin noise, 2 arcmin beam).

As a final note, we find that for low noise levels, a large number of iterations is required because our iterative process is ill-conditioned. Indeed, it is possible that there are eigenvalues of **R** that are so close to unity that their corresponding modes have not converged even after tens or hundreds of
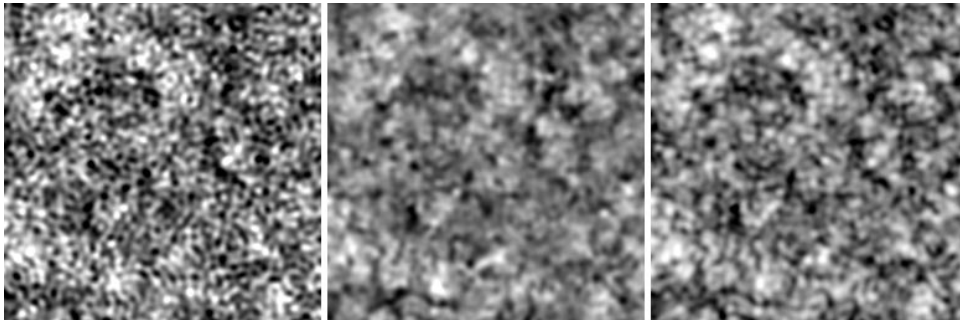

FIG. 4. A simulated reconstruction of the lensing convergence using polarization and reference expt. C parameters. In the left panel, we display the realization of the convergence field κ used to produce the simulated CMB. The reconstructions using the Wiener-filtered quadratic estimator and the iterative estimator are shown in the center and right panels, respectively. These frames are each $8°32'$ in angular width, corresponding to 1/16 of the simulated area; the scale ranges from black (diverging, $\kappa = -0.12$) through white (converging, $\kappa = +0.12$). Although all lensing multipoles up to $l = 3600$ are simulated, we have only displayed the $l \leq 1600$ modes in these figures for clarity. Field rotation was neglected in the calculations for this figure.
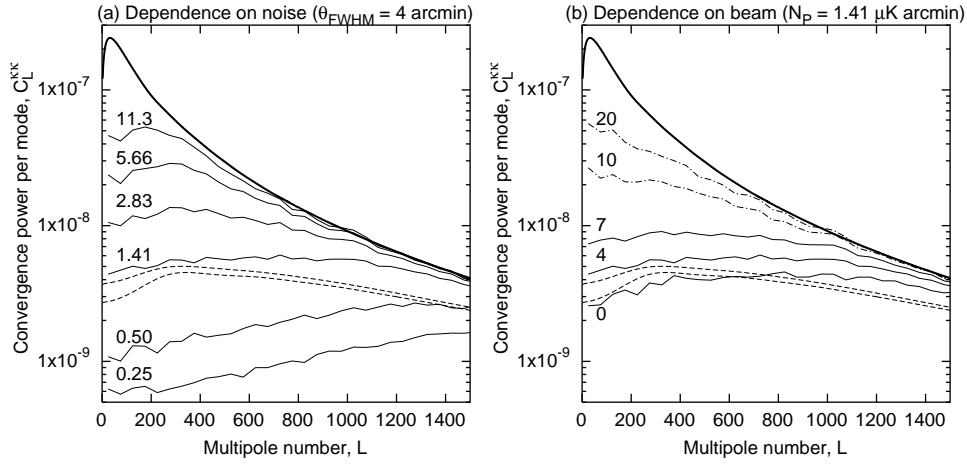
FIG. 5. The dependence of the mean squared error in lensing reconstruction, $\langle |\kappa_{\mathbf{l}} - \hat{\kappa}_{\mathbf{l}}|^2 \rangle$, on the instrument parameters. The baseline is Ref. expt. C, $\mathcal{N}_P = 1.41$ μK arcmin, $\theta_{FWHM} = 4$ arcmin. The thick solid line is the raw power spectrum $C_l^{\kappa\kappa}$; the thin solid lines indicate the mean squared error for the lensing reconstruction using the iterative estimator. As described in the text, the iterative estimator is unusable for wide-beam experiments (≥ 10 arcmin); we used the quadratic estimator for these cases (dot-dashed curves). The dashed lines indicate the ideal zero-noise reconstruction error from the quadratic estimator according to Eq. (38) with polarization only (top) and temperature +polarization (bottom). (a) Changing $\mathcal{N}_P$; units are μK arcmin. (b) Changing θ_{FWHM}; units are arcmin.

iterations; if this is the case, then it should be possible to improve upon our results by increasing the number of iterations, or by finding an iterative scheme that converges faster than Eq. (66). This is allowed by the Fisher matrix noise limits, which are significantly lower than the achieved noise levels (see Fig. 3). We consider this possibility unlikely since we tried increasing the number of iterations in several of the simulations and found little improvement. Additionally, modes with eigenvalue $\lambda_{\mathbf{R}}$ close to unity correspond to flat directions of the curvature matrix $\mathcal{F}$ [see Eq. (69)]; such directions, however, cannot be reconstructed accurately regardless of how many iterations are used.

B. Extracting the convergence power spectrum

We compute the lensing power spectrum from simulated data by solving Eq. (61). The approach, once again, is itera-

tive: we adjust the power spectrum $C_l^{\kappa\kappa}$ until $v_\mu = \langle v_\mu \rangle_{LSS}$. [Note that both the left and right sides of Eq. (61) depend on $C_l^{\kappa\kappa}$.] We will attempt here to compute the binned power spectrum, i.e. we choose a basis for the convergence power spectrum given by

$$\mathcal{M}_\mu^l = \begin{cases} 1 & \mu\Delta l < l \leq (\mu+1)\Delta l, \\ 0 & \text{otherwise,} \end{cases} \tag{70}$$

where Δl is the bin width and μ ranges from 0 through $N_{bin} - 1$. We are thus attempting to reconstruct the power spectrum in N_{bin} bins, equally spaced out to maximum multipole $l_{\max} = N_{bin}\Delta l$.

