Approximate likelihood approaches for detecting the influence of primordial gravitational waves in cosmic microwave background polarization

Zhen Pan,[1,*] Ethan Anderes,[2,†] and Lloyd Knox[1,‡]

[1]Department of Physics, University of California, One Shields Avenue, Davis, California 95616, USA
[2]Department of Statistics, University of California, One Shields Avenue, Davis, California 95616, USA

(Received 3 September 2017; published 17 May 2018)

One of the major targets for next-generation cosmic microwave background (CMB) experiments is the detection of the primordial B-mode signal. Planning is under way for Stage-IV experiments that are projected to have instrumental noise small enough to make lensing and foregrounds the dominant source of uncertainty for estimating the tensor-to-scalar ratio r from polarization maps. This makes delensing a crucial part of future CMB polarization science. In this paper we present a likelihood method for estimating the tensor-to-scalar ratio r from CMB polarization observations, which combines the benefits of a full-scale likelihood approach with the tractability of the quadratic delensing technique. This method is a pixel space, all order likelihood analysis of the quadratic delensed B modes, and it essentially builds upon the quadratic delenser by taking into account all order lensing and pixel space anomalies. Its tractability relies on a crucial factorization of the pixel space covariance matrix of the polarization observations which allows one to compute the full Gaussian approximate likelihood profile, as a function of r, at the same computational cost of a single likelihood evaluation.

DOI: 10.1103/PhysRevD.97.103512

I. INTRODUCTION

The inflation paradigm has successfully explained the origin of primordial density perturbations that grew into the cosmic microwave background (CMB) anisotropies and large scale structure we observe [e.g. [1–6]]. A key prediction of inflation is the background of primordial gravitational waves (GWs) or tensor-mode perturbations [e.g. [7–11]], which imprints a unique polarization pattern, called a primordial B mode, on the CMB anisotropies [12–17]. Further, detection of a nearly scale-invariant background of GWs would severely challenge noninflationary models [e.g. [18–22]]. The strength of primordial gravitational waves or tensor-mode power is commonly quantified by the tensor-to-scalar ratio r. Joint analysis of BICEP2/Keck Array and *Planck* data yields an upper bound $r < 0.12$ at 95% confidence level [23], the bound is slightly tightened when the *Planck* high-ℓ polarization data are included [24], and BICEP2/Keck Collaboration gives the latest upper bound $r < 0.09$ at 95% confidence level [25]. Fourth generation experiments, including COrE, LiteBird, and CMB Stage-IV, are expected to constrain r with uncertainty $\sigma(r) \simeq 0.001$ [26–32].

The primordial B modes are contaminated by several sources: emission from galactic dust and other foregrounds [33–40], instrumental noise, and gravitational lensing of scalar CMB perturbations. The B modes generated by gravitational lensing of the CMB have been detected [23,41–45]. The lensed B-mode power spectrum is nearly a constant at small multipoles ($\ell \lesssim 1000$) and therefore manifests as an effective white noise with amplitude ~ 5 μK-arc min [46,47]. For CMB Stage-IV, we expect to decrease the instrumental noise to ~ 1 μK-arc min [29]. Then, the lensing B noise (and foregrounds) would become the dominant noise source limiting the primordial B-mode survey.

Fortunately, the lensing B noise is well understood. Up to leading order, one can effectively delense observed B modes by utilizing a quadratic combination of observed E modes and an estimate of the lensing potential field $\phi^{\rm est}$ [47–51]. We find that the validity of the quadratic delenser crucially depends on a partial cancellation of higher order lensing terms (see Sec. IV C for details). However, in the regime of low instrumental noise and small lensing potential field estimate uncertainty, higher-order lensing terms, ignored by the quadratic delensing technique, can have an appreciable effect. These higher order terms not only induce a delensing bias but also contain information on primordial B modes. In addition, experimental complexities such as nonstationary noise and sky cuts become nontrivial for spectral-based methods such as the quadratic delenser.

As an alternative, a full-scale likelihood analysis of the tensor-to-scalar ratio r can, in principle, optimally account

*zhpan@ucdavis.edu
†anderes@ucdavis.edu
‡lknox@ucdavis.edu

for all the Gaussian and non-Gaussian information in the CMB observations. Unfortunately, a full likelihood analysis requires computation resources beyond what is available in the near future. In this paper, we introduce a likelihood approximation which is modified from the full-scale likelihood, so as to be computationally tractable.

We start with introducing a Gaussian likelihood incorporating all the 2-point information. A key element of our likelihood analysis is the covariance matrix of the polarization maps. For each data pair (d_i, d_j) (d can be Q or U), its covariance depends on the primordial polarization power spectra C_ℓ^{EE} and $C_\ell^{BB,r}$, lensing potential field $\phi(x)$ (and instrumental noise N^{QQ} and N^{UU}), where the E-mode power C_ℓ^{EE} has been well constrained [e.g. [52,53]], while the primordial B-mode signal has not been detected. We assume the tensor perturbations to be scale-invariant and Gaussian; scale invariance is a good approximation to the prediction of the single-field slow-roll inflation, for which the tensor spectral index $n_{\rm T} = -r/8$, with $n_{\rm T} = 0$ corresponding to scale invariance. Hence all the primordial B-mode information is encoded in the single parameter r, the tensor-to-scalar ratio at $k = 0.05\ {\rm Mpc}^{-1}$. Then the covariance matrix $\tilde{\Sigma}_{r,\phi}$ depends on the unknown parameter r and the underlying lensing potential field $\phi(x)$, where $\phi(x)$ can be estimated from exterior tracers, e.g. cosmic infrared background (CIB) [45,54–58] or from intrinsic CMB via, e.g. quadratic estimators [59,60] or Bayesian approach [61–66]. We obtain a covariance matrix Σ_r depending only on r by marginalizing $\tilde{\Sigma}_{r,\phi}$ over uncertainties in the lensing potential field estimate $\phi^{\rm est}$. With this full covariance matrix Σ_r, it is then straightforward to compute the likelihood of r for given data vector d by approximating d as a Gaussian vector.

In principle, this Gaussian likelihood method can exploit all the 2-point r information from the polarization maps, but usually the computation resource demands are still excessive. For example, to constrain r from some polarization maps with p pixels, we need to compute the full likelihood profile $L(r|d)$ as a function of r, which in practice requires computing the likelihood on a range of r values, say 50 values evenly distributed in the interval [0, 0.2]. For each different r, we need to compute the quadratic form $d^{\intercal}\Sigma_r^{-1}d$ and the determinant $\det(\Sigma_r)$, due to the r dependence of the covariance matrix. In any realistic experiments with $p \gtrsim 10^4$, it is a huge amount of work to compute and invert the covariance matrix of dimension $2p \times 2p$ for that many r values, where the factor 2 comes from two observables Q and U on each pixel.

In this paper, we present a modified Gaussian likelihood tailoring the full-scale likelihood analysis so as to be computationally tractable. The method consists of two parts. In the first part, we decompose the covariance matrix Σ_r as $\Sigma^{\rm en} + r\Sigma^{\rm b}$, where $\Sigma^{\rm en}$ is the contribution from E modes and instrumental noise, and $r\Sigma^{\rm b}$ is the contribution from B modes. This decomposition allows us to compute the covariance Σ_r, as a function of r, at the same computational cost of a single covariance matrix computation. In the second part, we suppress data size by tracking only s high signal-to-noise modes, say the large-scale quadratic delensed B modes. We project out the lensing-generated B modes and obtain the delensed modes $B^{\rm del}$ from the polarization data d via a projection matrix v, $({}^0B_{\boldsymbol{\ell}}^{\rm del})_s = (v^{\intercal})_{s\times 2p} d_{2p}$, with $s \sim 500$ and the upper left index ^{0} denoting the projected data vector limited to the s lowest frequency modes available. Then, the covariance matrix of the projected data vector ${}^0B_{\boldsymbol{\ell}}^{\rm del}$ is given by $v^{\intercal}\Sigma_r v$, with which the computation of the r likelihood $L(r|{}^0B_{\boldsymbol{\ell}}^{\rm del})$ given the projected data vector turns out to be tractable. This method can be naturally extended to incorporate higher frequency modes, as we describe in Sec. III.

We make no attempt here to address contamination from foregrounds. At every frequency, and even in the cleanest patches of sky, foreground emission is brighter than the tensor signal of interest [23,67]. Foregrounds also have the potential to contaminate the signals one relies on for delensing [68]. We effectively assume here that this contamination can be reduced to negligible levels by use of observations at multiple frequencies. The noise levels we assume in the following should be understood as post-foreground-cleaning noise in CMB maps.

The work presented here focuses exclusively on the problem of CMB delensing, as distinct from lensing estimation. Our approach uses approximate likelihood techniques to model, rather than remove, the higher order lensing residuals present after linear order E template subtraction. This contrasts with other higher order methods which remove lensing through iterative techniques that necessarily fuse delensing and internal lensing estimation together into a single procedure. Two benefits follow from our approach: (1) it makes higher order delensing techniques available in modular form which can be used with different combinations of lensing estimates; i.e., from external sources and/or internal estimates of the lensing map (including those obtained by iterative procedures or posterior samples), (2) it allows us to perform a comparison where the only difference in the two methods is the delensing aspect, simplifying interpretation; i.e., we know any differences are due to the delensing method rather than the lensing potential estimation method.

The paper is organized as follows. We introduce the quadratic delenser and the likelihood-based delenser in Sec. II and Sec. III, respectively. In Sec. IV, we apply the two r constraining techniques on simulations mimicking Stage III and IV CMB surveys, and compare their r constraints. We conclude with Sec. V. For reference, we derive the analytic expression for the covariance matrices of the polarization maps, and the eigenvalue method for inverting large matrices in Appendixes A, B and C, respectively.

Throughout this paper, we follow the notation of Lewis and Challinor [46], which is different from that of Hu and Okamoto [60] by factors of 2π.

II. QUADRATIC DELENSER

For simplicity, we assume no contamination of foregrounds throughout this paper. Then the observed B modes can generally be expressed as

$$B^{\rm obs} = B^r + B^{\rm len} + N^B, \tag{1}$$

where B^r, $B^{\rm len}$, N^B are primordial B signal, lensing B noise, and instrumental B noise, respectively. To constrain the primordial B signal, delensing is essential, where we obtain an estimate of the lensing B noise and subtract it off from the observed B modes. Here, we introduce a quadratic delenser.

