

# Optimal constraints on primordial gravitational waves from the lensed CMB

Julien Carron\*

*Department of Physics & Astronomy, University of Sussex, Brighton BN1 9QH, United Kingdom*

(Received 13 September 2018; published 13 February 2019)

We demonstrate how to obtain optimal constraints on a primordial gravitational wave component in lensed cosmic microwave background (CMB) data under ideal conditions. We first derive an estimator of the tensor-to-scalar ratio,  $r$ , by using an error-controlled close approximation to the exact posterior, under the assumption of Gaussian primordial CMB and lensing deflection potential. This combines fast internal iterative lensing reconstruction with optimal recovery of the unlensed CMB. We evaluate its performance on simulated low-noise polarization data targeted at the recombination peak. We carefully demonstrate our  $r$ -posterior estimate is optimal and shows no significant bias, making it the most powerful estimator of primordial gravitational waves from the CMB. We compare these constraints to those obtained from  $B$ -mode band-power likelihood analyses on the same simulated data, before and after map-level quadratic estimator delensing, and iterative delensing. Internally, iteratively delensed band powers are only slightly less powerful on average (by less than 10%), promising close-to-optimal constraints from a stage IV CMB experiment.

DOI: [10.1103/PhysRevD.99.043518](https://doi.org/10.1103/PhysRevD.99.043518)

## I. INTRODUCTION

After the completion of the *Planck* Cosmic Microwave Background (CMB) mission [1], the major target of the CMB community has now become precise measurement of the CMB polarization. The magnetic (B) part of the CMB polarization [2,3] on degree scales is a unique signature of the stochastic background of primordial gravitational waves produced during inflation [4,5]. Constraints on the tensor-to-scalar power spectrum ratio,  $r$ , are expected to increase by 2 orders of magnitude in precision within the next decade: CMB Stage IV (CMB-S4<sup>1</sup>) has forecast sensitivity down to  $r \sim 5 \times 10^{-4}$ , using delensed  $B$ -mode band powers after foreground cleaning [6]. Lensing of the CMB photons by large-scale structures generates  $B$  polarization that effectively appears as an approximately white cosmic variance noise [7,8]. To reach such tight constraints on the primordial signal, successful delensing of this  $5 \mu\text{K}$ -arc min noise is mandatory. Currently and for the next few years, the most faithful lensing tracer at the scales relevant for  $B$ -mode delensing is the cosmic infrared background (CIB) [9], able to achieve 40% delensing on 60% of the sky [10], and possibly more in areas that are clean from galactic dust. It is also possible to combine the CIB with other large-scale structure tracers [11,12], or with the CMB internal reconstruction [10], in order to increase its fidelity to the CMB lensing field. Lensing estimates for

CMB-S4 will be dominated by the internal reconstruction, using polarization quadratic estimators [13] or more powerful iterative estimators, first introduced by Ref. [14]. At the low instrumental noise levels expected for CMB-S4, iterative internal estimation from CMB polarization has been demonstrated on simulated data to give lensing reconstructions that are more than 90% cross-correlated to the true lensing [15,16].

One may ask whether it could be possible, at least in principle, to do even better than these forecasts. The lensing deflections introduce non-Gaussianities in the form of higher-order statistics in the CMB temperature and polarization [8], which are used to reconstruct the lensing signal [17]. Delensing will remove part of the non-Gaussianity, but only imperfectly, and some amount of information must remain beyond the power spectra. Hence, it is plausible that there may be room for alternative statistics that compress more information than delensed  $B$ -mode band powers.

This paper has two main purposes. The first is to demonstrate how to obtain directly the posterior probability density (PDF) for  $r$ , from lensed CMB data. The posterior contains all the information on  $r$ , and constraints based on it are optimal. The second is to compare this optimal method to band-power likelihood analysis. Finding a posterior width in agreement with naive expectations will confirm current forecasting methods and our understanding of how well CMB experiments can constrain primordial gravitational waves.

Our approach uses an approximate, analytic marginalization of the large-scale structure lensing to build the

\*j.carron@sussex.ac.uk  
<sup>1</sup><https://cmb-s4.org>.

statistics of interest, here the tensor-to-scalar ratio,  $r$ . This analytic marginalization is a fairly natural choice and has been used already in a different context by Ref. [14], in which the aim was to obtain an optimal estimator of the lensing spectrum. In this paper, we generalize the analytic marginalization to an error-controlled variational approximation and provide a rigorous discussion of its accuracy, showing that corrections are negligible for our purposes and the experimental configurations investigated. Variational principles have a long history in cosmology, at least dating back to Peebles's minimum action principle to reconstruct large-scale motions [18].