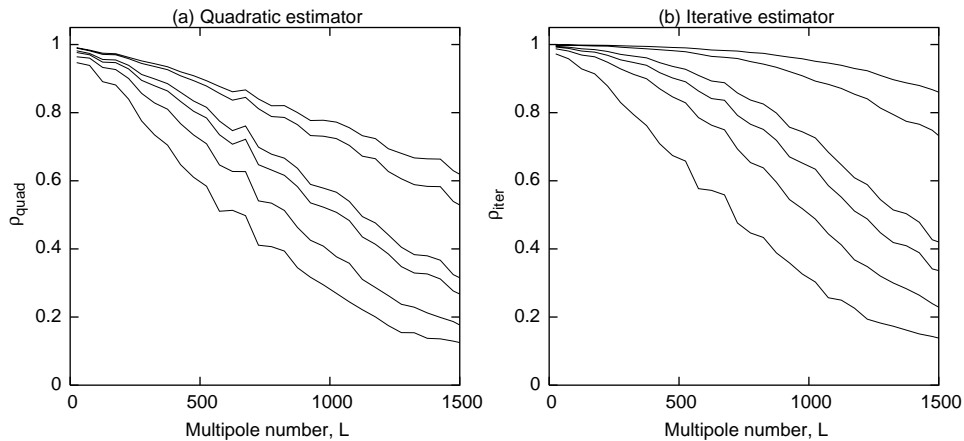Our iterative algorithm for solving Eq. (61) is

FIG. 6. The correlation coefficient $\rho = \langle \kappa \hat{\kappa} \rangle / \langle \kappa^2 \rangle^{1/2} \langle \hat{\kappa}^2 \rangle^{1/2}$ between the estimated and reconstructed lensing convergences as a function of multipole l, as determined in simulations. The correlation coefficients for the quadratic estimator are shown in panel (a); those for the iterative estimator are shown in panel (b). In both of these panels, the eight curves are for reference expts. A, B, C, D, E, and F (bottom to top) from Table I. Field rotation was not included in the calculations for this figure.
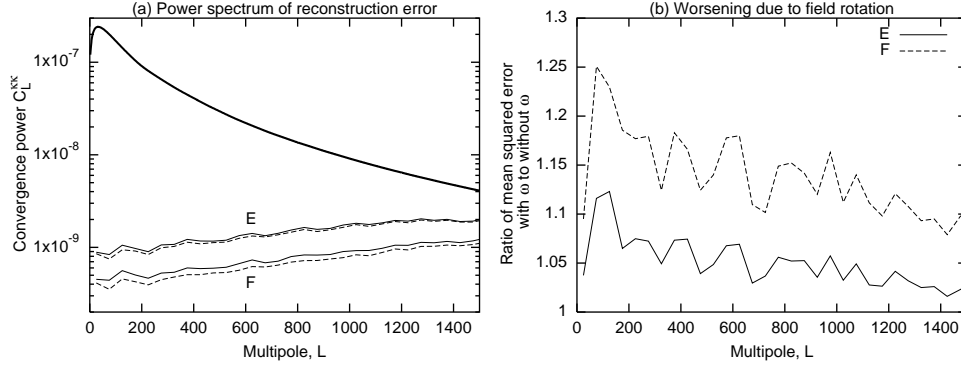
FIG. 7. The effect of field rotation on lensing estimation for Ref. expts. E and F. (a) Power spectra of the convergence κ (thick solid line), convergence error $\kappa - \hat{\kappa}$ with field rotation (thin solid lines), and convergence error $\kappa - \hat{\kappa}$ without field rotation (thin dashed lines). (b) The worsening of the reconstruction due to the presence of field rotation, as measured by the ratio of power spectra of the convergence errors: $\langle |(\kappa - \hat{\kappa})_l|^2 \rangle_{\text{with } \omega} / \langle |(\kappa - \hat{\kappa})_l|^2 \rangle_{\text{without } \omega}$. The same CMB, LSS, and noise (scaled appropriately to the experiment) realizations were used for all the simulations in this figure.

$$c_{\mu,n+1} = c_{\mu,n}^{1-\zeta_{(p)}} \left[c_{\mu,n} + \frac{d_\mu}{2\langle v_\mu(c_{\nu,n})\rangle_{LSS[c_{\nu,n}]}} \right.$$
$$\left. \times \left(\frac{v_\mu(c_{\nu,n})}{\langle v_\mu(c_{\nu,n})\rangle_{LSS[c_{\nu,n}]}} - 1 \right) \right]^{\zeta_{(p)}}, \quad (71)$$

where n represents the iteration number, and d_μ is the number of modes that fall into the μth band. It is readily apparent that the final values $c_{\nu,\infty}$ will satisfy $v_\mu = \langle v_\mu \rangle_{LSS}$.

In order to compute the convergence power spectrum estimator, the expected value $\langle v_\mu(c_{\nu,n})\rangle_{LSS[c_{\nu,n}]}$ must be determined; the simplest method for doing this is via Monte Carlo simulations. Since in the end we are solving the equation $v_\mu = \langle v_\mu \rangle_{LSS}$, we want to make sure that the Monte Carlo–induced error in the right-hand side of this equation is small compared with the statistical error in the left-hand side (which depends only on the data and on $c_{\nu,n}$). It is trivial to see that after N_{MC} Monte Carlo simulations, the variance in determination of the right-hand side is $1/N_{MC}$ of the statistical variance in the left-hand side. Therefore, we expect that if we use N_{MC} Monte Carlo simulations to determine $\langle v_\mu \rangle_{LSS}$, then the variance of our determination of the convergence power spectrum will increase by a factor of $1+1/N_{MC}$. A reasonable choice, then, is to take $N_{MC}=3$, which results in 15% increase in the variance of the power spectrum estimator over the case of $N_{MC}=\infty$ (exact computation of $\langle v_\mu \rangle_{LSS}$). The uncertainty in the power spectrum estimation can then be estimated from Eq. (62) with the correction for N_{MC}:

$$\sigma_{c_\mu} = \frac{C_l^{\kappa\kappa\ 2}}{C_l^{\kappa*\kappa*}} \sqrt{\frac{2}{d_\mu}(1+N_{MC}^{-1})}. \quad (72)$$

(Note that this is the standard Gaussian formula for error bars, except that it is corrected for N_{MC} and is written in terms of the filtered power spectrum $C_l^{\kappa*\kappa*}$ instead of the noise power.)

In Fig. 8(a), we show a determination of the convergence power spectrum from simulated data using Ref. expt. C noise parameters. The choice of bins was $N_{bin}=32$, $\Delta l=50$, $l_{\max}=1600$, and the survey area was 0.355 steradians (2048 arcmin $\times$ 2048 arcmin, with toroidal periodic boundary conditions). We initialized the power spectrum estimation with the white spectrum $C_l^{\kappa\kappa}=1\times10^{-9}$, corresponding to $c_{\mu,0}=10^{-9}$. We used $\zeta_{(p)}=0.5$ for the first two iterations of Eq. (71), which are sufficient to bring the estimated power spectrum $c_{\mu,n=2}$ to the correct order of magnitude. Once this "ballpark" estimation has been completed, we used $\zeta_{(p)}=1$ for the subsequent ($n\geq2$) iterations.

An examination of Fig. 8(a) shows that the power spectrum estimator Eq. (71) has been successful in reproducing the qualitative features of the power spectrum; however, the power has evidently been overestimated at the high-l end. We can perform a quantitative analysis of the performance of the power spectrum estimator using the χ^2 test, using the Gaussian error estimate of Eq. (72). The χ^2 value for the $l<1000$ region (where the power spectrum determination should be cosmic-variance limited) is $\chi^2=25.46$ for 20 degrees of freedom ($p=0.18$), indicating that Eq. (72) appears to be giving a reasonable estimate of the uncertainty on the power spectrum.