Accurate to the leading order of the lensing potential $\phi_{\boldsymbol{\ell}}$, $B^{\rm len}$ is the convolution of the lensing potential and primordial E modes [46], i.e.,

$$B^{\rm len}_{\boldsymbol{\ell}} = \int \frac{d^2\boldsymbol{\ell}'}{2\pi} \boldsymbol{\ell}' \cdot (\boldsymbol{\ell} - \boldsymbol{\ell}') \sin(2\varphi_{\boldsymbol{\ell},\boldsymbol{\ell}'}) E_{\boldsymbol{\ell}'} \phi_{\boldsymbol{\ell}-\boldsymbol{\ell}'}. \tag{2}$$

Usually, the underlying lensing potential is not known *a priori*, but can be estimated from either intrinsic CMB or from external tracers. From an estimated lensing potential $\phi^{\rm est}_{\boldsymbol{\ell}}$ and observed modes $E^{\rm obs}_{\boldsymbol{\ell}}$, we construct a quadratic estimate of the lensing B noise

$$B^{\rm len,est}_{\boldsymbol{\ell}} = \int \frac{d^2\boldsymbol{\ell}'}{2\pi} f_{\boldsymbol{\ell},\boldsymbol{\ell}'} \boldsymbol{\ell}' \cdot (\boldsymbol{\ell} - \boldsymbol{\ell}') \sin(2\varphi_{\boldsymbol{\ell},\boldsymbol{\ell}'}) E^{\rm obs}_{\boldsymbol{\ell}'} \phi^{\rm est}_{\boldsymbol{\ell}-\boldsymbol{\ell}'}, \tag{3}$$

where $\varphi_{\boldsymbol{\ell},\boldsymbol{\ell}'} = \varphi_{\boldsymbol{\ell}} - \varphi_{\boldsymbol{\ell}'}$, $E^{\rm obs}$ is the observed E modes (to the lowest order, the difference between lensed E and primordial E can be neglected), and the weighting function $f_{\boldsymbol{\ell},\boldsymbol{\ell}'}$ is to be determined by minimizing the residual, $B^{\rm res}_{\boldsymbol{\ell}} = B^{\rm len}_{\boldsymbol{\ell}} - B^{\rm len,est}_{\boldsymbol{\ell}}$. If we define the correlation coefficient of $\phi_{\boldsymbol{\ell}}$ and $\phi^{\rm est}_{\boldsymbol{\ell}}$

$$\rho_\ell = \frac{C^{\phi,\phi^{\rm est}}_\ell}{\sqrt{C^{\phi\phi}_\ell C^{\phi^{\rm est}\phi^{\rm est}}_\ell}}, \tag{4}$$

the optimal weight at leading order was proved to be [47]

$$f_{\boldsymbol{\ell},\boldsymbol{\ell}'} = \frac{C^{EE}_{l'}}{C^{EE}_{l'} + N^{EE}_{l'}} \rho^2_{|\boldsymbol{\ell}-\boldsymbol{\ell}'|}, \tag{5}$$

which enables a minimal residual power spectrum

$$C^{BB,{\rm res}}_{\boldsymbol{\ell}} = \int \frac{d^2\boldsymbol{\ell}'}{(2\pi)^2} [\boldsymbol{\ell}' \cdot (\boldsymbol{\ell} - \boldsymbol{\ell}') \sin(2\varphi_{\boldsymbol{\ell},\boldsymbol{\ell}'})]^2 \times C^{EE}_{\boldsymbol{\ell}'} C^{\phi\phi}_{\boldsymbol{\ell}-\boldsymbol{\ell}'} (1 - f_{\boldsymbol{\ell},\boldsymbol{\ell}'}). \tag{6}$$

After subtracting off the template $B^{\rm len,est}_{\boldsymbol{\ell}}$, we obtain a quadratic delensed B-mode map

$$B^{\rm del}_{\boldsymbol{\ell}} = B^{\rm obs}_{\boldsymbol{\ell}} - B^{\rm len,est}_{\boldsymbol{\ell}} = B^r_{\boldsymbol{\ell}} + B^{\rm res}_{\boldsymbol{\ell}} + N^B_{\boldsymbol{\ell}}, \tag{7}$$

and its power spectrum

$$C^{BB,{\rm del}}_\ell = C^{BB,r}_\ell + C^{BB,{\rm res}}_\ell + N^{BB}_\ell. \tag{8}$$

From the delensed B modes, one can better constrain r due to the suppressed lensing B noise. Note that in the evaluation of the residual lensing B power $C^{BB,{\rm res}}_\ell$ of Eq. (6) we have made two approximations: 1) we keep only the linear order lensing in $B^{\rm len}$; 2) we completely ignore the lensing in $E^{\rm obs}$.

III. GAUSSIAN LIKELIHOOD DELENSER

In contrast to the quadratic delenser, the likelihood analysis works on observables $Q^{\rm obs}$ and $U^{\rm obs}$ in pixel space. Concatenating the polarization data on all pixels yields a length-$2p$ data vector

$$d = [Q^{\rm obs}(x_1) \cdots Q^{\rm obs}(x_p), U^{\rm obs}(x_1) \cdots U^{\rm obs}(x_p)]^{\intercal}, \tag{9}$$

with p being the number of pixels. We first evaluate the covariance matrix of the data vector, which depends on the primordial polarization power spectra C^{EE}_ℓ and $C^{BB,r}_\ell$, lensing potential field $\phi(x)$ (and instrumental noise). With C^{EE}_ℓ being well determined, and $C^{BB,r}_\ell$ being parametrized by the tensor-to-scalar ratio r, we marginalize the covariance matrix $\tilde{\Sigma}_{r,\phi}$ over uncertainties in $\phi(x)$ estimate, and obtain a covariance matrix Σ_r depending only on r. Then it is straightforward to compute the approximate likelihood of r for given data d by approximating d as a Gaussian vector, i.e.,

$$-2\log L(r|d) = d^{\intercal}\Sigma_r^{-1} d + \log\det\Sigma_r, \tag{10}$$

up to a constant term.

A. Comparison with the quadratic delenser

Before delving into the details of the Gaussian likelihood delenser, it would be useful to do a brief comparison with the quadratic delenser (see Table I):

(i) The quadratic delenser works on the delensed modes $B^{\rm del}_{\boldsymbol{\ell}}$ in Fourier space, and approximates these modes as stationary and Gaussian, i.e.,

TABLE I. A brief comparison of the two delensers.

Delenser	Quadratic delenser	Gaussian likelihood
Working space	Fourier	Pixel
Power spectrum/ covariance matrix	Leading order	All order
Nonstationarity	✗	✓
Non-Gaussianity	✗	✗

$$\begin{pmatrix} \vdots \\ B_\ell^{\rm del} \\ \vdots \end{pmatrix} \sim N\left(0, \begin{bmatrix} \ddots & 0 & 0 \\ 0 & C_\ell^{BB,{\rm del}} & \ddots \\ 0 & 0 & 0 \end{bmatrix}\right),$$

where the power spectrum $C_\ell^{BB,{\rm del}}$, derived in Eq. (8), only takes into account the leading order in ϕ. Therefore, the quadratic delenser exploits the 2-point information in a biased way by ignoring the nonstationarity and higher-order lensing in the power spectrum.

(ii) The Gaussian likelihood delenser works on the observables d in pixel space and approximates the data vector d as Gaussian after marginalizing over uncertainties in the ϕ estimate. In the computation of the covariance matrix Σ_r, all-order lensing is taken into account and no stationarity assumption is made. Therefore, the Gaussian likelihood delenser naturally incorporates all the 2-point information.

The Gaussian likelihood is potentially favored in several aspects, but usually is computationally excessive. As explained in the Introduction, the bottleneck of the likelihood analysis is the covariance matrix Σ_r related computation, which is of large size $2p \times 2p$, and is a function of r. Here we introduce a modified Gaussian likelihood method. The method consists of two parts, covariance decomposition and data compression, where the former allows us to compute the covariance matrix Σ_r, as a function of r, at the computation cost of a single covariance matrix computation, and the latter allows us to compress the covariance matrix by tracking a small number of high S/N modes.

B. Covariance decomposition

1. Decomposition

To avoid repeating the computation of the covariance matrix Σ_r for each different r, we find it is possible to single out the r dependence by decomposing the covariance matrix as

$$\Sigma_r = \Sigma^{\rm en} + r\Sigma^{\rm b}, \tag{11}$$

where $\Sigma^{\rm en}$ is the contribution from E modes and instrumental noise, and $r\Sigma^{\rm b}$ is the contribution from primordial B modes. With this decomposition, we can obtain the covariance matrix Σ_r as a function of r, as long as the r-independent components $\Sigma^{\rm en}$ and $\Sigma^{\rm b}$ are obtained.

For the covariance decomposition of Eq. (11), we first decompose observables $Q^{\rm obs}$ and $U^{\rm obs}$ as linear combinations of E modes and B modes. Stokes parameters Q and U are related to coordinate independent quantities E and B via [12–14,16]

$$\begin{aligned} Q_\ell &= -\cos(2\varphi_\ell)E_\ell + \sin(2\varphi_\ell)B_\ell, \\ U_\ell &= -\sin(2\varphi_\ell)E_\ell - \cos(2\varphi_\ell)B_\ell. \end{aligned} \tag{12}$$

We define the following modulated E/B modes

$$\begin{aligned} SE_\ell &\equiv -\sin(2\varphi_\ell)E_\ell, \qquad SB_\ell \equiv +\sin(2\varphi_\ell)B_\ell, \\ CE_\ell &\equiv -\cos(2\varphi_\ell)E_\ell, \qquad CB_\ell \equiv -\cos(2\varphi_\ell)B_\ell, \end{aligned} \tag{13}$$

then the observables $Q^{\rm obs}$ and $U^{\rm obs}$ are consequently expressed as

$$\begin{aligned} Q^{\rm obs}(x) &= \widetilde{CE}(x) + \sqrt{r}\widetilde{SB^0}(x) + N^Q(x), \\ U^{\rm obs}(x) &= \widetilde{SE}(x) + \sqrt{r}\widetilde{CB^0}(x) + N^U(x), \end{aligned} \tag{14}$$

where B^0 denotes fiducial B modes with unity power spectrum $C_\ell^{BB,r=1}$, tildes denote lensed fields $\tilde{X}(x) = X(x + \nabla\phi(x))$ $(X = CE, SE, CB, SB)$, and $N^{Q,U}$ is the Q/U noise.