We work in the flat-sky approximation. All our simulations use the *Planck* 2015 cosmology [19], with power spectra generated with the CAMB [20] software. We use square maps of area  $645 \text{ deg}^2$ , assuming periodic boundary conditions for simplicity, with pixels of 1.5 arc min on a side. Foreground cleaning is a major challenge to the quest for primordial gravitational waves. We do not consider these complications in this paper, assuming throughout we are working with foreground-cleaned maps with white noise power spectra. We always use white Gaussian noise levels of  $\sqrt{2} \cdot 1.5 \mu\text{K}\cdot\text{arc min}$  in polarization, an instrument beam of 3 arc min, and consider CMB multipoles below  $\ell = 3000$  only. The minimum multipole we probe in our flat-sky implementation of the sky patches is  $\ell_{\min} = 14$ , excluding the reionization peak. We consider three different levels of tensor modes, with tensor-to-scalar ratio  $r_{\text{in}}$  defined at the pivot scale  $k = 0.05/\text{Mpc}$  and a vanishing tensor spectral index. The first has  $r_{\text{in}} = 0.05$ , close to current constraints  $r_{0.002} < 0.062$  (cf. 95%) from BICEP2/KECK array BK15 data in combination with *Planck* [21,22]. For the configuration just described, the nominal band powers are enough for a strong detection. Second,  $r_{\text{in}} = 0.01$ , in which case the nominal band powers cannot detect the waves decisively but the delensed band powers can. Third, a vanishing amplitude  $r_{\text{in}} = 0.0$ , where in all cases only upper limits can be placed from the data.

The paper is built as follows. Section II describes our approximation scheme to the exact posterior density function, and Sec. III gives details on our numerical implementation. In Sec. IV, we discuss our nominal and delensed band-powers likelihood and implementation. We present in Sec. V our results and summarize and conclude in Sec. VI. One Appendix details a couple of technical points for completeness.

## II. POSTERIOR FOR TENSOR-TO-SCALAR RATIO

Given CMB Stokes parameter polarization data  $X^{\text{dat}} = (Q^{\text{dat}}, U^{\text{dat}})$ , and for a uniform prior on  $r$ , the posterior is proportional to the likelihood

$$p(r|X^{\text{dat}}) \propto p(X^{\text{dat}}|r). \quad (2.1)$$

Owing to the lensing by large-scale structures, the CMB probability density function on the right-hand side is non-Gaussian in  $X^{\text{dat}}$  and does not have a simple analytical description. We may write, however, marginalizing over possible lensing deflection maps,

$$p(X^{\text{dat}}|r) = \int \mathcal{D}\phi p_\phi[\phi] p(X^{\text{dat}}|\phi, r), \quad (2.2)$$

where  $p_\phi$  is the probability density of the lensing potential, defined by the large-scale structure evolution. When the lensing map is known, the lensed CMB is still Gaussian, since the deflections just remap points on the sky to very good approximation [23]. Hence, we can consider  $p(X^{\text{dat}}|\phi, r)$  Gaussian in  $X^{\text{dat}}$ , with covariance determined by the lensed spectra of the CMB together with the specified amount of the tensor mode, transfer function, and noise covariance matrix. Neglecting for simplicity the tiny cross-correlation between  $\phi$  and the CMB  $E$  polarization [24], which is too small to impact our results,<sup>2</sup> we can write explicitly

$$\ln p(X^{\text{dat}}|\phi, r) = -\frac{1}{2} X^{\text{dat},\dagger} \text{Cov}_\phi^{-1} X^{\text{dat}} - \frac{1}{2} \ln \det \text{Cov}_\phi, \quad (2.3)$$

with pixel-space covariance matrix

$$\text{Cov}_\phi = \mathcal{B} D_\phi C^{\text{unl, fid}} D_\phi^\dagger \mathcal{B}^\dagger + N. \quad (2.4)$$

Here,  $\mathcal{B}$  is the transfer function including instrument beam,  $D_\phi$  is the lensing deflection operator that maps the unlensed CMB Stokes parameters to the ones deflected by  $\nabla\phi$ ,  $C^{\text{unl, fid}}$  is our set of *unlensed* fiducial CMB spectra, and  $N$  is the noise matrix. In position space and on the flat sky,  $D_\phi$  has the explicit representation  $(D_\phi X)(\mathbf{x}) \equiv X(\mathbf{x} + \nabla\phi(\mathbf{x}))$ .

The model specified by Eq. (2.3) neglects the very small effect of lensing of the polarization generated from reionization, so a single, common lensing deflection  $D_\phi$  can be used. The CMB spectra contain the dependence on tensor modes

$$C_\ell^{\text{unl, fid}} \equiv C_\ell^{\text{scal, fid}} + r C_\ell^{\text{tens, fid}}. \quad (2.5)$$

We pick the prior on lensing maps  $p_\phi$  in Eq. (2.2) to be Gaussian with power  $C_L^{\phi\phi, \text{fid}}$ , calculated from the nonlinear matter power.<sup>3</sup> Owing to nonlinear evolution in the late Universe,  $p_\phi$  is not exactly Gaussian. However, nonlinear effects are weak at scales relevant for the lensing  $B$  modes (the nonlinear contribution to the degree-scale lensing  $B$  power is only a few percent), and generally the effects of

<sup>2</sup>We do include properly all cross-correlations including  $C_\ell^{\phi E}$  in all our simulations.