The same is not true of the high-l region $1000\leq l<1600$, for which we compute $\chi^2=53.05$ for 12 degrees of freedom ($p=4\times10^{-7}$). It is readily apparent from Fig. 8 that the failure of the χ^2 test is due to an upward bias in the power spectrum estimator Eq. (71). This bias occurs because, regardless of c_μ, our power spectrum estimator assumes that there is no convergence power at $l\geq1600$. However, it still detects the B modes induced by this short-wavelength convergence power, and introduces excess convergence power at $l<1600$ to reproduce these B modes — hence $C_l^{\kappa\kappa}$ is overestimated. This bias can be removed by either estimating the modes to higher l or including the aliasing of power from those modes into the modes we estimate as an additional source of noise [26]. This explanation of the upward bias is confirmed by Fig. 8(b), in which we have artificially "turned
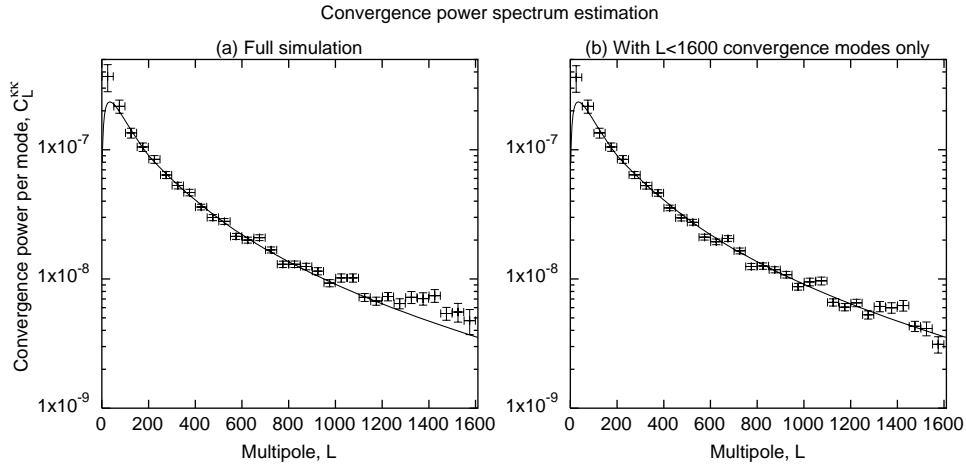
Convergence power spectrum estimation

FIG. 8. (a) Simulated convergence power spectrum estimation from Ref. expt. C with solid angle 0.355 steradians. The solid curve is the fiducial model $C_l^{\kappa\kappa}$; the points are the convergence power spectrum measured from simulated data after 5 iterations. (b) The same, except that the $l \geqslant 1600$ convergence modes were ignored in producing the simulated data (however, exactly the same CMB+LSS+noise realization was used). The horizontal error bars indicate the widths of the bins, while the vertical error bars are the 1σ measurement uncertainties according to Eq. (72). Note that the vertical error bars include the Monte Carlo error associated with using $N_{MC}=3$ simulations to determine $\langle v_\mu \rangle$; if we had calculated $\langle v_\mu \rangle$ exactly ($N_{MC} \gg 1$), the vertical error bars would be 13% smaller (see text for details).

off'' the lensing effect for convergence modes at $l \geqslant 1600$, then produced a simulated data set and applied the power spectrum estimator Eq. (71). In order to make the comparison between the original simulation and the restricted (i.e. short-wavelength lensing turned off) simulation as simple as possible, we have used the same CMB, noise, and LSS realizations for both. One can see by comparing Figs. 8(a) and 8(b) that there is little effect at low l, where the power spectrum estimation is limited by cosmic variance. However, at high l, one can see that the bias present in the original simulation has disappeared in the restricted simulation, thereby confirming that the bias was due to high-l convergence power. The χ^2 for the $1000 \leqslant l < 1600$ range has been reduced to $\chi^2=31.36$ (12 degrees of freedom, $p=1.7 \times 10^{-3}$), which is still indicative of underestimation of the uncertainty in the c_μ. Thus we conclude that in this regime, either the Gaussian error estimate Eq. (72) is underestimating the error by a factor of $\sim \sqrt{31/12} \approx 1.6$, or the error bars are correlated, or the iteration of Eq. (71) has not completely converged.

In a real lensing experiment, the underlying primary power spectrum C_l^{EE} is unknown and only the lensed power spectrum is directly observable (and even our knowledge of this is limited by instrument noise and cosmic variance). Thus a slightly more complicated version of the above analysis will be necessary to simultaneously solve for C_l^{EE} and $C_l^{\kappa\kappa}$. (Although since in the regime we are examining, $l <3000$, the E power spectrum is dominated by primary anisotropies rather than lensing, we do not expect a degeneracy between these two quantities.) It will also be necessary to estimate the convergence power spectrum well beyond the region of interest in order to avoid the upward bias described here. Since the signal-to-noise ratio at high l is low, it will be necessary to use wider bins (i.e. larger Δl) in this region. The choice of exactly which bins to use must be determined by the characteristics of the specific experiment.

VII. CONCLUSIONS

Weak gravitational lensing of the CMB allows us to reconstruct the (projected) mass distribution in the universe, thereby probing large-scale structure and its power spectrum. Since the window functions for lensing peak at redshift z of order unity, lensing offers the possibility of using the CMB to study the low-redshift universe [1–5]. "Cleaning" of lensing from CMB maps is potentially valuable for studying the primary CMB, particularly for inflationary gravitational wave searches using the low-l B-mode polarization [11,12]. Since the primary CMB polarization is expected to contain only E modes on the relevant angular scales (l of order 10^3), while lensing transfers some of the CMB polarization power into B modes [9], all B modes that we see on these scales are due to lensing (or foregrounds). Thus the CMB B-mode polarization allows much better lensing reconstruction than is possible using temperature data alone. It is thus of interest to consider optimal methods of reconstructing the lensing field from CMB polarization data; in this paper, we have investigated this problem in detail and improved significantly on the previous quadratic estimator methods [10]. We have shown that this improvement can be up to an order of magnitude in mean squared error over the zero-noise reconstruction error for the quadratic estimator.

We make several comments concerning the present calculations. First of all, our lensing estimator, Eq. (21), while statistically superior to the quadratic estimator, still does not achieve the Cramer-Rao bound on reconstruction accuracy. We have argued that this results in part from "curvature corrections," fluctuations in the curvature matrix that render the Cramer-Rao bound impossible to achieve (more generally, this should also serve as a warning against blindly assuming that the statistical errors in any measurement are given by $\mathbf{F}^{-1}$.). We expect that our lensing reconstruction estimator is near-optimal since it is an approximation to the maximum-

likelihood estimator and our iterative estimator shows no signs of incomplete convergence, however the possibility of further improvement has not been ruled out.