With above decomposition, we find the data vector d is Gaussian with covariance $\tilde{\Sigma}_{r,\phi}$ for given r and $\phi(x)$, i.e., $d \sim N(0, \tilde{\Sigma}_{r,\phi})$, where

$$\tilde{\Sigma}_{r,\phi} \equiv \begin{bmatrix} \tilde{\Sigma}^{Q^{\rm obs},Q^{\rm obs}} & \tilde{\Sigma}^{Q^{\rm obs},U^{\rm obs}} \\ \tilde{\Sigma}^{Q^{\rm obs},U^{\rm obs}} & \tilde{\Sigma}^{U^{\rm obs},U^{\rm obs}} \end{bmatrix}_{r,\phi}, \tag{15}$$

and the covariance matrix is naturally expressible in the form of Eq. (11), i.e.,

$$\begin{aligned} \tilde{\Sigma}^{Q^{\rm obs},Q^{\rm obs}} &= \tilde{\Sigma}^{CE,CE} + r\tilde{\Sigma}^{SB^0,SB^0} + \Sigma^{NQ,NQ}, \\ \tilde{\Sigma}^{Q^{\rm obs},U^{\rm obs}} &= \tilde{\Sigma}^{CE,SE} + r\tilde{\Sigma}^{SB^0,CB^0}, \\ \tilde{\Sigma}^{U^{\rm obs},U^{\rm obs}} &= \tilde{\Sigma}^{SE,SE} + r\tilde{\Sigma}^{CB^0,CB^0} + \Sigma^{NU,NU}. \end{aligned} \tag{16}$$

2. Marginalization

In a more practical case, we only have an estimate of lensing potential $\phi_\ell^{\rm est}$ which is a noisy version of the true ϕ_ℓ, i.e., $\phi_\ell^{\rm est} = \phi_\ell + n_\ell^\phi$, where n_ℓ^ϕ is the uncertainty of the ϕ estimate and its power spectrum $N_\ell^{\phi\phi}$ is usually an output of the lensing estimator used. For an unbiased estimator with Gaussian uncertainty, one can write $n_\ell^\phi \sim N(0, N_\ell^{\phi\phi})$. Then the correlation coefficient of ϕ and $\phi^{\rm est}$ defined in Eq. (4) now is explicitly known as

$$\rho_\ell = \sqrt{\frac{C_\ell^{\phi\phi}}{C_\ell^{\phi\phi}+N_\ell^{\phi\phi}}}. \tag{17}$$

In this context, one can treat $\phi^{\rm est}$ as data and compute the posterior on ϕ given $\phi^{\rm est}$, i.e.,

$$P(\phi_{\boldsymbol{\ell}}|\phi^{\rm est}_{\boldsymbol{\ell}}) \sim N(\phi^{\rm WF}_{\boldsymbol{\ell}}, C^{\epsilon\epsilon}_\ell), \tag{18}$$

with

$$\begin{aligned}\phi^{\rm WF}_{\boldsymbol{\ell}} &= \frac{C_\ell^{\phi\phi}}{C_\ell^{\phi\phi}+N_\ell^{\phi\phi}}\phi^{\rm est}_{\boldsymbol{\ell}} = \rho_\ell^2\phi^{\rm est}_{\boldsymbol{\ell}},\\ C^{\epsilon\epsilon}_\ell &= \frac{C_\ell^{\phi\phi}}{C_\ell^{\phi\phi}+N_\ell^{\phi\phi}}N_\ell^{\phi\phi} = \rho_\ell^2 N_\ell^{\phi\phi}.\end{aligned} \tag{19}$$

Therefore a sample $\phi_{\boldsymbol{\ell}} \sim P(\phi_{\boldsymbol{\ell}}|\phi^{\rm est}_{\boldsymbol{\ell}})$ can be written as

$$\phi_{\boldsymbol{\ell}} = \phi^{\rm WF}_{\boldsymbol{\ell}} + \epsilon^{\rm WF}_{\boldsymbol{\ell}}, \tag{20}$$

with $\epsilon_{\boldsymbol{\ell}} \sim N(0, C^{\epsilon\epsilon}_\ell)$.

Two different sets of variables are involved here: raw variables $(\phi_{\boldsymbol{\ell}}, n^\phi_{\boldsymbol{\ell}}, N^{\phi\phi}_{\boldsymbol{\ell}})$ and Wiener filter variables $(\phi^{\rm WF}_{\boldsymbol{\ell}}, \epsilon^{\rm WF}_{\boldsymbol{\ell}}, C^{\epsilon\epsilon}_{\boldsymbol{\ell}})$. Usually, it is more numerically stable to work on the Wiener filter variables. Marginalizing $\tilde\Sigma_{r,\phi}$ in Eq. (15) over $\epsilon^{\rm WF}$, we obtain a covariance matrix only depending on r, i.e., $\Sigma_r \equiv \langle\tilde\Sigma_{r,\phi}\rangle_{\epsilon^{\rm WF}}$, where its analytic form is presented in Appendixes A and B. The computation of its inverse matrix Σ_r^{-1} is presented in Appendix C.

C. Data compression

1. Idea

To compress the data, we project the original length-$2p$ data vector d to a small number of length-s ($s \ll 2p$) vectors $\hat d^i$ ($i = 0, \ldots, i_{\rm max}$). If we properly choose the projected vectors to be independent and of high S/N, then the total likelihood is given by

$$\log L(r|d) \approx \sum_{i=0}^{i_{\rm max}} \log L(r|\hat d^i). \tag{21}$$

We denote $(\hat d^i)_s = (v^\intercal)^i_{s\times 2p} d_{2p}$, where v^i is the projection matrix to be determined later, then $\hat d^i \sim N(0, (v^\intercal)^i\Sigma_r v^i)$ and the likelihood $L(r|\hat d^i)$ is given by

$$\begin{aligned}-2\log L(r|\hat d^i) = {}& d^\intercal[(v^\intercal)^i\Sigma_r v^i]^{-1}d\\ &+ \log\det[(v^\intercal)^i\Sigma_r v^i],\end{aligned} \tag{22}$$

up to a constant term.

Since the primordial B modes at large scales are less contaminated by the lensing B noise, a natural choice is to project the polarization data to the large-scale quadratic delensed modes defined in Eq. (7), i.e., $\hat d^i = {}^iB^{\rm del}_{\boldsymbol{\ell}}$, where the upper index ^{i} denotes the projected data vector limited to the i-th s lowest frequency modes available (Fig. 1).[1] These projected vectors ${}^iB^{\rm del}_{\boldsymbol{\ell}}$ are not completely independent. We will confirm the validity of ignoring the cross correlation via simulations in Sec. IV.

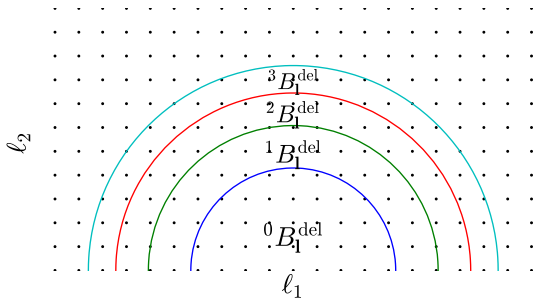

FIG. 1. The modes covered by each different projected vector ${}^iB^{\rm del}_{\boldsymbol{\ell}}(i = 0, 1, 2, \ldots)$. Here we only show the modes with $\ell_2 \geq 0$, since our observables $Q^{\rm obs}$ and $U^{\rm obs}$ are real numbers.

We find that the modified Gaussian likelihood method works better if we incorporate the same number of E modes and delensed B modes in each projected vector, i.e.,

$$[\hat d^i]_{2s} = \begin{bmatrix} ({}^iB^{\rm del}_{\boldsymbol{\ell}})_s \\ ({}^iE^{\rm obs}_{\boldsymbol{\ell}})_s \end{bmatrix} = \begin{bmatrix} (v_b^\intercal)^i_{s\times 2p} \\ (v_e^\intercal)^i_{s\times 2p}\end{bmatrix} d_{2p} = (v^\intercal)^i_{2s\times 2p} d_{2p}. \tag{23}$$

2. Projection matrix

In this subsection, we focus on the computation of the projection matrices v^i. As described in Sec. II, the quadratic delenser is actually a linear operator, i.e.,

$$\begin{aligned} d = (Q^{\rm obs}, U^{\rm obs})^\intercal &\xrightarrow{\text{Eq. (12)}} (E^{\rm obs}, B^{\rm obs}),\\ &\xrightarrow{\text{Eq. (3)}} (B^{\rm len,est}, B^{\rm obs}),\\ &\xrightarrow{\text{Eq. (7)}} B^{\rm obs} - B^{\rm len,est} = B^{\rm del}.\end{aligned} \tag{24}$$

Therefore we can formally write the quadratic delensing as $(B^{\rm del}_{\boldsymbol{\ell}})_{2p} = \mathcal{P}_{2p\times 2p} d_{2p}$, where $\mathcal{P}$ is a concatenation of the three linear operations above, and its matrix elements can be found by recording the impulse response of the delensed modes to each element in the data vector. For example, to compute the 1st column of $\mathcal{P}$, we do the quadratic delensing to a single-element "data vector" $\delta_1 = (1, 0, \ldots, 0)^\intercal_{2p}$ and denote the corresponding delensed modes as $(B^{\rm del}_{\boldsymbol{\ell}}|_{\delta_1})$, i.e.,

[1]The large-scale delensed B modes are no longer the highest S/N modes, when foregrounds, contaminating the primordial B modes more at large scales, are considered. Our methodology is flexible. In principle, modes could be selected that minimizes noise and residual foreground contamination.

$$(B^{\rm del}_{\ell}|_{\delta_1})_{2p} = \mathcal{P}_{2p\times 2p}(\delta_1)_{2p} = \text{1st col of } \mathcal{P}. \quad (25)$$

where $(B^{\rm del}_{\ell}|_{\delta_1})$ is obtained via the three successive operations of Eq. (24). In this way, we obtain the matrix $\mathcal{P}$.

It is clear that the projection matrices of vectors ${}^iB^{\rm del}_{\ell}$ correspond to row blocks of $\mathcal{P}$. Explicitly, we write $(B^{\rm del}_{\ell})_{2p} = \mathcal{P}_{2p\times 2p} d_{2p}$ as

$$\begin{pmatrix} ({}^0B^{\rm del}_{\ell})_s \\ ({}^1B^{\rm del}_{\ell})_s \\ ({}^2B^{\rm del}_{\ell})_s \\ \cdots \end{pmatrix} = \mathcal{P}_{2p\times 2p} d_{2p} = \begin{pmatrix} ({}^0\mathcal{P})_{s\times 2p} \\ ({}^1\mathcal{P})_{s\times 2p} \\ ({}^2\mathcal{P})_{s\times 2p} \\ \cdots \end{pmatrix} d_{2p}, \quad (26)$$

and, therefore, we obtain $(v^{\rm I}_b)^i_{s\times 2p} = ({}^i\mathcal{P})_{s\times 2p}$. The projection matrices $(v^{\rm I}_e)^i$ can be obtained in a similar but easier way, since only the first operation in Eq. (24) is involved.

IV. SIMULATIONS

In this section we present simulations that demonstrate the likelihood-based higher order delensing methodology in comparison to first order E template delensing [47]. Although higher order iterative estimates of lensing/delensing have been forecasted to outperform first order E template delensing [69], their implementation is only just now being developed and highly coupled to the lensing estimation methodology. Indeed, there are arguably only three prototype implementations existing in the current literature, all derived from iterative maximization of the log posterior of lensing (see [62,65,66]), none of which are yet sufficiently developed for use in obtaining r constraints.

Since the focus of this paper is exclusively on the problem of CMB delensing, as distinct from lensing estimation, we simplify our simulations by stipulating that lensing potential uncertainty arises from a simple additive noise model $\phi(x) + n^{\phi}(x)$, with $\phi(x)$ being the true lensing potential field and $n^{\phi}(x)$ being independent isotropic Gaussian noise. The spectral density used for $n^{\phi}(x)$ is set to the projected error from the EB quadratic lensing estimate [60] which mimics a two survey approach for r constraints: one survey for reconstructing the lensing map, the second for r constraints via delensing. Although the EB quadratic estimate is known to be suboptimal, as compared to projections in the literature of iterative techniques [69], we use the EB quadratic estimate uncertainty due to the fact that far less is known about the bias and uncertainty associated with higher order lensing estimates which may lead to overly optimistic projections of lensing uncertainty.