<sup>3</sup>We follow standard practice of denoting lensing multipoles with  $L$  and CMB multipoles with  $\ell$ .

non-Gaussian lenses are not expected to bias the lensing reconstruction from polarization [25,26]. Hence, this choice is unlikely to significantly bias our results. In practice, this assumption could simply be explicitly tested by performing the reconstruction of this paper using lensing maps from realistic  $N$ -body simulations.

The integrand in Eq. (2.2) peaks at the most probable lensing map, given the data and candidate value  $r$  for the tensor amplitude. Let  $\phi^*(X^{\text{dat}}, r)$  be this most probable potential map. We may then write

$$e^{-S[\phi]} \equiv p_\phi[\phi] p(X^{\text{dat}}|\phi, r) \quad (2.6)$$

and expand the action  $S[\phi]$  to second order around  $\phi^*$ ,

$$S[\phi] \sim S[\phi^*(r)] + \frac{1}{2}(\phi - \phi^*(r))H_r[\phi^*(r)](\phi - \phi^*(r)). \quad (2.7)$$

Both the lensing potential  $\phi^*$  and the curvature  $H_r$  at this point have explicit dependence on the data map and explicit (through the CMB spectra) and implicit [through  $\phi^*(r)$ ] dependence on  $r$ . In the following, we suppress the data dependence for notational clarity.

The lensing map marginalization, Eq. (2.2), becomes trivial under this approximation. The result is, up to terms independent from  $r$ ,

$$\ln p(r|X^{\text{dat}}) \simeq -S[\phi^*(r)] - \frac{1}{2} \ln \det H_r[\phi^*(r)], \quad (2.8)$$

which forms the basis of our investigations. Leading corrections or alternative expansions are discussed below in Sec. II A.

In the remainder of this subsection, we just note how  $S[\phi^*]$  is related to the  $B$ -mode power, by looking at its dependence on  $r$ . Let us introduce the Wiener-filtered CMB maps  $X^{\text{WF}} = (E^{\text{WF}}, B^{\text{WF}})$  through

$$X^{\text{WF}} \equiv C^{\text{unl, fid}} D_\phi^\dagger \mathcal{B}^\dagger \text{Cov}_\phi^{-1} X^{\text{dat}}. \quad (2.9)$$

These are the most probable unlensed CMB maps given the observed Stokes data and given the fiducial model entering the covariance  $\text{Cov}_\phi$ . In particular, since the scalar unlensed  $B$  power vanishes,  $B^{\text{WF}}$  is the most probable (maximum *a posteriori*) map of the  $B$  tensor modes, assuming the fiducial value of  $r$  and lensing map  $\phi$  are the truth. By definition,

$$\left. \frac{\delta S[\phi]}{\delta \phi} \right|_{\phi=\phi^*(r)} = 0; \quad (2.10)$$

hence,  $dS/dr = \partial S/\partial r$  for this lensing map. For this reason, the prior term does not contribute to the action  $r$  derivative, only the likelihood defined in Eq. (2.3). It then follows (neglecting the tensor  $E$  contribution) that

$$\frac{dS[\phi^*(r)]}{d \ln r} = \frac{1}{2} \sum_\ell \frac{|B_\ell^{\text{WF}}(r)|^2 - \langle |B_\ell^{\text{WF}}(r)|^2 \rangle}{r C_\ell^{BB, \text{tens}}}. \quad (2.11)$$

The first term on the right-hand side comes from the quadratic part in the CMB likelihood. The sum  $\sum_\ell$  runs over all two-dimensional frequencies of our sky patch. The second term (the average of the first over data realizations) comes from the log-determinant  $\ln \det \text{Cov}_\phi$   $r$  derivative.<sup>4</sup>

All dependence on the lensing map prior in the reconstructed  $r$  posterior in Eq. (2.11) is absorbed into the lensing map reconstruction  $\phi^*$ . If the lensing map were exactly known, then  $B^{\text{WF}}$  is the properly delensed  $B$ -mode map, there is no marginalization over the lensing map, and Eq. (2.11) directly gives the (derivative of  $\log$ ) $r$  posterior by trivial comparison to the expected power.

### A. Corrections and alternative approximations

Equation (2.8) is an approximation of the posterior PDF, that relies on the Gaussianity of the reconstructed lensing map. Lens reconstruction from the CMB is nonlinear in the data, and some non-Gaussian features are expected, and visible, in standard lens reconstructions with perfectly Gaussian true input lensing [27]. If the approximation is not good enough, resulting constraints might be biased, or suboptimal, possibly both. Hence, it is very useful to be able to assess the size of the leading corrections. We discuss now how corrections can be evaluated, which also gives us alternative expansion schemes. Section III D later on demonstrates by using these tools that Eq. (2.8) is accurate and unbiased.