Secondly, we have assumed negligible primary B mode polarization here (although the formalism described herein is trivially modified to include significant primary B modes, the results would be qualitatively different). In the absence of vector or tensor perturbations, this is correct; if vector or tensor perturbations are present, then one must consider their effect on lensing reconstruction. In the case of inflationary gravitational waves, primary B modes are generated mostly on very large angular scales; the arcminute-scale anisotropies used for lensing reconstruction are uncontaminated [10]. (Formally, if we were doing a lensing reconstruction with the objective of cleaning lensing contamination of the tensor-induced reionization bump at $l < 20$, we would set $N_l^{BB} = \infty$ for $l < 20$ so that the lensing reconstruction does not remove tensor B modes.) A more rigorous investigation of the effect on inflationary gravitational wave searches is deferred to future work.

Thirdly, the real CMB is contaminated by foregrounds — an important issue for all CMB anisotropy experiments. One advantage of using CMB polarization for lensing reconstruction is that whereas the small-scale CMB temperature field is heavily contaminated by scattering-induced secondary anisotropies such as the thermal and kinetic Sunyaev-Zel'dovich effects, Ostriker-Vishniac effect, and patchy reionization, these effects are much smaller for polarization [27]. However, polarized point sources and galactic foregrounds are still a serious concern. These have very different frequency dependence than the blackbody fluctuations characteristic of the CMB, and this property has been exploited to remove them; unfortunately, their fluctuation spectrum, degree of polarization, non-Gaussianity, and variations in frequency dependence are poorly understood. Galactic foregrounds do not correlate with the cosmological signals, and in this sense the residuals from their subtraction act like instrument noise contaminating the B-mode (and, to a lesser extent, E-mode) polarization. The foreground power spectrum is likely to be different from that of instrument noise and is variable across the sky; nevertheless, if the covariance matrix of the foregrounds (or residuals after foreground subtraction) can be determined, then we can add the foreground covariance to the instrument noise covariance matrix $\mathbf{N}$. (If the statistical properties of the foreground residuals cannot be determined or at least constrained, then any cosmological analysis is pointless regardless of the methods used.) Polarized point sources produce Poisson noise; also since many of them are extragalactic, one could be concerned about their correlation with LSS and hence the lensing signal. We leave a detailed study of foregrounds and their impact on lensing reconstruction to future investigation. We note that the predicted levels of foreground contamination from dust and synchrotron galactic emission are at a level of a few μK arcmin prior to any frequency cleaning [28], comparable to the noise levels discussed here. Frequency cleaning should reduce this, at the expense of amplifying instrument noise. If foreground removal is inadequate, this may result in anomalies in the final results such as unphysical correlations between the con-

vergence maps and CMB polarization, variation of the convergence power spectrum between "clean" and "dirty" portions of the sky, correlation of the convergence maps with synchrotron or dust emission, etc.

In summary, we have shown that taking into account the full likelihood function allows improved reconstruction of the lensing of the CMB polarization field over that achieved by quadratic statistics. For purposes of computing the lensing power spectrum or cross-correlating CMB lensing with another tracer of the cosmological density field, the most important improvement is at high l where earlier approaches do not reconstruct the convergence at high signal-to-noise. (At low l, the reconstruction is already cosmic variance limited.) If one's objective is to clean out the lensing effect in search of primordial gravitational waves, then the relevant quantity is the residual error in the reconstruction, and it is important to reduce this even if the convergence has been mapped at high signal-to-noise; hence improvement at all multipoles is useful. We conclude that the likelihood-based estimators developed here offer the best prospective so far to extract the full amount of information from future high-resolution CMB polarization experiments.

APPENDIX A: QUADRATIC ESTIMATOR

In our simulations, we have compared the error of our iterative estimator, Eq. (27), with that of the quadratic estimator, Eq. (37). Here we show that the latter estimator corresponds to the optimally weighted quadratic estimator, as proposed by Ref. [10]. Statistically isotropic noise is assumed throughout.

We begin by expanding Eq. (38) using the formula for $\mathbf{f}_{\mathbf{l}}^{\kappa}$ given by Eq. (33). The off-diagonal elements vanish by symmetry, while the diagonal elements are

$$
\begin{aligned}
F_l^{\text{(quad)}} = \frac{1}{2} \sum_{\mathbf{l}_1} \ \mathrm{Tr}\{ [\mathbf{f}_{-\mathbf{l}}^{\kappa}]_{-\mathbf{l}_1,\mathbf{l}_2} [\langle \hat{\mathbf{C}} \rangle_{LSS}]_{l_2}^{-1} \\
\times [\mathbf{f}_{\mathbf{l}}^{\kappa}]_{\mathbf{l}_2, -\mathbf{l}_1} [\langle \hat{\mathbf{C}} \rangle_{LSS}]_{l_1}^{-1} \},
\end{aligned}
\tag{A1}
$$

where we have defined $\mathbf{l}_2 = \mathbf{l} - \mathbf{l}_1$, and the inverses are 3×3 matrix inverses (using the $\{T, E, B\}$ basis). Next we note that to first order in κ, the correlation between two Fourier modes of temperature or polarization is

$$
\langle \hat{\mathbf{x}}_{\mathbf{l}_1} \hat{\mathbf{x}}_{-\mathbf{l}_2}^{\dagger} \rangle = [\mathbf{f}_{\mathbf{l}}^{\kappa}]_{\mathbf{l}_1, -\mathbf{l}_2} \kappa_{\mathbf{l}} + O(\kappa^2).
\tag{A2}
$$

A general quadratic estimator for the convergence κ is then constructed as

$$\hat{\kappa}_{\mathbf{l}} = \sum_{\mathbf{l}_1} \hat{\mathbf{x}}^{\dagger}_{-\mathbf{l}_2} [\Xi_{-\mathbf{l}}]_{-\mathbf{l}_2,\mathbf{l}_1} \hat{\mathbf{x}}_{\mathbf{l}_1}, \qquad (A3)$$

where $\Xi_{-\mathbf{l}}$ is the weight matrix, which we assume without loss of generality to be Hermitian (since the anti-Hermitian part does not contribute to $\hat{\kappa}_{\mathbf{l}}$). We further require it to satisfy $[\Xi_{-\mathbf{l}}]_{-\mathbf{l}_2,\mathbf{l}_1} = [\Xi_{\mathbf{l}}]^{\dagger}_{\mathbf{l}_2,-\mathbf{l}_1}$. (This guarantees that the estimate $\hat{\kappa}$ is a real field.) We can construct the optimally weighted unbiased (to first order) estimator for κ by minimizing the variance of the estimator (neglecting the trispectrum contribution)