The fiducial cosmology we use in this section is a flat ΛCDM cosmology with a baryon density $\omega_{\rm b} = 0.02246$, a cold dark matter density $\omega_{\rm c} = 0.1185$, a reionization optical depth $\tau = 0.079$, an angular size of sound horizon at recombination $100\theta_{\star} = 1.0410$, an amplitude and a spectral index of the primordial scalar the perturbation power spectrum $\ln(10^{10}A_{\rm s}) = 3.192, n_{\rm s} = 0.9686$, and a tensor-to-scalar ratio r in the range of $[0.001, 0.1]$. For each different r, we simulate 500 realizations of primordial polarization fields $Q(x)$ and $U(x)$, then lense these fields via the same lensing potential field $\phi(x)$, following the lensing approach proposed by Louis *et al.* [70]. All the power spectra used in simulations are computed from the Boltzmann code CLASS [71].

A. Two surveys

We consider a survey strategy consisting of two different surveys of the same area of sky, differing in angular resolution. The main goal of the higher-resolution survey is to allow for a reconstruction of the lensing potential. Such reconstructions benefit from reaching an angular scale comparable to the typical lensing deflection angle of ∼2 arc min. In contrast, the primordial B-mode signal is on fairly large angular scales of greater than a degree. In principal, one high-resolution survey could be used both for the lensing reconstruction and for sensitivity to the primordial B-mode signal. However, there are advantages to using a survey dedicated to the large-scale signals. These advantages do not appear in the idealized analyses that we perform here, as they are related to systematic error control and foreground cleaning, as we now explain briefly. The large-scale survey can be achieved with a smaller telescope with a simplified optics chain. Having a smaller telescope facilitates boresight rotation, which BICEP2/Keck have used for null tests to bound certain systematic errors. Foreground cleaning is also likely to be more of a challenge at larger angular scales than it is for the smaller angular scales with the bulk of the lensing information and serves as a further driver of differences in optimal design for the two surveys. For these reasons a two-survey approach is likely to be a part of the straw man concepts for the CMB Stage-IV instrument soon to emerge from the CMB Stage-IV Concept Definition Taskforce.

In this paper, we simulate three scenarios. We consider a scenario "N" in which the B-mode instrument noise power is larger than the B-mode lensing power, and two scenarios "La" and "Lb" with the opposite situation. The more

TABLE II. The three scenarios we simulated.

		$\Delta_{\rm T}$ (μK-arc min)	$\theta_{\rm FWHM}$	$f_{\rm sky}$
Lb	ϕ	0.5	$2'$	2.7%
	B	0.5	$10'$	
La	ϕ	1	$2'$	2.7%
	B	1	$10'$	
N	ϕ	10	$2'$	2.7%
	B	10	$10'$	

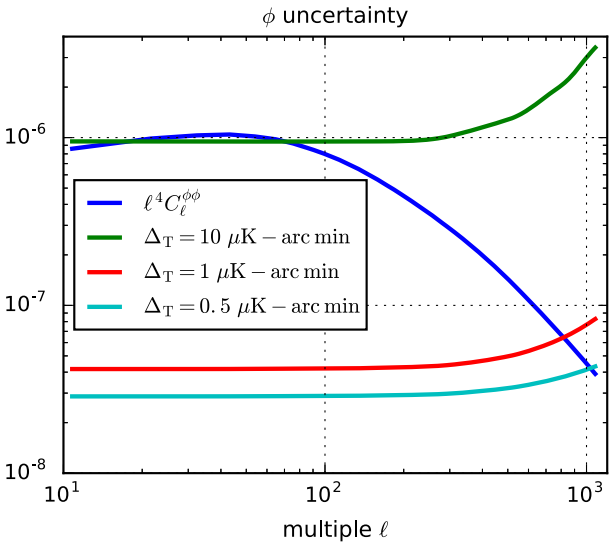

FIG. 2. The $\phi(x)$ reconstruction noises for Scenario N ($\Delta_T = 10\ \mu$K-arc min), La ($\Delta_T = 1\ \mu$K-arc min), and Lb ($\Delta_T = 0.5\ \mu$K-arc min) surveys.

sensitive scenarios La and Lb are motivated by potential CMB-S4 scenarios. Each scenario consists of two surveys, a high-resolution survey for ϕ reconstruction and a low-resolution survey capturing the B-mode signal, covering the same patch of the sky (see Table II for the survey configurations in detail). The noise levels specified here should be understood as noise levels in a foreground-cleaned CMB map, as mentioned in the Introduction. For the high-resolution surveys, the lensing potential reconstruction noise expected from the EB quadratic estimator [59,60,72] is shown in Fig. 2.

B. r constraints

For each simulated CMB realization, we first reconstruct the lensing potential field from the ϕ survey (*high* resolution survey) using the EB quadratic estimator, then use the reconstructed lensing field $\phi^{\mathrm{est}}(x)$ to delense the *low* resolution polarization maps using the quadratic delenser (Sec. II) and the modified Gaussian likelihood method (Sec. III), and finally compare their r constraints from the two delensers.

In Fig. 3, we show the detection level $r/\sigma(r)$ and bias level $\mathrm{Bias}(r)/r$ obtained from the quadratic delenser and from the modified Gaussian likelihood method, where $\sigma(r)$ and $\mathrm{Bias}(r)$ are the standard error and the average bias of the 500 best-fit r values (from 500 CMB realizations), respectively. For Scenario N, both methods obtain similar r detection levels, while the modified Gaussian likelihood method shows its advantages in the Scenario La and Lb. We find that in the regime of low map noise ($\lesssim 1\ \mu$K-arc min), the bias of the modified Gaussian likelihood method is appreciably smaller than that of the quadratic delenser (see next subsection for the detailed bias analysis for the quadratic delenser).

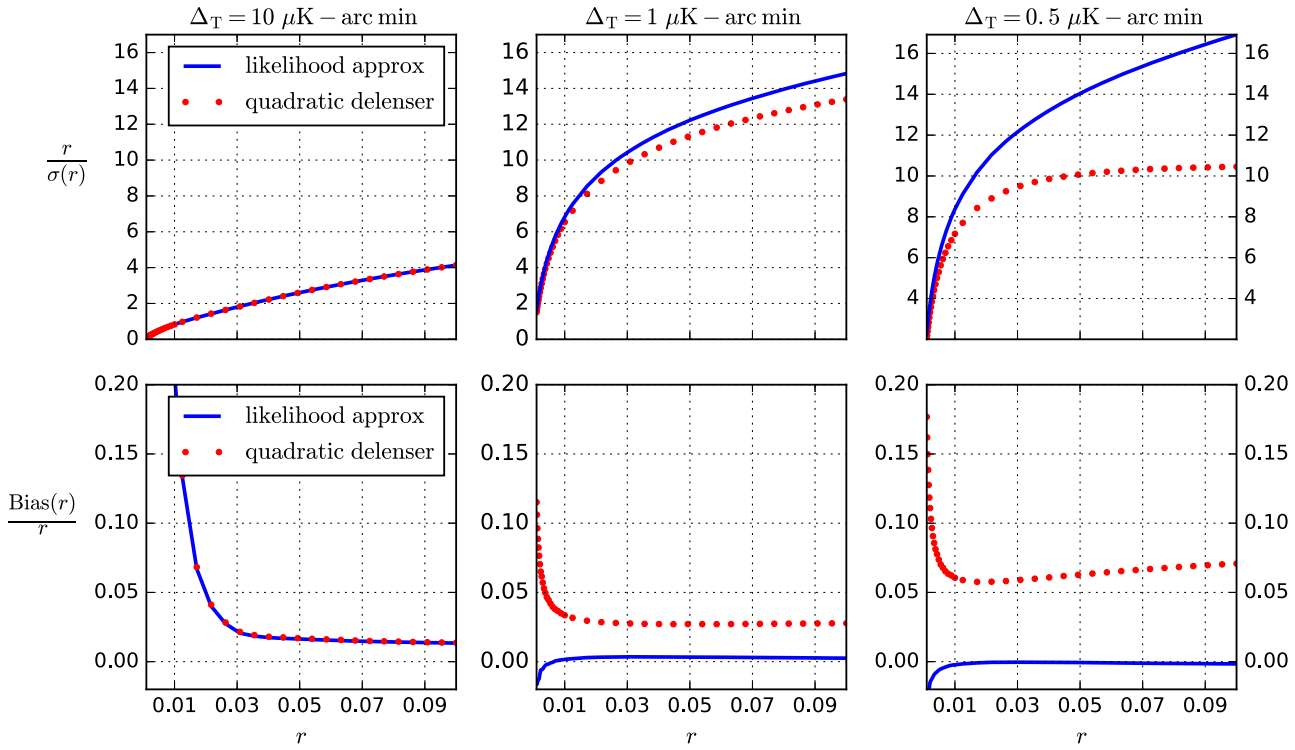

FIG. 3. Upper three panels show the detection levels $r/\sigma(r)$ expected from surveys of Scenario N, La, and Lb, and lower three panels show the corresponding bias levels $\mathrm{Bias}(r)/r$.

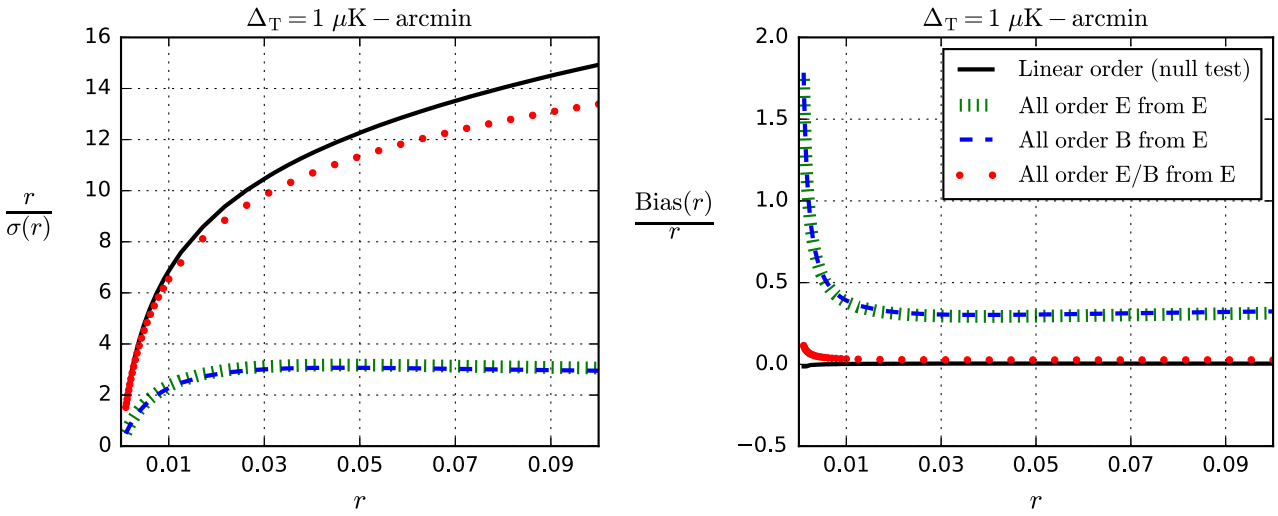

FIG. 4. Bias analysis of the quadratic delenser via simulations under different assumptions: (black/solid lines) null test assuming $E^{\rm obs} = E + N^E$ and $B^{\rm obs} = B + \delta^1 B_{\rm from\,E} + N^B$; (green/bar lines) all order E from E test assuming $E^{\rm obs} = E + \delta E_{\rm from\,E} + N^E$, $B^{\rm obs} = B + \delta^1 B_{\rm from\,E} + N^B$; (blue/dashed lines) all order B from E test assuming $E^{\rm obs} = E + N^E$ and $B^{\rm obs} = B + \delta B_{\rm from\,E} + N^B$; (red/dots) all order E/B from E test assuming $E^{\rm obs} = E + \delta E_{\rm from\,E} + N^E$ and $B^{\rm obs} = B + \delta B_{\rm from\,E} + N^B$.