We can proceed as follows. The exact posterior in Eq. (2.2) integrates  $e^{-S[\phi]}$ . Using an arbitrary trial action  $S_t$  as an approximation to  $S$ , we may write the exact identity

$$\int \mathcal{D}\phi e^{-S[\phi]} = \left( \int \mathcal{D}\phi e^{-S_t[\phi]} \right) \langle e^{-\Delta S[\phi]} \rangle_t, \quad (2.12)$$

where the average is with respect to probability density  $e^{-S_t}$  and  $\Delta S = S - S_t$  is the mismatch between the true and trial actions. If choosing for  $S_t$  the natural expansion Eq. (2.7), the first term on the right-hand side (the normalization factor of the density  $e^{-S_t}$ ) is our approximation Eq. (2.8), and the second term encapsulates all errors. The point is that whenever  $S_t$  is chosen quadratic in  $\phi$  this error term is an average over Gaussian lensing maps. We can simulate these maps and estimate this term by averaging over simulations. Unless the approximation is extremely good, in practice, we can never probe directly the exponential with a reasonable number of Monte Carlos, but

<sup>4</sup>This term is most easily derived by realizing that, on average, the likelihood variation vanishes,  $\langle \partial_r \ln p(X^{\text{dat}}|\phi, r) \rangle_{X^{\text{dat}}} = \partial_r \langle p(X^{\text{dat}}|\phi, r) \rangle_{X^{\text{dat}}} = 0$ , since  $p$  is a properly normalized probability density for any value of  $r$ . Hence, Eq. (2.11) must average to zero.

we obtain in this paper leading contributions. Using a standard cumulant expansion, we can write the asymptotic expansion

$$\ln \langle e^{-\Delta S} \rangle_t \sim \langle -\Delta S \rangle_t + \frac{1}{2} (\langle (\Delta S)^2 \rangle_t - \langle \Delta S \rangle_t^2) + \dots \quad (2.13)$$

In particular, all expansions of the form

$$S_t[\phi] = S[\phi^*] + \frac{1}{2} (\phi - \phi^*)^\dagger H (\phi - \phi^*) \quad (2.14)$$

for an arbitrary  $H$  matrix result in (including here only the leading cumulant corrections)

$$\ln p(r|X^{\text{dat}}) \sim -S[\phi^*] - \frac{1}{2} \ln \det H - \kappa_1 + \frac{1}{2} \kappa_2, \quad (2.15)$$

where  $\kappa_1$  and  $\kappa_2$  are the mean and variance of  $\Delta S$  from a Gaussian ensemble of lensing maps with inverse covariance  $H$ . This provides alternative expansion schemes and useful consistency checks, in which some of the difficulties of dealing with the exact curvature can be eliminated by using a simpler matrix, at the cost of a larger cumulant correction (see Sec. III D).

### III. POSTERIOR EVALUATION

This section describes the implementation of the different terms in the posterior, Eq. (2.8). Two large log-determinants must be evaluated. The determinant of the CMB data covariance matrix is discussed in Sec. III A and is in fact fairly harmless. The second, the determinant of the lensing curvature matrix, discussed in Sec. III B, is more complex and is the most expensive step of the entire implementation. However, we find that its dependence on the data realization is very weak and can be neglected. For this reason, it only needs to be calculated once for a simulation suite. All terms must be evaluated at the best lensing map  $\phi^*$ . The production of this map together with the Wiener-filtered CMB maps  $X^{\text{WF}}$  is described in Sec. III C. Finally, we give in Sec. III D details on our generation of Gaussian lensing map samples with large and dense covariance matrices, as required for evaluation of the cumulant corrections.

#### A. Data covariance determinant

How can one estimate log-determinants of very large, dense matrices like  $\text{Cov}_\phi$ ? There is no trivial diagonalization, so the large size of the matrix renders brute force methods totally useless. Possibilities include integral representations combined with Monte Carlo evaluation of the diagonal [28]. We are only interested in the  $r$  dependence. This allows us to simplify the problem quite a bit. We first note that

$$c(r) \equiv -\frac{\partial}{\partial r} \frac{1}{2} \ln \det \text{Cov}_{\phi^*(r)} = \frac{1}{2} \text{Tr} \text{Cov}_{\phi^*(r)}^{-1} \partial_r \text{Cov}_{\phi^*(r)},$$

where  $\partial_r \text{Cov}_\phi$  only depends on the covariance matrix constructed from tensor CMB spectra.<sup>5</sup> Using Monte Carlo simulations, we can write an unbiased estimator as

$$\hat{c}(r) = \frac{1}{2} X^\dagger \text{Cov}_{\phi^*(r)}^{-1} \partial_r \text{Cov}_{\phi^*(r)} X, \quad (3.1)$$

where  $X$  are unit spectra random variables with  $\langle X_i X_j^\dagger \rangle = \delta_{ij}$ . It is easily seen that the estimator is unbiased. Its Monte Carlo (MC) noise variance (when using Gaussian  $X$ ) is simply the Fisher information on  $r$ , divided by the number of simulations used. We can further reduce the variance with the help of a reference ideal covariance matrix, denoted with a subscript 0, for which we know the determinant: consider the modified estimator