$$V_l = \langle |\hat{\kappa}_{\mathbf{l}}|^2 \rangle_{LSS}$$

$$\approx \sum_{\mathbf{l}_1} \mathrm{Tr}\{[\langle \hat{\mathbf{C}} \rangle_{LSS}]_{l_2} [\Xi_{-\mathbf{l}}]_{-\mathbf{l}_2,\mathbf{l}_1} [\langle \hat{\mathbf{C}} \rangle_{LSS}]_{l_1} [\Xi_{\mathbf{l}}]_{\mathbf{l}_1,-\mathbf{l}_2}\},$$
$$(A4)$$

subject to the constraint that the estimator be unbiased to first order (i.e. have unit response):

$$1 = \sum_{\mathbf{l}_1} \mathrm{Tr}\{[\mathbf{f}^{\kappa}_{\mathbf{l}}]_{\mathbf{l}_1,-\mathbf{l}_2} [\Xi_{-\mathbf{l}}]_{-\mathbf{l}_2,\mathbf{l}_1}\}. \qquad (A5)$$

We may compute the minimum of Eq. (A4) constrained by Eq. (A5) using the method of Lagrange multipliers. The equation $\delta V_l + \lambda^{-1}\delta 1 = 0$ becomes

$$\sum_{\mathbf{l}_1} \mathrm{Tr}\Bigg\{ 2[\langle \hat{\mathbf{C}} \rangle_{LSS}]_{l_2} [\delta\Xi_{-\mathbf{l}}]_{-\mathbf{l}_2,\mathbf{l}_1} [\langle \hat{\mathbf{C}} \rangle_{LSS}]_{l_1} [\Xi_{\mathbf{l}}]_{\mathbf{l}_1,-\mathbf{l}_2}$$

$$+ \frac{1}{\lambda}[\mathbf{f}^{\kappa}_{\mathbf{l}}]_{\mathbf{l}_1,-\mathbf{l}_2} [\delta\Xi_{-\mathbf{l}}]_{-\mathbf{l}_2,\mathbf{l}_1}\Bigg\} = 0. \qquad (A6)$$

The solution to this (allowing $\delta\Xi_{-\mathbf{l}}$ to be arbitrary) is

$$[\Xi_{-\mathbf{l}}]_{-\mathbf{l}_2,\mathbf{l}_1} = \frac{1}{2\lambda}[\langle \hat{\mathbf{C}} \rangle_{LSS}]^{-1}_{l_2} [\mathbf{f}^{\kappa}_{\mathbf{l}}]_{\mathbf{l}_2,-\mathbf{l}_1} [\langle \hat{\mathbf{C}} \rangle_{LSS}]^{-1}_{l_1}.$$
$$(A7)$$

The correct normalization λ is obtained by substitution into Eq. (A5); it is easily seen to be $\lambda = F^{(\mathrm{quad})}_l$. The variance of this estimator in the absence of lensing, determined by substitution into Eq. (A4), is $1/F^{(\mathrm{quad})}_l$. The quadratic estimator we have used, Eq. (37), is then seen to be a Wiener-filtered version of Eq. (A7) with the optimized choice for Ξ, Eq. (A7), and its covariance Eq. (39) then follows from the theory of Wiener-filtering.

Hu and Okamoto [10] derive a quadratic estimator using essentially the same method outlined in this appendix. While they have chosen to separately optimize the different components of Ξ (TT, TE, EE, TB, and EB) and then combine these to form a "minimum variance" estimator, the end result of the optimal filtering must be the same. (Note that while our covariance response function $\mathbf{f}$ is the same as Hu and Okamoto's $\mathbf{f}$ aside from a factor of $l^2/2$ due to use of Φ vs κ as the fundamental field, we have used Ξ in place of their F/A to avoid confusion with the Fisher matrix.)

APPENDIX B: LENSING B-MODES AND IDEALIZED RECONSTRUCTION

The purpose of this appendix is to investigate the question of whether, in the absence of noise and field rotation, the equation $B_{\mathrm{unlensed}} = 0$ could be used to completely reconstruct the convergence field. We show that with probability 1, it is possible to reconstruct most of the convergence modes. There may remain a small number of convergence modes that cannot be reconstructed by this method. If we impose periodic boundary conditions, the fraction of the convergence modes that are in this category is at most of order $1/l_{\mathrm{max}}$; however, there may be fewer of these degenerate modes, or possibly none at all. We have not investigated more realistic survey topologies but we would expect the general result to be similar on scales small compared to the angular width of the survey. However, this seems mostly an academic point since zero noise is of course unrealistic, and there can be many almost-degenerate modes that spoil a reconstruction based on $B_{\mathrm{unlensed}} = 0$.

The B mode induced by lensing is, to first order,

$$B_{\mathbf{l},\mathrm{lensing}} = \frac{1}{\sqrt{4\pi}} \sum_{\mathbf{l}'} \left(\frac{2}{l'^2}\right) \mathbf{l}' \cdot (\mathbf{l} - \mathbf{l}') \sin 2\alpha \, E_{\mathbf{l}-\mathbf{l}'} \kappa_{\mathbf{l}'}$$

$$= \sum_{\mathbf{l}'} \mathcal{T}_{\mathbf{l},\mathbf{l}'} \kappa_{\mathbf{l}'}, \qquad (B1)$$

where $\mathcal{T}$ is a transfer matrix that is a linear function of E (and once again $\alpha = \phi_{\mathbf{l}} - \phi_{\mathbf{l}'}$). In the absence of noise, we may set $B_{\mathbf{l},\mathrm{lensing}}$ equal to the observed polarization $\hat{B}_{\mathbf{l}}$. We see that if N Fourier modes are considered, there are N linear equations for N unknowns $\kappa_{\mathbf{l}}$. (We do not consider $\mathbf{l} = 0$ modes since there does not exist a κ_0 mode, and lensing has no effect on zero-wave-number CMB modes.) Thus any convergence mode that cannot be reconstructed must be associated with a degenerate direction of $\mathcal{T}$. It is clear that for some realizations of the primary CMB, e.g. $E = 0$, $\mathcal{T}$ is massively degenerate. We thus wish to explore whether these singular realizations are "likely" or have probability zero. We will assume here that $C^{EE}_l > 0$ for all of the E modes so that "probability zero" and "measure zero" can be taken to be equivalent.