For Scenario La with map noise $\Delta_T = 1$ μK-arc min and sky coverage $f_{\rm sky} = 2.7\%$, we expect to detect the primordial B-mode signal at $\sim 1\sigma$ level for $r = 0.001$ and at $\sim 15\sigma$ level for $r = 0.1$. The lower noise Scenario Lb with map noise $\Delta_T = 0.5$ μK-arc min and the same sky coverage, only marginally increases the detection level, due to the saturation of cosmic variance.

C. Bias analysis for the quadratic delenser

In this subsection, we aim to quantify the bias of the quadratic delenser introduced by ignoring the lensing in E modes and higher order lensing in B modes.[2] For clarity, we use the following notation to denote the connection between lensed and primordial variables

$$\tilde{E} = E + \delta E_{\rm from\,E} + \delta E_{\rm from\,B},$$
$$\tilde{B} = B + \delta B_{\rm from\,E} + \delta B_{\rm from\,B}, \tag{27}$$

where $\delta X_{\rm from\,Y}$ is the lensing in (lensed) X from (primordial) Y. In addition, $\delta E_{\rm from\,B}$ and $\delta B_{\rm from\,B}$ are much smaller than their counterparts $\delta E_{\rm from\,E}$ and $\delta B_{\rm from\,E}$, so we simply ignore them in this subsection.

(i) First we do a null test. In accordance with the two approximations made in the quadratic delenser (Sec. II), we completely drop lensing in E modes and only keep linear order lensing in B modes, i.e., we simulate polarization maps assuming $E^{\rm obs} = E + N^E$ and $B^{\rm obs} = B + \delta^1 B_{\rm from\,B} + N^B$, where $N^{E/B}$ is the E/B map noise, and $\delta^1 B_{\rm from\,E}$ is the linear order lensing in B from E. As expected, we find the quadratic delenser is not biased in this context (Fig. 4, black/solid lines).[3,4]

(ii) To scrutinize the bias introduced by ignoring lensing in E modes, we keep all order lensing in E modes and linear order lensing in B modes; i.e., we simulate polarization maps assuming $E^{\rm obs} = E + \delta E_{\rm from\,E} + N^E$ and $B^{\rm obs} = B^r + \delta^1 B_{\rm from\,B} + N^B$. In this context, the quadratic delenser is highly biased (Fig. 4, green/bar lines).

(iii) In the same way, to test the bias introduced by ignoring high order lensing terms in B modes, we ignore lensing in E modes and keep all order lensing in B modes; i.e., we do simulations assuming $E^{\rm obs} = E + N^E$ and $B^{\rm obs} = B + \delta B_{\rm from\,B} + N^B$. In this context, we also find the quadratic delenser is highly biased. More interestingly, we find that the bias level almost exactly matches that of ignoring lensing in E modes (Fig. 4, blue/dashed lines).

(iv) The final step is to check the interaction between the two bias terms from (iii) and (iv). For this purpose, we keep all order lensing in E modes and all order lensing in B modes; i.e., we simulate polarization

[2]In principle, ignoring the nonstationarity of the delensed B modes also induces some bias to the r constraint. But we will see this bias is negligible.

[3]From the null test, where we ignore the nonstationarity of the delensed B modes, we conclude that the bias induced by ignoring the nonstationarity is negligible.

[4]Comparing the detection level of the null test (solid line in the left panel of Fig. 4), and the detection level of the modified Gaussian likelihood (solid line in the second panel of Fig. 3), we find the two matches exactly. Therefore, we confirm the validity of the two major approximations used in the modified Gaussian likelihood: only keeping a few projected data vectors, and ignoring the cross relation between different projected data vectors.

maps assuming $E^{\rm obs} = E + \delta E_{\rm fromE} + N^E$, and $B^{\rm obs} = B + \delta B_{\rm fromE} + N^B$. We find that the two bias contributions cancel to a high precision and therefore the net bias is strongly suppressed (Fig. 4, red/dots).

To make sense of the bias cancellation, we do a simple magnitude analysis. In the quadratic delenser, we delense the B modes via a quadratic template subtraction $B^{\rm res} = B_{\rm fromE} - E^{\rm obs} * \phi^{\rm est}$, and assume a residual power spectrum $C_\ell^{BB,{\rm res}} = \langle |\delta^1 B_{\rm fromE} - E * \phi^{\rm est}|^2_\ell \rangle$, where $*$ denotes the convolution defined in Eq. (2), (3), in the $\delta^1 B_{\rm fromE}$ term we ignore the second (and higher) order lensing in B modes $\delta^2 B_{\rm fromE}$, and in the $E * \phi$ term we ignore the difference of E and $\tilde{E}$. Therefore the template subtraction used has an error $\delta^2 B_{\rm fromE} - \delta^1 E_{\rm fromE} * \phi$, where both error terms are of the same order $O(E\phi^2)$ considering that

$$
\begin{aligned}
&(\delta^2 B_{\rm fromE})_{\boldsymbol{\ell}} \\
&= -\frac{1}{2}\int \frac{d^2\boldsymbol{\ell}_1 d^2\boldsymbol{\ell}_2}{(2\pi)^2}[(\boldsymbol{\ell}_1+\boldsymbol{\ell})\cdot(\boldsymbol{\ell}_1+\boldsymbol{\ell}_2)] \\
&\quad\times[(\boldsymbol{\ell}_1+\boldsymbol{\ell})\cdot\boldsymbol{\ell}_2] E_{\boldsymbol{\ell}_1+\boldsymbol{\ell}} \sin(2\varphi_{\boldsymbol{\ell}_1+\boldsymbol{\ell},\boldsymbol{\ell}}) \phi^*_{\boldsymbol{\ell}_1+\boldsymbol{\ell}_2} \phi_{\boldsymbol{\ell}_2},
\end{aligned}
\tag{28}
$$

and

$$
\begin{aligned}
(\delta^1 E_{\rm fromE})_{\boldsymbol{\ell}} = &\int \frac{d^2\boldsymbol{\ell}'}{2\pi}\boldsymbol{\ell}'\cdot(\boldsymbol{\ell}'+\boldsymbol{\ell}) \\
&\times E_{\boldsymbol{\ell}'+\boldsymbol{\ell}} \cos(2\varphi_{\boldsymbol{\ell}'+\boldsymbol{\ell},\boldsymbol{\ell}}) \phi^*_{\boldsymbol{\ell}'}.
\end{aligned}
\tag{29}
$$

To summarize, in the quadratic delenser,

$$
B^{\rm del}_{\boldsymbol{\ell}} = B^r_{\boldsymbol{\ell}} + B^{\rm res}_{\boldsymbol{\ell}} + N^B_{\boldsymbol{\ell}}, \tag{30}
$$

we have ignored lensing in E modes and high order lensing in B modes when estimating the residual power spectrum $\langle |B^{\rm res}_{\boldsymbol{\ell}}|^2 \rangle$. We find that each of the two approximations introduces a strong bias in the r estimate, while the two bias contributions cancel to a high precision, and the validity of the quadratic delenser sensitively depends on the cancellation.

According to the above analysis, the bias in the residual power estimate in principle is independent of primordial B-mode signal B^r, therefore we naively expect a r-independent bias ${\rm Bias}(r)$ and therefore a bias level ${\rm Bias}(r)/r$ decaying with growing r, which is indeed the behavior we observe for $r \lesssim 0.01$ (Figs. 3 and 4). But the bias level does not die down for even greater r, since the r constraints become more sensitive to higher frequency regime where the bias is stronger. Here, we give an informal analysis of the bias level behavior. From a single delensed B mode $B^{\rm del}_{\boldsymbol{\ell}}$, we can estimate the primordial B-mode power spectrum with root variance $\Delta C_{\boldsymbol{\ell}} = C^{BB,r}_{\boldsymbol{\ell}} + C^{BB,{\rm res}}_{\boldsymbol{\ell}} + N^{BB}_{\boldsymbol{\ell}}$ and consequently estimate r with mean value

$$
r^{\rm est}_{\boldsymbol{\ell}} = \frac{|B^{\rm del}_{\boldsymbol{\ell}}|^2 - C^{BB,{\rm res}}_{\boldsymbol{\ell}} - N^{BB}_{\boldsymbol{\ell}}}{C^{BB,r=1}_{\boldsymbol{\ell}}}, \tag{31}
$$

and with root variance $\sigma_{\boldsymbol{\ell}}(r) = \Delta C_{\boldsymbol{\ell}}/C^{BB,r=1}_{\boldsymbol{\ell}}$. These estimators from different modes can be added with inverse-variance weighting $r^{\rm est} = \sum_{\boldsymbol{\ell}} W_{\boldsymbol{\ell}} r^{\rm est}_{\boldsymbol{\ell}}$, where

$$
W_{\boldsymbol{\ell}}(r) = \frac{\frac{1}{\sigma^2_{\boldsymbol{\ell}}(r)}}{\sum_{\boldsymbol{\ell}} \frac{1}{\sigma^2_{\boldsymbol{\ell}}(r)}}. \tag{32}
$$

It is straightforward to understand that $W_{\boldsymbol{\ell}}(r)$ increases with r for large $|\boldsymbol{\ell}|$ where $C^{BB,{\rm res}}_{\boldsymbol{\ell}} + N^{BB}_{\boldsymbol{\ell}}$ dominates $\Delta C_{\boldsymbol{\ell}}$, and decreases with r for small $|\boldsymbol{\ell}|$ where $C^{BB,r}_{\boldsymbol{\ell}}$ dominates $\Delta C_{\boldsymbol{\ell}}$. In addition, we know that the quadratic delenser is a biased estimator, i.e.,

$$
\langle |B^{\rm del}_{\boldsymbol{\ell}}|^2 \rangle = C^{BB,{\rm res}}_{\boldsymbol{\ell}} + N^{BB}_{\boldsymbol{\ell}} + C^{BB,r}_{\boldsymbol{\ell}} + C^{BB,{\rm bias}}_{\boldsymbol{\ell}}, \tag{33}
$$

and

$$
\langle r^{\rm est}_{\boldsymbol{\ell}} \rangle = \frac{C^{BB,r}_{\boldsymbol{\ell}} + C^{BB,{\rm bias}}_{\boldsymbol{\ell}}}{C^{BB,r=1}_{\boldsymbol{\ell}}} = r + r^{\rm bias}_{\boldsymbol{\ell}}, \tag{34}
$$

where $r^{\rm bias}_{\boldsymbol{\ell}}$ increases with $|\boldsymbol{\ell}|$. Therefore, we have $\langle r^{\rm est} \rangle = \sum_{\boldsymbol{\ell}} W_{\boldsymbol{\ell}} \langle r^{\rm est}_{\boldsymbol{\ell}} \rangle = r + \sum_{\boldsymbol{\ell}} W_{\boldsymbol{\ell}}(r) r^{\rm bias}_{\boldsymbol{\ell}} = r + {\rm Bias}(r)$, with ${\rm Bias}(r)$ increasing with r. It also explains the increasing bias level ${\rm Bias}(r)$ with decreasing map noise $N^{BB}_{\boldsymbol{\ell}}$ (see Fig. 3). Note that we do not expect the quadratic delenser to exactly match the inverse-variance weighted estimator described above, but the latter should be a good proxy for interpreting the bias behavior.