$$\hat{c}(r) - \hat{c}_0(r) + c_0(r), \quad (3.2)$$

the last term  $c_0(r) \equiv \langle \hat{c}_0(r) \rangle$  being known analytically. With a good isotropic approximation to the covariance, we might expect to be able to reduce the MC noise of the original estimate substantially. Figure 1 shows in the upper panel the Monte Carlo error  $\Delta \hat{c}(r)/|c(r)|$  as a function of  $r$ , using the raw estimator  $\hat{c}(r)$  (blue line) of Eq. (3.1), with the lensed spectra as reference (orange) in Eq. (3.2) and using unlensed spectra (green) as reference. More specifically, neglecting again the tensor  $E$  contribution, these last two choices correspond to

$$c_0^{\text{len}}(r) = \frac{1}{2} \sum_{\ell} \frac{(2\ell + 1) C_\ell^{\text{BB,tens}}}{r C_\ell^{\text{BB,tens}} + C_\ell^{\text{BB,len}} + C_\ell^{\text{BB,noise}}} \quad (3.3)$$

in the lensed case and

$$c_0^{\text{unl}}(r) = \frac{1}{2} \sum_{\ell} \frac{(2\ell + 1) C_\ell^{\text{BB,tens}}}{r C_\ell^{\text{BB,tens}} + C_\ell^{\text{BB,noise}}} \quad (3.4)$$

in the unlensed case. In these two equations,  $2\ell + 1$  is shorthand notation for the exact number of modes of our flat-sky patch, and  $C_\ell^{\text{BB,noise}}$  is the beam-deconvolved noise spectrum. In the unlensed case, it is apparent that with a single simulation we can reach well below percent accuracy on the log-determinant.

In fact, for our experimental configuration at least, the isotropic, unlensed spectra determinant approximation is extremely accurate, and the small deviation from it can be very well captured by the leading term of a perturbative expansion in powers of  $\phi$ . This is shown in the lower panel of Fig. 1, which displays the relative deviation of the exact determinant  $c(r)$  from its isotropic counterpart calculated

<sup>5</sup>The derivative  $\partial_r$  does not act on  $\phi^*(r)$ ; we are interested in the explicit  $r$  dependence only, for the reasons explained at the end of Sec. II.

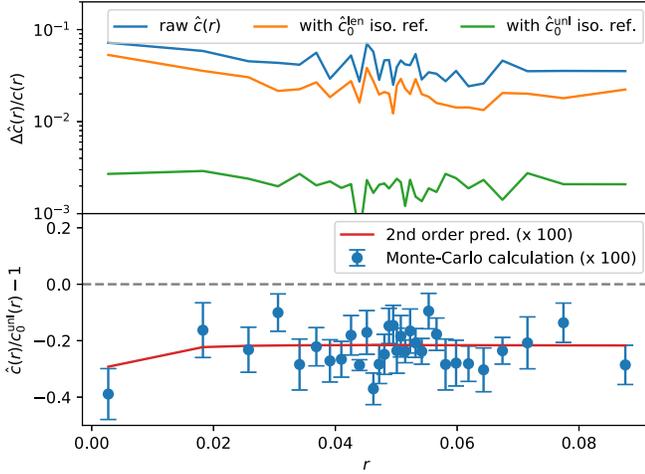


FIG. 1. Upper panel: The three curves show the Monte Carlo noise  $\Delta\hat{c}(r)/c(r)$  (one simulation equivalent) root variance for three estimators  $\hat{c}(r)$  of the lensed data covariance  $r$  derivative, Eq. (3.2). The estimator accelerated by the isotropic reference estimator using unlensed spectra and neglecting lensing (green) is already accurate to well below a percent with a single Monte Carlo estimate. Lower panel: The red curve shows the relative deviation of the analytic perturbative prediction to the isotropic, unlensed approximation  $c_0^{\text{unl}}(r)$  to  $\hat{c}(r)$  (multiplied by a factor 100), Eq. (3.4). Blue points show estimates of the exact  $c(r)$  from independent Monte Carlo simulations at each  $r$  point, where the iterative lensing solution  $\phi^*(r)$  was obtained using ten iterations starting from the quadratic estimator. It is apparent that the exact result is both very close to the isotropic estimation and correctly captured by the analytical perturbative expansion. The error bars, independent from point to point, are the empirical standard deviations across nine Monte Carlo simulations. The figure was built using a data realization with input  $r^{\text{in}} = 0.05$  where the  $r$  grid is densest.

with unlensed spectra (blue points with error bars) and the perturbative prediction (red line). More explicitly, the perturbative prediction is

$$\frac{1}{2} \ln \det \text{Cov}_\phi = \frac{1}{2} \ln \det \text{Cov}_{\phi=0} + \frac{1}{2} \sum_L R_L |\phi_L|^2, \quad (3.5)$$

where the linear, mean-field response matrix  $R$  is obtained in the Appendix and the first term on the right-hand side is evaluated with the unlensed CMB spectra. At first sight, it might appear counterintuitive that the unlensed spectra prediction is so accurate, but this is easily explained. In the limit of vanishing instrumental noise and perfect resolution, the data covariance (2.3) reduces to

$$\text{Cov}_\phi \xrightarrow[\text{high res}]{\text{low noise}} D_\phi C^{\text{unl, fid}} D_\phi^\dagger.$$