In order to do this, we consider the characteristic polynomial of $\mathcal{T}$:

$$P(\lambda;E) = \det(\mathcal{T} - \lambda\mathbf{1}) = \sum_{n=0}^{N} a_n(E)\lambda^n. \qquad (B2)$$

The determinant of an $N \times N$ matrix is a polynomial of degree N in the entries of the matrix, hence each $a_n(E)$ is a polynomial of order at most $N - n$ in the $E_{\mathbf{l}}$. We know from linear algebra that the roots and multiplicities of $P(\lambda;E)$ (viewed as a polynomial in λ) are precisely the eigenvalues and multiplicities of $\mathcal{T}$; in particular, the number of degenerate ($\lambda = 0$) modes is equal to the smallest value of n for which $a_n(E) \neq 0$. Now suppose it were the case for some n that $a_n(E) \neq 0$ with nonzero probability (recall that the primary CMB polarization E is a random variable). This implies

that $a_n(E) \neq 0$ is generic, i.e. only a small (measure zero) set of values of E give $a_n(E) = 0$. The significance of this result is that if we can exhibit even one possible polarization field E for which $a_n(E) \neq 0$, it follows that $a_n(E) \neq 0$ with unit probability for the real primary CMB polarization field. A similar statement holds for the number of degenerate convergence modes: if we can exhibit a possible polarization field with n degenerate modes, then it follows that with unit probability, the lensing field as reconstructed from the real CMB will have at most n degenerate modes. Conceptually, this means that the generic lensing reconstruction using $B_{\text{unlensed}} = 0$ cannot be more degenerate than any special case we exhibit. (It may, however, be less degenerate.)

We consider here the following very simple realization: take a sky with area 4π, square (with sidelength $\sqrt{4\pi}$), and with periodic boundary conditions. Suppose that only the E mode $E_{\mathbf{L}}$ where $\mathbf{L} = (2\pi/\sqrt{4\pi}, 0)$ (i.e. the longest-wavelength mode in the x direction) is nonzero. Now consider a degenerate convergence mode, i.e. one that does not contribute to B_{lensing}. From Eq. (B1), we see that the $B = 0$ requirement forces all of the convergence modes $\kappa_{\mathbf{l}'}$ to be zero except those for which $\sin 2\alpha = 0$, i.e. those for which $\mathbf{l}'$ is either parallel to or perpendicular to $\mathbf{l}' + \mathbf{L}$. The latter is impossible given the boundary conditions and the former requires $\mathbf{l}'$ to lie in the x direction. Thus, out of $O(l_{\max}^2)$ convergence modes, only the $O(l_{\max})$ modes with wavevector in the x direction cannot be reconstructed. Hence no more than a fraction $O(1/l_{\max})$ of the convergence modes are degenerate (i.e. cannot be reconstructed from $B_{\text{unlensed}} = 0$), and by the argument of the previous paragraph this must hold with probability 1 for the actual realization of the primary polarization field. Note that this is only an upper limit and the actual number of degenerate modes may be smaller, or even zero.

The problem of lensing reconstruction using $B_{\text{unlensed}} = 0$ has been considered previously using real-space methods by Ref. [5]. They derive the following equation for the lensing-induced B mode:

$$\frac{1}{2}\nabla^2 B_{\text{lensing}} = \gamma_U \nabla^2 Q - \gamma_Q \nabla^2 U + \nabla\gamma_U \cdot \nabla Q - \nabla\gamma_Q \cdot \nabla U,$$
(B3)

where (as above) γ_Q and γ_U are second derivatives of Φ. This is therefore a third-order partial differential equation for Φ. Reference [5] then performs a two-dimensional Taylor expansion of Φ and finds that some of the coefficients are not fixed by Eq. (B3). They thus determine that there exists a class of lensing potential modes that do not produce B modes purely by counting the number of equations and the number of variables to be calculated. The relationship between our approach and that of Ref. [5] is that we express $\mathcal{T}$ in the Fourier basis [Eq. (B1)], whereas they have expressed $\mathcal{T}$ in the Taylor polynomial $\{x^j y^k\}_{j,k=0}^{\infty}$ basis. These bases are not equivalent due to the differing boundary conditions assumed (the Fourier basis imposes periodic boundary conditions whereas the Taylor polynomial basis does not), and this leads to different conclusions regarding the number and character of degenerate modes. In the Fourier basis there are the same

number of equations as variables. Thus one cannot conclude from the counting argument alone that there is a degeneracy in the case of full-sky coverage. (The analysis on the sphere would be slightly more intricate since the sky is curved and the boundary conditions have different topology than in the flat-sky approximation; however, on scales $l \gg 1$ small compared to the curvature scale, we expect that the results presented here will still apply.)

APPENDIX C: CURVATURE CORRECTIONS

Our purpose in this appendix is to investigate in greater detail the mathematical structure of the curvature corrections, i.e. the increase in uncertainty in lensing reconstruction due to fluctuations of the curvature matrix. We show that the curvature correction has another interpretation: it represents the increased noise in the reconstruction of $\kappa_{\mathbf{l}}$ due to the presence of other lensing modes, $\kappa_{\mathbf{l}'}$ (where $\mathbf{l}' \neq \pm \mathbf{l}$), an effect studied in detail in Kesden *et al* [21], where it was called the "first-order noise contribution" and denoted by $N^{(1)}_{XX', X''X'''}(L)$. Here we show that in fact the curvature correction contains the likelihood analysis manifestation of the $N^{(1)}$ of Ref. [21].

We can compute the curvature corrections to this as follows. If we define $\delta\mathcal{F} = \mathcal{F} - \mathbf{F}$, and retain our approximation from Eq. (32) that $\mathbf{F}$ is independent of κ, then we have

$$\mathbf{S} = \langle (\delta\mathcal{F} + \mathbf{F} + \mathbf{C}^{\kappa\kappa - 1})^{-1} \rangle_{LSS} \sim \mathbf{S}_0 + \mathbf{S}_0 \langle \delta\mathcal{F} \, \mathbf{S}_0 \, \delta\mathcal{F} \rangle_{LSS} \mathbf{S}_0$$
$$- \cdots,$$
(C1)

where $\mathbf{S}_0 = (\mathbf{F} + \mathbf{C}^{\kappa\kappa - 1})^{-1}$. [To derive this equation, we have merely Taylor-expanded in $\delta\mathcal{F}$, then taken the CMB + noise + LSS ensemble average, and noted that by definition $\delta\mathcal{F}$ vanishes when ensemble-averaged over CMB + noise realizations. Note that because we have taken the expectation value, Eq. (C1) should be viewed as an asymptotic expansion rather than a Taylor expansion.] The mean squared error $\mathbf{S}$ picks up "curvature correction" terms involving $\delta\mathcal{F}$ that cause it to not equal the naive result $\mathbf{S}_0$. Note that curvature corrections to $\mathbf{S}$ only increase $\mathbf{S}$, they cannot decrease it (in the sense that $\mathbf{S} - \mathbf{S}_0$ has all eigenvalues ≥ 0; equivalently the diagonal elements $\mathbf{S}_{jj} \geq [\mathbf{S}_0]_{jj}$ in all orthonormal bases). This is true to all orders in $\delta\mathcal{F}$ because the inverse of the mean of a set of positive definite Hermitian matrices is smaller than the mean of the inverses (in this same sense).