D. Nonstationary noise

The modified Gaussian likelihood works not only as a correction to the quadratic template subtraction estimator but also shows its advantage in dealing with realistic experiment complexities, e.g. nonstationary noise and sky cuts. Here we explore an example of nonstationary noise with pixel dependent noise, i.e., $\langle n(\mathbf{x}) n(\mathbf{y}) \rangle = \sigma^2(\mathbf{x}) \Delta_{\rm P}^2 \delta_D(\mathbf{x}-\mathbf{y})$, with $\Delta_{\rm P} = \sqrt{2}\Delta_{\rm T} = \sqrt{2}\ \mu$K-arc min, and $\sigma(\mathbf{x})$ a pixel-dependent modulation (Fig. 5). We expect the likelihood-based estimator to work robustly in the presence of nonstationary noise, as long as we take the pixel dependent noise into account when calculating the covariance matrix of noise (see Appendix B). But the nonstationary noise becomes troublesome for the quadratic delenser in Fourier space.[5]

[5]In the case of nonstationary noise, the noise power spectrum loses the protection of symmetry; i.e., $\langle n_{\boldsymbol{\ell}} n_{\boldsymbol{\ell}'} \rangle = N_{\boldsymbol{\ell},\boldsymbol{\ell}'}$ now depends on both multipoles instead of their linear combination $\boldsymbol{\ell}, \boldsymbol{\ell}'$. If we were to correctly use the quadratic delenser, then the residual power evaluation in Eq. (6) becomes difficult, and is out of the scope of this paper. Here we simply (but incorrectly) assume the stationary noise power spectrum in Eq. (6), and test how the nonstationary noise biases the r constraint from the quadratic delenser.

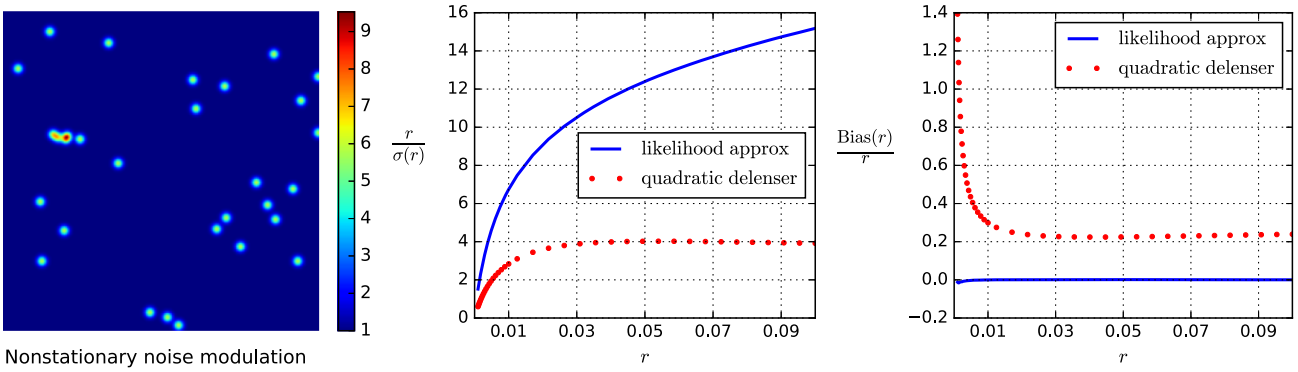

FIG. 5. The impact of nonstationary map noise on the r constraints for the Scenario La experiments ($\Delta_T = 1$ μK-arc min). Left panel: the nonstationary noise modulation $\sigma(\mathbf{x})$. Middle/Right panel: the detection/bias level of r constraints inferred from the two delensers.

Applying the two estimators on simulations with nonstationary noise, we find that the modified Gaussian likelihood method works as well as in the case of stationary noise, while the quadratic delenser is significantly biased (Fig. 5).

V. SUMMARY AND CONCLUSIONS

Delensing is a crucial part for future CMB experiments aiming to detect a primordial B-mode signal. Up to linear order, one can effectively delense observed B-modes by utilizing a quadratic combination of observed E-modes and an estimate of the lensing potential. This is the underlying idea of the quadratic delenser. However, in the regime of small map noise, the lensing in E modes, and higher order lensing in B modes ignored by the quadratic delenser, significantly bias the r constraint. We investigated the bias induced by each of the two approximations via simulations, finding that each of two approximations induce a large bias, while the two bias terms partly cancel and therefore the net bias is moderately suppressed. The validity of the quadratic delenser sensitively depends on the cancellation.

Alternatively, a full-scale likelihood analysis of the tensor-to-scalar ratio r can, in principle, optimally account for all the r information in the CMB observations and remedy possible bias problems. Unfortunately, a full likelihood analysis requires computation resources beyond what is available in the near future. In this paper, we presented a modified Gaussian likelihood method. This method consists of two parts, covariance decomposition and data compression. In the first part, we decomposed the covariance matrix in the form of $\Sigma_r = \Sigma^{\mathrm{en}} + r\Sigma^{\mathrm{b}}$, which allows us to compute the covariance matrix Σ_r, as a function of r, at the computational cost of a single covariance matrix evaluation. In the second part, we compressed the data size by keeping only $s \sim 500$ high signal-to-noise modes, say the large-scale quadratic delensed B modes. We obtained these B modes from polarization data d via a projection matrix v, $({}^0B^{\mathrm{del}}_{\ell})_s = (v^{\intercal})_{s\times 2p} d_{2p}$, and applied the likelihood analysis on the projected data vector. This method can be naturally extended to incorporate higher frequency modes.

Finally, we applied the quadratic delenser and the modified Gaussian likelihood method on simulated CMB observations mimicking experiments of Scenario N, La, and Lb, and compare the resulting r constraints. We found that the two methods have similar performance in constraining r for Scenario N, while the quadratic delenser does not perform as well for the lower-noise Scenarios La and Lb due to a strong r constraint bias in the regime of low map noise. For Scenario La, we expected to detect the primordial B-mode signal at $\sim 1\sigma$ level for $r = 0.001$, and at $\sim 15\sigma$ level for $r = 0.1$, from the modified Gaussian likelihood method. For Scenario Lb with even lower map noise and the same sky coverage, the detection level only marginally increases due to the saturation of cosmic variance. Therefore it would be valuable to optimize the survey configurations $(\Delta_T, f_{\mathrm{sky}})$ for the coming CMB experiments given a fixed amount of survey time [29].

We also explored the impact of realistic experiment complexities: in the presence of nonstationary noise, the modified Gaussian likelihood method also works robustly as long as we slightly modify the noise covariance matrix to take into account the pixel dependent noise.

To realize the delensing method we present here as part of a complete analysis pipeline from observed maps to constraints on r, a number of additional steps are required. First, one needs to estimate the lensing potential from the observed maps. Such estimates in general will have some degree of correlation with the CMB signal in the maps themselves, although this can be avoided by separating the spatial frequencies used for lensing from those used to detect the tensor signal. They will also be non-Gaussian to some degree, and the impact of this distribution will need to be understood—perhaps via simulations. Also, as mentioned in the introduction, foregrounds need to be taken into account. There are at least two different paths to pursue: include a model of foreground residuals in the foreground-cleaned map, or extend the likelihood of the map to a likelihood of maps at multiple frequencies, with a model of the foreground contributions. We leave these developments to future work.

ACKNOWLEDGMENTS

We thank the referee for his/her insightful comments on this manuscript. Z. P. is supported by UC Davis Dissertation Year Fellowship. E. A. acknowledges support from NSF CAREER Grant No. DMS-1252795. This work made extensive use of the NASA Astrophysics Data System and of the astro-ph preprint archive at arXiv.org.

APPENDIX A: SIGNAL COVARIANCE MATRIX

There are eight different terms in the map covariance matrix Σ_r: $\{\tilde{\Sigma}^{SE,SE}, \tilde{\Sigma}^{CE,SE}, \tilde{\Sigma}^{CE,CE}\}$, $\{\tilde{\Sigma}^{SB,SB}, \tilde{\Sigma}^{CB,SB}, \tilde{\Sigma}^{CB,CB}\}$, and $\{\Sigma^{NQ,NQ}, \Sigma^{NU,NU}\}$ [see Eqs. (16)]. In this subsection, we show how the marginalization over uncertainty in the ϕ estimate is done for the six lensed signal terms, and leave the two noise terms to the next subsection. Take $\tilde{\Sigma}^{XX}(X = SE)$ as an example,

$$
\begin{aligned}
\langle \tilde{\Sigma}^{XX}_{r,\phi} \rangle_{\epsilon^{\rm WF}} &= \langle \tilde{X}(\mathbf{x})\tilde{X}(\mathbf{y})\rangle_{\epsilon^{\rm WF}} \\
&= \langle X(\mathbf{x}+\nabla\phi^{\rm WF}(\mathbf{x})+\nabla\epsilon^{\rm WF}(\mathbf{x}))\times X(\mathbf{y}+\nabla\phi^{\rm WF}(\mathbf{y})+\nabla\epsilon^{\rm WF}(\mathbf{y}))\rangle_{\epsilon^{\rm WF}} \\
&= \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} e^{i\boldsymbol{\ell}\cdot(\mathbf{x}-\mathbf{y}+\nabla\phi^{\rm WF}(\mathbf{x})-\nabla\phi^{\rm WF}(\mathbf{y}))} C^{XX}_\ell \langle e^{i\boldsymbol{\ell}\cdot(\nabla\epsilon^{\rm WF}(\mathbf{x})-\nabla\epsilon^{\rm WF}(\mathbf{y}))}\rangle_{\epsilon^{\rm WF}} \\
&= \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} e^{i\boldsymbol{\ell}\cdot(\mathbf{x}-\mathbf{y}+\nabla\phi^{\rm WF}(\mathbf{x})-\nabla\phi^{\rm WF}(\mathbf{y}))} C^{XX}_\ell \exp\left\{-\frac{1}{2}\boldsymbol{\ell}\cdot[\Sigma^\epsilon(0)-\Sigma^\epsilon(\mathbf{x}-\mathbf{y})]\cdot\boldsymbol{\ell}\right\} \\
&\simeq \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} e^{i\boldsymbol{\ell}\cdot(\mathbf{x}-\mathbf{y}+\nabla\phi^{\rm WF}(\mathbf{x})-\nabla\phi^{\rm WF}(\mathbf{y}))} C^{XX}_\ell \left(1-\frac{1}{2}\boldsymbol{\ell}\cdot[\Sigma^\epsilon(0)-\Sigma^\epsilon(\mathbf{x}-\mathbf{y})]\cdot\boldsymbol{\ell}\right) \\
&= \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} e^{i\boldsymbol{\ell}\cdot(\mathbf{x}-\mathbf{y}+\nabla\phi^{\rm WF}(\mathbf{x})-\nabla\phi^{\rm WF}(\mathbf{y}))} C^{XX}_\ell - \frac{1}{2}\sum_{p,q=1}^{2}[\Sigma^\epsilon(0)-\Sigma^\epsilon(\mathbf{x}-\mathbf{y})]_{p,q}\int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2}\ell_p\ell_q e^{i\boldsymbol{\ell}\cdot(\mathbf{x}-\mathbf{y}+\nabla\phi^{\rm WF}(\mathbf{x})-\nabla\phi^{\rm WF}(\mathbf{y}))} C^{XX}_\ell \\
&= {\rm Cov}(X(\mathbf{w}),X(0)) + \frac{1}{2}\sum_{p,q=1}^{2}[\Sigma^\epsilon(0)-\Sigma^\epsilon(\mathbf{x}-\mathbf{y})]_{p,q}\partial_{p,q}{\rm Cov}(X(\mathbf{w}),X(0)),
\end{aligned}
\tag{A1}
$$