Averaged over lensing maps, this gives the CMB spectra *lensed* by  $C^{\phi\phi}$ . However, taking the determinant, one expects the factorization (e.g., after imposing a sufficiently high band limit, making all these operators square matrices)

$$\ln \det \text{Cov}_\phi \xrightarrow[\text{high res}]{\text{low noise}} \ln \det D_\phi D_\phi^\dagger + \ln \det C^{\text{unl, fid}}.$$

The first determinant has no dependence on  $r$  and is an irrelevant constant. Hence, usage of the unlensed spectra gives under these conditions the exact result irrespective of the amount of lensing in the data. Figure 1 demonstrates that in our instrumental configuration the mean-field response  $R$  in Eq. (3.5) captures very well the tiny residual coupling between these terms, the transfer function, and the noise matrix.

## B. Hessian determinant

The Hessian  $H$  is defined as the second variation of the log-posterior  $S$  of the lensing map, Eq. (2.7). Operationally speaking, there are three terms  $H \equiv H^{\text{dat}} + H^{\text{Cov}} + H^{\text{Pri}}$ . The first originates from the quadratic piece of the CMB likelihood and is data realization dependent,

$$H_{LL'}^{\text{dat}} \equiv \frac{1}{2} \frac{\delta^2}{\delta\phi_L \delta\bar{\phi}_{L'}} X^{\text{dat}} \text{Cov}_\phi^{-1} X^{\text{dat}} \Big|_{\phi=\phi^*(r)}; \quad (3.6)$$

the second also comes from the likelihood but is data independent,

$$H_{LL'}^{\text{Cov}} \equiv \frac{1}{2} \frac{\delta^2}{\delta\phi_L \delta\bar{\phi}_{L'}} \ln \det \text{Cov}_\phi \Big|_{\phi=\phi^*(r)}; \quad (3.7)$$

and the third is the lensing map prior curvature, dominant on small scales in which the data do not constrain the lensing deflection field. Under our choice of Gaussian statistics for  $\phi$ , this prior term is trivially given by

$$H_{LL'}^{\text{Pri}} = \frac{\delta_{LL'}}{C_L^{\phi\phi, \text{fid}}}. \quad (3.8)$$

The calculation of the total Hessian determinant is the main difficulty in getting the  $r$  posterior. It plays a key role in shaping the final posterior and cannot be neglected but is difficult to obtain exactly. However, as shown further below, the dependence of the log-determinant on the data realization can be neglected for all practical purposes. Hence, this calculation only needs to be performed once for each model.

In general, we solve for  $\ln \det H$  by coupling the same trace probing method described in Sec. III A to a double layered conjugate gradient inversion. We use for  $H^{\text{Cov}}$  our accurate approximation from Sec. III A for the covariance determinant, with the trivial result

$$H_{LL'}^{\text{Cov}} = \delta_{LL'} R_L. \quad (3.9)$$

The main operational difficulty lies in the application of  $H^{\text{dat}}$  to an arbitrary lensing potential map  $\phi$ . The details of this calculation are deferred to the Appendix. There, it is shown that one can apply the Hessian matrix to a lensing potential vector at the cost of Wiener filtering one pair of

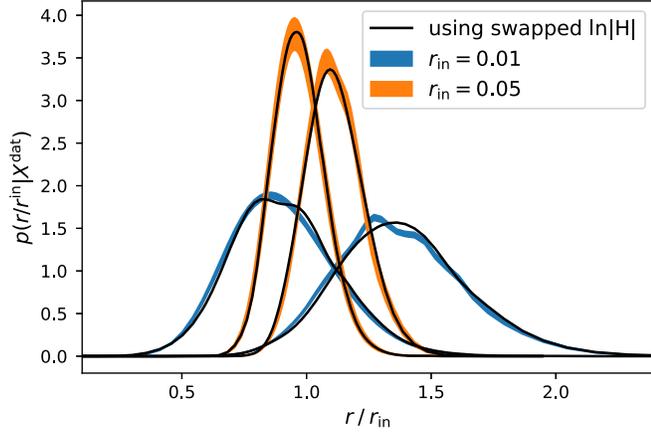


FIG. 2. Data dependence of the Hessian log-determinant. Comparison of the  $r$ -posterior reconstruction on two simulated data maps spanning  $645 \text{ deg}^2$  with  $2.1 \mu\text{K}$ -arc min polarization sensitivity. The filled colors use the expensive, realization-dependent curvature determinant estimate. The width shows the  $2\sigma$  Monte Carlo uncertainties from the determinant derivative calculation. For each realization, errors on opposite sides of the peak of the posterior are strongly anticorrelated. The black lines show the same reconstructions but swap the log-determinant between the two realizations (black lines), with virtually identical results. For the cases we investigated, the determinant needs only be calculated on a single data realization.