In order to compute the second-order curvature correction explicitly, we must understand the fluctuations in the curvature matrix. For simplicity, we evaluate $\delta\mathcal{F}$ at $\kappa = 0$. In this case, we find

$$\mathcal{F}_{\mathbf{l},\mathbf{l}'} \equiv \mathcal{F}[\kappa_{\mathbf{l}}, \kappa_{\mathbf{l}'}] = \frac{\partial^2}{\partial\kappa_{\mathbf{l}}^* \partial\kappa_{\mathbf{l}'}} \left(\frac{1}{2}\ln \det \hat{\mathbf{C}}_g + \frac{1}{2}\hat{\mathbf{x}}^\dagger \hat{\mathbf{C}}_g^{-1} \hat{\mathbf{x}} \right)$$

$$= \frac{1}{2}\hat{\mathbf{x}}^\dagger \hat{\mathbf{C}}^{-1} \left(2\frac{\partial\hat{\mathbf{C}}}{\partial\kappa_{-\mathbf{l}}}\hat{\mathbf{C}}^{-1}\frac{\partial\hat{\mathbf{C}}}{\partial\kappa_{\mathbf{l}'}} - \frac{\partial^2\hat{\mathbf{C}}}{\partial\kappa_{-\mathbf{l}}\partial\kappa_{\mathbf{l}'}} \right) \hat{\mathbf{C}}^{-1}\hat{\mathbf{x}}$$

$$- \frac{1}{2}\mathrm{Tr}\left[\hat{\mathbf{C}}^{-1}\left(\frac{\partial\hat{\mathbf{C}}}{\partial\kappa_{-\mathbf{l}}}\hat{\mathbf{C}}^{-1}\frac{\partial\hat{\mathbf{C}}}{\partial\kappa_{\mathbf{l}'}} - \frac{\partial^2\hat{\mathbf{C}}}{\partial\kappa_{-\mathbf{l}}\partial\kappa_{\mathbf{l}'}} \right) \right], \quad (C2)$$

where the second line has the $\hat{\mathbf{C}}$'s evaluated at $\kappa=0$. Subtracting out the average value $\mathbf{F}_{\mathbf{l},\mathbf{l}'}$ over CMB+noise realizations yields

$$\delta\mathcal{F}_{\mathbf{l},\mathbf{l}'}=-\frac{1}{2}\hat{\mathbf{x}}^{\dagger}\hat{\mathbf{C}}^{-1}\mathbf{J}_{(-\mathbf{l},\mathbf{l}')}\hat{\mathbf{C}}^{-1}\hat{\mathbf{x}}+\frac{1}{2}\mathrm{Tr}(\hat{\mathbf{C}}^{-1}\mathbf{J}_{(-\mathbf{l},\mathbf{l}')}),$$
(C3)

where we have defined

$$\begin{aligned}\mathbf{J}_{(-\mathbf{l},\mathbf{l}')}=\hat{\mathbf{C}}\frac{\partial^{2}[\hat{\mathbf{C}}^{-1}]}{\partial\kappa_{-\mathbf{l}}\partial\kappa_{\mathbf{l}'}}\bigg|_{\kappa=0}&=\left(\frac{\partial\hat{\mathbf{C}}}{\partial\kappa_{-\mathbf{l}}}\hat{\mathbf{C}}^{-1}\frac{\partial\hat{\mathbf{C}}}{\partial\kappa_{\mathbf{l}'}}\right.\\&\left.+\frac{\partial\hat{\mathbf{C}}}{\partial\kappa_{\mathbf{l}'}}\hat{\mathbf{C}}^{-1}\frac{\partial\hat{\mathbf{C}}}{\partial\kappa_{-\mathbf{l}}}-\frac{\partial^{2}\hat{\mathbf{C}}}{\partial\kappa_{-\mathbf{l}}\partial\kappa_{\mathbf{l}'}}\right)\bigg|_{\kappa=0}.\end{aligned}$$
(C4)

Note that $\mathbf{J}_{(-\mathbf{l},\mathbf{l}')}=\mathbf{J}_{(\mathbf{l}',-\mathbf{l})}=\mathbf{J}_{(\mathbf{l},-\mathbf{l}')}^{\dagger}$.

Using this relation, and noting that at $\kappa=0$ we have $\partial\hat{\mathbf{C}}/\partial\kappa_{\mathbf{l}}=\mathbf{f}_{\mathbf{l}}^{\kappa}$, we can use Wick's theorem to compute

$$\langle\delta\mathcal{F}_{\mathbf{l},\mathbf{l}'}\delta\mathcal{F}_{\mathbf{l}_{1},\mathbf{l}'_{1}}\rangle=\frac{1}{4}\mathrm{Tr}(\hat{\mathbf{C}}^{-1}\mathbf{J}_{(-\mathbf{l},\mathbf{l}')}\hat{\mathbf{C}}^{-1}\mathbf{J}_{(-\mathbf{l}',\mathbf{l})}).$$
(C5)

In the case of statistically isotropic noise, Eq. (C5) allows us to compute the covariance of the reconstruction $\mathbf{S}$ using Eq.

(C1). In harmonic space, the off-diagonal elements of $\mathbf{S}$ vanish by symmetry whereas the diagonal elements are given by

$$S_{l}=[S_{0}]_{l}+\frac{1}{4}[S_{0}]_{l}^{2}\sum_{\mathbf{l}'}[S_{0}]_{l'}\mathrm{Tr}(\hat{\mathbf{C}}^{-1}\mathbf{J}_{(-\mathbf{l},\mathbf{l}')}\hat{\mathbf{C}}^{-1}\mathbf{J}_{(-\mathbf{l}',\mathbf{l})}),$$
(C6)

which is a summation over quadrilateral configurations of the modes $\mathbf{l}$, $\mathbf{l}'$, and the mode over which we sum when computing the trace.