where we have used cumulant expansion at the 4th equal sign, $[\Sigma^\epsilon(\mathbf{x}-\mathbf{y})]_{p,q}$ is the covariance of $\nabla\epsilon^{\rm WF}$, i.e.,

$$
[\Sigma^\epsilon(\mathbf{x}-\mathbf{y})]_{p,q} = \langle \nabla_p\epsilon^{\rm WF}(\mathbf{x})\nabla_q\epsilon^{\rm WF}(\mathbf{y})\rangle_{\epsilon^{\rm WF}} = \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2}\ell_p\ell_q e^{i\boldsymbol{\ell}\cdot(\mathbf{x}-\mathbf{y})} N^{\phi\phi}_\ell, \tag{A2}
$$

and ${\rm Cov}(X(\mathbf{w}),X(0))$ is the covariance of X at separation $\mathbf{w} = \mathbf{x}-\mathbf{y}+\nabla\phi^{\rm WF}(\mathbf{x})-\nabla\phi^{\rm WF}(\mathbf{y})$,

$$
{\rm Cov}(X(\mathbf{w}),X(0)) = \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} e^{i\boldsymbol{\ell}\cdot\mathbf{w}} C^{XX}_\ell. \tag{A3}
$$

The above two dimensional integrals [Eqs. (A2)–(A3)] can be simplified as one dimensional integrals as follows. Take Eq. (A3) as an example,

$$
\begin{aligned}
{\rm Cov}(X(\mathbf{w}),X(0)) &= 4\partial_1^2\partial_2^2\int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} e^{i\boldsymbol{\ell}\cdot\mathbf{w}} C^{\mathcal{E}\mathcal{E}}_\ell \\
&\equiv 4\partial_1^2\partial_2^2 K^{\mathcal{E}}(w),
\end{aligned}
\tag{A4}
$$

where we have used $C^{XX}_\ell = 4\ell_1^2\ell_2^2(C^{EE}_\ell/\ell^4)$ and defined $C^{\mathcal{E}\mathcal{E}}_\ell \equiv C^{EE}_\ell/\ell^4$. Exploiting the integral representation of Bessel functions, we rewrite $K^{\mathcal{E}}(\mathbf{w})$ as a one dimensional integral

$$
K^{\mathcal{E}}(\mathbf{w}) = \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} e^{i\boldsymbol{\ell}\cdot\mathbf{w}} C^{\mathcal{E}\mathcal{E}}_\ell = \frac{1}{2\pi}\int J_0(\ell w) C^{\mathcal{E}\mathcal{E}}_\ell \ell\, d\ell, \tag{A5}
$$

which has no angular dependence. For derivative calculation, we define $\hat{K}(w^2) \equiv K^{\mathcal{E}}(\mathbf{w})$, then

$$
\begin{aligned}
\partial_1^2\partial_2^2 K^{\mathcal{E}}(\mathbf{w}) &= \partial_1^2\partial_2^2\hat{K}(w^2) \\
&= 16 w_1^2 w_2^2 \hat{K}^{(4)}(w^2) \\
&\quad + 8(w_1^2+w_2^2)\hat{K}^{(3)}(w^2) + 4\hat{K}^{(2)}(w^2).
\end{aligned}
\tag{A6}
$$

Using the property

$$
\frac{d}{dz} z^{-s} J_s(z) = -z^{-s} J_{s+1}(z), \tag{A7}
$$

the n-th order derivative $\hat{K}^{(n)}$ is explicitly expressed as

$$\hat{K}^{(n)}(w^2) = \frac{1}{2\pi}\int \left(-\frac{\ell}{2w}\right)^n J_n(\ell w)\ell d\ell. \quad \text{(A8)}$$

Collecting Eqs. (A4), (A6), (A8), $\mathrm{Cov}(X(\mathbf{w}), X(0))$ is decomposed into a few one dimensional integrals. The calculation of $\partial_{p,q}\mathrm{Cov}(X(\mathbf{w}), X(0))$ and $[\Sigma^e(\mathbf{x}-\mathbf{y})]_{p,q}$ is conducted in the same way. For other lensed terms, the above formulas apply similarly.

APPENDIX B: NOISE COVARIANCE MATRIX

In Sec. A, we completely ignore the consequence of the finite beam size in the signal covariance evaluation, since the signal suppression by the beam convolution can be interpreted as the noise enhancement by the beam deconvolution. For noise field $n(\mathbf{x})$, we denote the deconvolved noise field as $X(\mathbf{x}) = \varphi_x^{-1}[n(\mathbf{x})]$, with

$$\varphi_x^{-1}[n(\mathbf{x})] = \int \frac{d^2\boldsymbol{\ell}}{2\pi} e^{i\boldsymbol{\ell}\cdot\mathbf{x}} \frac{n_{\boldsymbol{\ell}}}{\varphi_{\boldsymbol{\ell}}} = \int \frac{d^2\boldsymbol{\ell}}{2\pi}\frac{d^2\mathbf{x}'}{2\pi} e^{i\boldsymbol{\ell}\cdot(\mathbf{x}-\mathbf{x}')} \frac{n(\mathbf{x}')}{\varphi_{\boldsymbol{\ell}}}, \quad \text{(B1)}$$

where for Gaussian beam profile $\varphi(\mathbf{x}) = \frac{1}{2\pi\sigma_b^2}\exp(-\frac{\mathbf{x}^2}{2\sigma_b^2})$, $\varphi_{\boldsymbol{\ell}} = \exp(-\frac{l^2\sigma_b^2}{2})$, and $\sigma_b^2 = \theta^2_{\mathrm{FWHM}}/(8\ln 2)$. Then

$$\Sigma^{XX} = \langle X(\mathbf{x})X(\mathbf{y})\rangle = \int \frac{d^2\boldsymbol{\ell}}{2\pi}\frac{d^2\mathbf{x}'}{2\pi}\frac{d^2\mathbf{k}}{2\pi}\frac{d^2\mathbf{y}'}{2\pi} e^{i\boldsymbol{\ell}\cdot(\mathbf{x}-\mathbf{x}')} e^{i\mathbf{k}\cdot(\mathbf{y}-\mathbf{y}')} \frac{1}{\varphi_{\boldsymbol{\ell}}\varphi_{\mathbf{k}}}\langle n(\mathbf{x}')n(\mathbf{y}')\rangle. \quad \text{(B2)}$$

For simple white noise $\langle n(\mathbf{x})n(\mathbf{y})\rangle = \Delta_{\mathrm{P}}^2\delta_D(\mathbf{x}-\mathbf{y})$, we have

$$\langle X(\mathbf{x})X(\mathbf{y})\rangle = \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} e^{i\boldsymbol{\ell}\cdot(\mathbf{x}-\mathbf{y})}\frac{\Delta_{\mathrm{P}}^2}{\varphi_{\boldsymbol{\ell}}\varphi_{-\boldsymbol{\ell}}} = \Delta_{\mathrm{P}}^2\int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} e^{i\boldsymbol{\ell}\cdot(\mathbf{x}-\mathbf{y})} e^{\ell^2\sigma_b^2}, \quad \text{(B3)}$$

where Δ_{P} is polarization noise and we usually take $\Delta_{\mathrm{P}} = \sqrt{2}\Delta_{\mathrm{T}}$. For more realistic nonstationary noise $\langle n(\mathbf{x})n(\mathbf{y})\rangle = \sigma^2(\mathbf{x})\Delta_{\mathrm{P}}^2\delta_D(\mathbf{x}-\mathbf{y})$, the covariance matrix of the deconvolved noise field $X(\mathbf{x})$ is written as

$$\begin{aligned}
\langle X(\mathbf{x})X(\mathbf{y})\rangle &= \langle \varphi_x^{-1}[n(\mathbf{x})]\varphi_y^{-1}[n(\mathbf{y})]\rangle \\
&= \left\langle \varphi_x^{-1}[n(\mathbf{x})]\int \frac{d^2\boldsymbol{\ell}}{2\pi}\frac{d^2\mathbf{y}'}{2\pi} e^{i\boldsymbol{\ell}\cdot(\mathbf{y}-\mathbf{y}')}\frac{n(\mathbf{y}')}{\varphi_{\boldsymbol{\ell}}}\right\rangle \\
&= \varphi_x^{-1}\left[\int \frac{d^2\boldsymbol{\ell}}{2\pi}\frac{d^2\mathbf{y}'}{2\pi} e^{i\boldsymbol{\ell}\cdot(\mathbf{y}-\mathbf{y}')}\frac{1}{\varphi_{\boldsymbol{\ell}}}\langle n(\mathbf{x})n(\mathbf{y}')\rangle\right] \\
&= \varphi_x^{-1}\left[\sigma^2(\mathbf{x})\Delta_{\mathrm{P}}^2\int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} e^{i\boldsymbol{\ell}\cdot(\mathbf{y}-\mathbf{x})} e^{\frac{\ell^2\sigma_b^2}{2}}\right],
\end{aligned} \quad \text{(B4)}$$

where we have exchanged the order of deconvolution and ensemble average at the 3rd equal sign, since deconvolution is a linear operator.

$$\begin{aligned}
\langle X(\mathbf{x})X(\mathbf{y})\rangle &= \Delta_{\mathrm{P}}^2\int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2}\frac{d^2\mathbf{k}}{2\pi} e^{i\boldsymbol{\ell}\cdot\mathbf{x}} e^{i\mathbf{k}\cdot\mathbf{y}}\varphi_{\boldsymbol{\ell}}\varphi_{\mathbf{k}}\left[\int \frac{d^2\mathbf{x}'}{2\pi} e^{-i(\boldsymbol{\ell}+\mathbf{k})\cdot\mathbf{x}'}\delta^2(\mathbf{x}')\right] \\
&= \Delta_{\mathrm{P}}^2\int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2}\frac{d^2\mathbf{k}}{2\pi} e^{i\boldsymbol{\ell}\cdot\mathbf{x}} e^{i\mathbf{k}\cdot\mathbf{y}}\varphi_{\boldsymbol{\ell}}\varphi_{\mathbf{k}}(\delta^2)_{\boldsymbol{\ell}+\mathbf{k}} \\
&= \Delta_{\mathrm{P}}^2\int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} e^{i\boldsymbol{\ell}\cdot\mathbf{x}}\varphi_{\boldsymbol{\ell}}\int \frac{d^2\mathbf{k}}{2\pi} e^{i\mathbf{k}\cdot\mathbf{y}}\varphi_{\mathbf{k}}(\delta^2)_{\boldsymbol{\ell}+\mathbf{k}}.
\end{aligned} \quad \text{(B5)}$$