Stokes  $Q$ ,  $U$  maps. This operation is itself performed via conjugate gradient inversion. Using a simple diagonal preconditioner for the outer conjugate gradient inversion, we found that we need slightly above ten iterations and thus the inverse filtering of the same number of polarization maps to conservatively solve for  $\phi^\dagger H^{-1} \partial_r H \phi$  to five significant figures. The same operations must then be repeated for each Monte Carlo  $\phi$  used to probe the trace and at each point  $r$  of interest. For the area of  $\sim 645 \text{ deg}^2$  and the resolution of  $1.5 \text{ arc min}$  dealt with in this paper, a single probe of the trace at some  $r$  point takes a couple of minutes. The final Monte Carlo error on the PDFs depends on the  $r$  density of trace-probing MCs, rather than the number of MCs per  $r$  point. Hence, it can be advantageous to use a dense  $r$  grid in which each point is sparsely trace probed. For the PDFs shown in this work, in which we reconstruct accurately the entire PDF shape, we typically use  $64 r$  points, denser near the peak, and four MCs per point, which gives a crude but still reasonable estimate of the local error.

Figure 2 shows a couple of posterior reconstructions on simulated data: one pair with input  $r^{\text{in}} = 0.01$  (orange) and one pair with  $r^{\text{in}} = 0.05$  (blue). The filled colors show the reconstructions using the realization-dependent Hessian estimate, in which the width displays the  $2\sigma$  uncertainties resulting from the Monte Carlo determinant evaluation. These errors are independent only for the estimation of the posterior  $r$  derivative; after normalization to build the posterior itself, this produces a strong anticorrelation of errors between points on the opposite side of the peak on

this figure. The black lines show the corresponding curves, but with Hessian estimation swapped between the two realizations. Summary statistics are virtually identical, and this demonstrates that for all practical purposes we can neglect the data dependence of the log-determinant. However, the dependence on the lensing potential  $\phi^*$  remains, and we do not offer in this work a trivial, fully isotropic approximation.

### C. Iterative lens estimation

The first step in the  $r$ -posterior reconstruction is always the calculation of the iterative solution  $\phi^*$  together with the Wiener-filtered CMB maps  $X^{\text{WF}}$ . We follow Ref. [16] very closely, maximizing the lensing map posterior probability for the lensing deflections using a quasi-Newton iterative descent. At each iteration step, the current lensing estimate induces a mean-field ( $\langle \phi \rangle$ , in standard lensing estimation terminology) term to subtract from the quadratic estimate, originating from the first variation (with respect to  $\phi$ ) of the covariance matrix determinant. This term is small on all scales for lens reconstruction from polarization, and we use the same accurate analytical approximation discussed in Sec. III A and in the Appendix instead of simulations.

### D. Higher-order cumulants calculation

Evaluation of the correction terms in Eq. (2.13) requires sampling Gaussian lensing maps with inverse covariance given by the curvature matrix chosen for the posterior expansion. This matrix is very large and has no trivial structure and hence is difficult to sample from, and standard generic methods such as Cholesky decomposition are useless. It is possible, however, to apply this matrix to a lensing map in reasonable time, and powerful methods can be used to construct samples with the correct covariance structure using this property. We proceed by deforming continuously the inverse square root of  $H$  as follows [29]. We connect  $H$  to the identity matrix  $\mathbb{I}$ , defining

$$H_t \equiv \mathbb{I} + t(H - \mathbb{I}), \quad (3.10)$$

and introduce  $\phi_t \equiv H_t^{-1/2} \phi_0$ , where  $\phi_0$  is a unit variance Gaussian vector of the appropriate dimensionality. By definition,  $H_{t=1} = H$  and  $\phi_{t=1}$  has covariance  $H^{-1}$ . Differentiating gives the following ordinary differential equation (ODE) for  $\phi_t$ :

$$\frac{d\phi_t}{dt} = -\frac{1}{2}(H - \mathbb{I})H_t^{-1}\phi_t \quad (3.11)$$

The solution of this ODE at  $t = 1$  is then by construction the desired Gaussian sample.<sup>6</sup> We have used this machinery

<sup>6</sup>After discretization, the algorithm is reminiscent of the Van Cittert deconvolution of image analysis [30,31].

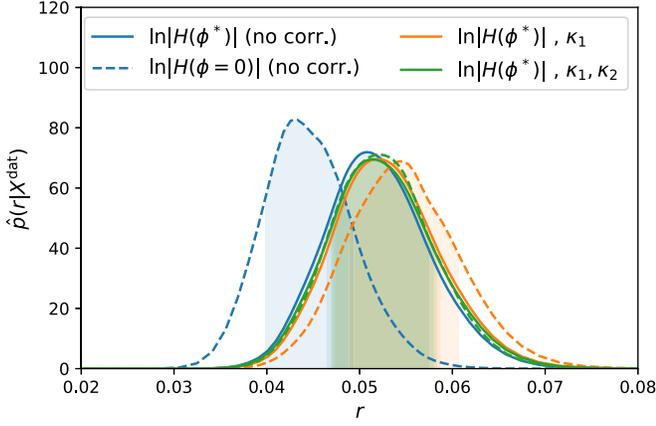


FIG. 3. Corrections to posterior and alternative expansion. Tensor-to-scalar ratio  $r$ -posterior estimates for one data realization (with input value  $r_{\text{in}} = 0.05$ ), without any cumulant correction (solid blue), including first (solid orange) and second (solid green) cumulant corrections. The dashed lines show the same curves, but use a different posterior expansion scheme, with a less accurate but simpler curvature matrix neglecting the lensing. The agreement between the two methods is very good and provides a good consistency check, but in the latter case, the inclusion of the cumulant corrections becomes mandatory. The dashed bands contain 68% of the probability.