To lowest order ($\delta\mathcal{F}^{2}$), the curvature correction is given by Eq. (C6). The correction to the mean inverse curvature V_{l}, i.e. to the covariance matrix of an unbiased estimator for κ, is related to the correction to S_{l} by noting that $S_{l}^{-1}=V_{l}^{-1}+C_{l}^{\kappa\kappa-1}$, hence

$$\Delta V_{l}=\frac{V_{l}^{2}}{S_{l}^{2}}\Delta S_{l}\approx\frac{1}{4}V_{l}^{2}\sum_{\mathbf{l}'}[S_{0}]_{l'}\mathrm{Tr}(\hat{\mathbf{C}}^{-1}\mathbf{J}_{(-\mathbf{l},\mathbf{l}')}\hat{\mathbf{C}}^{-1}\mathbf{J}_{(-\mathbf{l}',\mathbf{l})}).$$
(C7)

We now pass to the "linear approximation" in which the second derivative $\partial^{2}\hat{\mathbf{C}}/\partial\kappa_{\mathbf{l}}^{*}\partial\kappa_{\mathbf{l}'}$ is neglected. (This was found to be a valid approximation for temperature-based lensing estimation on scales $l<3500$ [17], although it is unclear whether this is also true in the present context.) Substituting Eq. (C4) then yields

$$\Delta V_{l}\approx\frac{1}{2}V_{l}^{2}\sum_{\mathbf{l}'}[S_{0}]_{l'}\mathrm{Tr}(\hat{\mathbf{C}}^{-1}\mathbf{f}_{\mathbf{l}}^{\kappa}\hat{\mathbf{C}}^{-1}\mathbf{f}_{\mathbf{l}'}^{\kappa}\hat{\mathbf{C}}^{-1}\mathbf{f}_{-\mathbf{l}}^{\kappa}\hat{\mathbf{C}}^{-1}\mathbf{f}_{-\mathbf{l}'}^{\kappa}+\hat{\mathbf{C}}^{-1}\mathbf{f}_{\mathbf{l}}^{\kappa}\hat{\mathbf{C}}^{-1}\mathbf{f}_{-\mathbf{l}}^{\kappa}\hat{\mathbf{C}}^{-1}\mathbf{f}_{\mathbf{l}'}^{\kappa}\hat{\mathbf{C}}^{-1}\mathbf{f}_{-\mathbf{l}'}^{\kappa}).$$
(C8)

This should be compared with the first-order noise contribution $N^{(1)}$ of Ref. [21]. In our notation, and written in terms of the convergence rather than the potential, their Eq. (25) can be rewritten with the help of some algebra and the relation $\mathbf{f}_{\mathbf{l}}^{\kappa}{}^{\dagger}=\mathbf{f}_{-\mathbf{l}}^{\kappa}$ as

$$N_{TT,TT}^{(1)}(\kappa_{\mathbf{l}})=\frac{1}{2}V_{l}^{2}\sum_{\mathbf{l}'}C_{l'}^{\kappa\kappa}\sum_{\mathbf{l}_{1}}\frac{[f_{\mathbf{l}}]_{\mathbf{l}_{1},-\mathbf{l}_{2}}^{\kappa}[f_{\mathbf{l}'}]_{-\mathbf{l}_{2},-\mathbf{l}_{2}-\mathbf{l}'}^{\kappa}[f_{-\mathbf{l}}]_{-\mathbf{l}_{2}-\mathbf{l}',\mathbf{l}_{1}-\mathbf{l}'}^{\kappa}[f_{-\mathbf{l}'}]_{\mathbf{l}_{1}-\mathbf{l}',\mathbf{l}_{1}}^{\kappa}}{[\langle\hat{\mathbf{C}}\rangle_{LSS}]_{l_{1}}^{\kappa}[\langle\hat{\mathbf{C}}\rangle_{LSS}]_{l_{2}}^{\kappa}[\langle\hat{\mathbf{C}}\rangle_{LSS}]_{l_{2}+\mathbf{l}'}^{\kappa}[\langle\hat{\mathbf{C}}\rangle_{LSS}]_{\mathbf{l}_{1}-\mathbf{l}'}^{\kappa}}.$$
(C9)

This is the first term of Eq. (C8), except that it only includes temperature information (hence we have multiplication of numbers rather than 3×3 matrices), and the residual power spectrum $[S_{0}]_{l'}$ has been replaced with the raw convergence power spectrum $C_{l'}^{\kappa\kappa}$. (The latter difference arises because Ref. [21] computed the first-order noise $N^{(1)}$ for the quadratic estimator; when the Bayesian estimator is used, the contaminating modes $\{\kappa_{\mathbf{l}'}\}$ have their power reduced from $C_{l'}^{\kappa\kappa}$ to $[S_{0}]_{l'}$ since the estimated lensing field $\hat{\kappa}$ is used to delens the CMB.) Thus we see that the first-order noise arises in the likelihood formalism as a curvature correction, which is not taken into account in the Fisher matrix for lensing reconstruction.

The question naturally arises as to the interpretation of the second term in the curvature correction, Eq. (C8). We note that within the linear approximation,

$$\langle\hat{\mathbf{C}}^{-1}\rangle_{LSS}=\hat{\mathbf{C}}_{(0,0)}^{-1}+\sum_{\mathbf{l}'}[S_{0}]_{\mathbf{l}'}\hat{\mathbf{C}}_{(0,0)}^{-1}\mathbf{f}_{\mathbf{l}'}^{\kappa}\hat{\mathbf{C}}_{(0,0)}^{-1}\mathbf{f}_{-\mathbf{l}'}^{\kappa}\hat{\mathbf{C}}_{(0,0)}^{-1}.$$
(C10)

The second term of Eq. (C8) is thus seen to be the correction to the Fisher matrix, Eq. (32), due to the lensing effect on the CMB power spectrum.

[1] F. Bernardeau, Astron. Astrophys. **324**, 15 (1997).

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

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

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

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

[6] H. V. Peiris and D. N. Spergel, Astrophys. J. **540**, 605 (2000).

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

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

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

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

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

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

[13] S. Cole and G. Efstathiou, Mon. Not. R. Astron. Soc. **239**, 195 (1989).

[14] U. Seljak, Astrophys. J. **463**, 1 (1996).

[15] J. Guzik, U. Seljak, and M. Zaldarriaga, Phys. Rev. D **62**, 043517 (2000).

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

[17] C. M. Hirata and U. Seljak, Phys. Rev. D **67**, 043001 (2003).

[18] U. Seljak and M. Zaldarriaga, Astrophys. J. **538**, 57 (2000).

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

[20] U. Seljak and M. Zaldarriaga, Astrophys. J. **469**, 437 (1996).

[21] M. Kesden, A. Cooray, and M. Kamionkowski, Phys. Rev. D **67**, 123507 (2003).

[22] A. Cooray and W. Hu, Astrophys. J. **574**, 19 (2002).

[23] S. M. Carroll, W. H. Press, and E. L. Turner, Annu. Rev. Astron. Astrophys. **30**, 499 (1992).

[24] J. A. Peacock and S. J. Dodds, Mon. Not. R. Astron. Soc. **280**, L19 (1996).

[25] T. Okamoto and W. Hu, Phys. Rev. D **67**, 083002 (2003).

[26] U. Seljak, Astrophys. J. **503**, 492 (1998).

[27] W. Hu, Astrophys. J. **529**, 12 (2000).

[28] M. Tegmark, D. J. Eisenstein, W. Hu, and A. de Oliveira-Costa, Astrophys. J. **530**, 133 (2000).