APPENDIX C: INVERSE OF COVARIANCE MATRIX

The inverse covariance matrix Σ_r^{-1} evaluation is the key to the r likelihood in Eq. (10). To avoid repeating the similar computation for every different r, we can single out the r dependence rewriting the covariance matrix in the form $\Sigma_r = \Sigma^{\mathrm{en}} + r\Sigma^{\mathrm{b}}$, where

$$\Sigma^{\mathrm{en}} = \begin{pmatrix} \tilde{\Sigma}^{CE,CE} + \Sigma^{NQ,NQ} & \tilde{\Sigma}^{CE,SE} \\ \tilde{\Sigma}^{CE,SE} & \tilde{\Sigma}^{SE,SE} + \Sigma^{NU,NU} \end{pmatrix},$$

$$\Sigma^{\mathrm{b}} = \begin{pmatrix} \tilde{\Sigma}^{SB^0,SB^0} & \tilde{\Sigma}^{CB^0,SB^0} \\ \tilde{\Sigma}^{CB^0,SB^0} & \tilde{\Sigma}^{CB^0,CB^0} \end{pmatrix}.$$

Both Σ^{en} and Σ^{b} are symmetric and positive definite. We first decompose Σ^{b} as $\Sigma^{\mathrm{b}} = V\Lambda V^{\intercal}$, with Λ being a diagonal

matrix composed of its eigenvalues, and V being a matrix composed of its eigenvectors. Now we do a little manipulation to the covariance matrix

$$\Sigma_r = \Sigma^{\rm en} + rV\Lambda V^\intercal = V\sqrt{\Lambda}\left(\sqrt{\Lambda^{-1}}V^\intercal\Sigma^{\rm en}V\sqrt{\Lambda^{-1}} + rI\right)\sqrt{\Lambda}V^\intercal, \quad \text{(C1)}$$

where we have used the orthogonality $V^\intercal = V^{-1}$. One more eigendecomposition, $\sqrt{\Lambda^{-1}}V^\intercal\Sigma^{\rm en}V\sqrt{\Lambda^{-1}} = \hat{V}\hat{\Lambda}\hat{V}^\intercal$, enables us further transform Σ_r as

$$\Sigma_r = V\sqrt{\Lambda}\hat{V}(\hat{\Lambda} + rI)\hat{V}^\intercal\sqrt{\Lambda}V^\intercal = V\sqrt{\Lambda}\hat{V}(\hat{\Lambda} + rI)(V\sqrt{\Lambda}\hat{V})^\intercal. \quad \text{(C2)}$$

Here we can obtain the inverse matrix Σ_r^{-1} at little cost, using the orthogonality of V and $\hat{V}$. And more beautifully, all the matrices V, Λ and $\hat{V}, \hat{\Lambda}$ have no r dependence; hence, we obtain the inverse covariance matrix as a function of r at the same computation cost of a single inverse matrix computation.

[1] V. Mukhanov and G. Chibisov, JETP Lett. **33**, 532 (1981).
[2] A. H. Guth, Phys. Rev. D **23**, 347 (1981).
[3] A. D. Linde, Phys. Lett. B **108**, 389 (1982).
[4] A. Albrecht and P. J. Steinhardt, Phys. Rev. Lett. **48**, 1220 (1982).
[5] J. E. Lidsey, A. R. Liddle, E. W. Kolb, E. J. Copeland, T. Barreiro, and M. Abney, Rev. Mod. Phys. **69**, 373 (1997).
[6] D. H. Lyth and A. Riotto, Phys. Rep. **314**, 1 (1999).
[7] A. A. Starobinskii, J. Exp. Theor. Phys. Lett. **30**, 682 (1979).
[8] V. A. Rubakov, M. V. Sazhin, and A. V. Veryaskin, Phys. Lett. **115B**, 189 (1982).
[9] R. Fabbri and M. D. Pollock, Phys. Lett. **125B**, 445 (1983).
[10] L. F. Abbott and M. B. Wise, Nucl. Phys. **B244**, 541 (1984).
[11] A. Starobinskii, Sov. Astron. Lett. **11**, 133 (1985).
[12] A. Stebbins, arXiv:9609149v1.
[13] M. Kamionkowski, A. Kosowsky, and A. Stebbins, Phys. Rev. Lett. **78**, 2058 (1997).
[14] M. Kamionkowski, A. Kosowsky, and A. Stebbins, Phys. Rev. D **55**, 7368 (1997).
[15] U. Seljak and M. Zaldarriaga, Phys. Rev. Lett. **78**, 2054 (1997).
[16] U. Seljak, Astrophys. J. **482**, 6 (1997).
[17] M. Zaldarriaga and U. Seljak, Phys. Rev. D **55**, 1830 (1997).
[18] J. Khoury, B. A. Ovrut, N. Seiberg, P. J. Steinhardt, and N. Turok, Phys. Rev. D **65**, 086007 (2002).
[19] J. Khoury, B. A. Ovrut, P. J. Steinhardt, and N. Turok, Phys. Rev. D **64**, 123522 (2001).
[20] J. Khoury, P. J. Steinhardt, and N. Turok, Phys. Rev. Lett. **91**, 161301 (2003).
[21] P. J. Steinhardt and N. Turok, Phys. Rev. D **65**, 126003 (2002).
[22] L. A. Boyle, P. J. Steinhardt, and N. Turok, Phys. Rev. D **69**, 127302 (2004).
[23] BICEP2/Keck Array and Planck Collaborations, Phys. Rev. Lett. **114**, 101301 (2015).
[24] Planck Collaboration XX, Astron. Astrophys. **594**, A20 (2016).
[25] BICEP2/Keck Array Collaborations, Phys. Rev. Lett. **116**, 031302 (2016).
[26] COrE Collaboration, arXiv:1102.2181.
[27] LiteBird Collaboration, J. Low Temp. Phys. **176**, 733 (2014).
[28] H. Ishino *et al.*, Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave, in *Proceedings of SPIE* (2016), Vol. 99040X.
[29] CMB-S4 Collaboration, arXiv:1610.02743.
[30] G. Cabass, L. Pagano, L. Salvati, M. Gerbino, E. Giusarma, and A. Melchiorri, Phys. Rev. D **93**, 063508 (2016).
[31] M. Kamionkowski and E. D. Kovetz, Annu. Rev. Astron. Astrophys. **54**, 227 (2016).
[32] J. Delabrouille *et al.* (f. t. C. Collaboration), J. Cosmol. Astropart. Phys. 04 (2018) 014.
[33] R. Hildebrand, J. Dotson, C. Dowell, D. Schleuning, and J. Vaillancourt, Astrophys. J. **516**, 834 (1999).
[34] B. T. Draine, arXiv:0304488.
[35] A. Benoît *et al.*, Astron. Astrophys. **424**, 571 (2004).
[36] M. J. Mortonson and U. Seljak, J. Cosmol. Astropart. Phys. 10 (2014) 035.
[37] M. D. Niemack *et al.*, J. Low Temp. Phys. **184**, 746 (2016).
[38] Planck Collaboration XXX, Astron. Astrophys. **586**, A133 (2016).
[39] Planck Collaboration L, arXiv:1606.07335.
[40] N. Krachmalnicoff, C. Baccigalupi, J. Aumont, M. Bersanelli, and A. Mennella, Astron. Astrophys. **588**, A65 (2016).
[41] D. Hanson *et al.*, Phys. Rev. Lett. **111**, 141301 (2013).
[42] POLARBEAR Collaboration, Phys. Rev. Lett. **112**, 131302 (2014).
[43] A. van Engelen *et al.*, Astrophys. J. **808**, 7 (2015).
[44] K. T. Story *et al.*, Astrophys. J. **810**, 50 (2015).
[45] Planck Collaboration XV, Astron. Astrophys. **594**, A15 (2016).
[46] A. Lewis and A. Challinor, Phys. Rep. **429**, 1 (2006).
[47] B. D. Sherwin and M. Schmittfull, Phys. Rev. D **92**, 043005 (2015).
[48] L. Knox and Y.-S. Song, Phys. Rev. Lett. **89**, 011303 (2002).
[49] M. Kesden, A. Cooray, and M. Kamionkowski, Phys. Rev. Lett. **89**, 011304 (2002).
[50] U. Seljak and C. M. Hirata, Phys. Rev. D **69**, 043005 (2004).
[51] G. Simard, D. Hanson, and G. Holder, Astrophys. J. **807**, 166 (2015).

[52] Planck Collaboration XI, Astron. Astrophys. **594**, A11 (2016).
[53] Planck Collaboration XIII, Astron. Astrophys. **594**, A13 (2016).
[54] Y.-S. Song, A. Cooray, L. Knox, and M. Zaldarriaga, Astrophys. J. **590**, 664 (2003).
[55] H. Dole, G. Lagache, J.-L. Puget, K. I. Caputi, N. Fernández-Conde, E. Le Floc'h, C. Papovich, P. G. Pérez-González, G. H. Rieke, and M. Blaylock, Astron. Astrophys. **451**, 417 (2006).
[56] Planck Collaboration XVIII, Astron. Astrophys. **571**, A18 (2014).
[57] P. Larsen, A. Challinor, B. D. Sherwin, and D. Mak, Phys. Rev. Lett. **117**, 151102 (2016).
[58] A. Manzotti *et al.*, Astrophys. J. **846**, 45 (2017).
[59] W. Hu, Astrophys. J. **557**, L79 (2001).
[60] W. Hu and T. Okamoto, Astrophys. J. **574**, 566 (2002).
[61] C. M. Hirata and U. Seljak, Phys. Rev. D **67**, 043001 (2003).
[62] C. M. Hirata and U. Seljak, Phys. Rev. D **68**, 083002 (2003).
[63] E. Anderes, L. Knox, and A. van Engelen, Phys. Rev. D **83**, 043523 (2011).
[64] E. Anderes, B. D. Wandelt, and G. Lavaux, Astrophys. J. **808**, 152 (2015).
[65] M. Millea, E. Anderes, and B. D. Wandelt, arXiv:1708.06753.
[66] J. Carron and A. Lewis, Phys. Rev. D **96**, 063510 (2017).
[67] Planck Collaboration XXX, Astron. Astrophys. **586**, A133 (2016).
[68] J. Carron, A. Lewis, and A. Challinor, J. Cosmol. Astropart. Phys. 05 (2017) 035.
[69] K. M. Smith, D. Hanson, M. Loverde, C. M. Hirata, and O. Zahn, J. Cosmol. Astropart. Phys. 06 (2012) 014.
[70] T. Louis, S. Næss, S. Das, J. Dunkley, and B. Sherwin, Mon. Not. R. Astron. Soc. **435**, 2040 (2013).
[71] J. Lesgourgues, arXiv:1104.2932.
[72] E. Anderes, Phys. Rev. D **88**, 083517 (2013).