to test corrections to our posterior approximation (2.8) on one data realization. In this case,

$$H_{LL'} = \frac{\delta^2 S[\phi]}{\delta\phi_L \delta\phi_{L'}} \Big|_{\phi=\phi^*(r)}. \quad (3.12)$$

For a consistency check, we also checked the alternative choice of curvature in which the lensing deflections are neglected,

$$H_{LL'} = \frac{\delta^2 S[\phi]}{\delta\phi_L \delta\phi_{L'}} \Big|_{\phi=0}. \quad (3.13)$$

The second case considerably speeds up the calculation of the curvature log-determinant, but the posterior is expected to be less accurate. Figure 3 shows the different  $r$ -posterior estimates, for one data realization with input  $r^{\text{in}} = 0.05$ . The solid lines show our baseline approximation, without corrections (blue), with the first cumulant (orange), and with the first two cumulants (green). We have used four Monte Carlo samples per  $r$  point, on a grid devised as just described in Sec. III B. We use in all cases a standard low-order Bulirsch-Stoer algorithm [32] to solve the ODE defined in Eq. (3.11). The bracketed inverse is solved with conjugate-gradient descent. The corrections have a completely negligible effect on the posterior. The dashed lines show the same curves with the choice of unlensed curvature. The agreement, including the corrections, is very good. However, the leading approximation is clearly worse,

and the corrections are essential in bringing the two expansion schemes in agreement. In the remainder of this paper, we stick to the lensed curvature matrix and safely neglect the cumulant corrections.

#### IV. NOMINAL AND DELENSED $B$ -MODE BAND POWERS

We now describe our reconstruction of  $r$  from raw and delensed band powers. The likelihood of CMB data temperature and polarization band powers has been discussed at length in several places already, and efficient parameterizations both at low and high multipoles are well known, e.g., Refs. [33,34]. Only preliminary work exists, however, for delensed band powers [35]. After performing consistency checks described in the next section, we decided to use the Hamimeche and Lewis likelihood [34], using analytical band-power predictions and covariance matrices as described below. We note that other possibilities such as the parametrization of Ref. [36] seem to work just as well. We consider  $B$ -mode multipoles only up to  $\ell_{\text{max}} = 200$  and use as a baseline the maximal number of bins (74) that our flat-sky mode structure allows when building the  $r$  posteriors.

All of our  $B$ -spectra likelihoods may eventually be written as

$$-2 \ln p(r | \hat{C}^{BB}) \equiv g(x_\ell) \hat{C}_\ell^{BB} \Sigma_\ell^{-1} \hat{C}_\ell^{BB} g(x_\ell),$$

with  $g(x) \equiv \text{sign}(x-1) \sqrt{2(x-\ln x-1)}$  and  $x_\ell(r) = \hat{C}_\ell^{BB} / C_\ell^{BB}(r)$ . These power spectra include the tensor contribution, residual lensing power, and beam-deconvolved noise. The predictions of the spectra  $C_\ell^{BB}(r)$  are evaluated analytically, using the perturbative expression for the lensing  $B$ -mode power to linear order in the appropriate lensing spectrum. For the nominal band powers, we use the fiducial lensing spectrum  $C_L^{\phi, \text{fid}}$ . In the case of the quadratic estimator or iteratively delensed band powers, we first empirically measure the cross-correlation coefficient squared  $\rho_L^2$  of the reconstructed lensing maps to the true input from a small set of simulations. We then use as a lensing spectrum  $C_L^{\phi, \text{fid}}(1 - \rho_L^2)$ . We neglect the tiny lensing effects on the tensor spectra. The lensing maps themselves are obtained as described in Sec. III C, with the difference that we exclude the tensor modes scales  $\ell \leq 200$  to build the tracers. Including these modes would dramatically complicate the analysis; if the delensed data and lensing tracer contain common multipoles, the delensed  $B$  modes would contain very strong, spurious delensing signature originating from disconnected four-points and six-points statistics of the CMB [37–39]. The reason is that the lensing estimator cannot distinguish between true and random lensing signatures in the data (such as shear or magnification) and all of these signatures are removed after delensing, leading to a spuriously unlensed-looking CMB at

the two-point level. The loss of signal to noise in the lensing tracer by excising the largest angular scales is small; the delensing efficiency  $\rho_L^2$  is reduced by roughly 1% on the most relevant scales  $L \sim 500$ .

To delens the CMB, we use a simple remapping technique. From the Wiener-filtered lensing deflection estimate  $\alpha$ , we simply remap the Stokes parameter (after beam deconvolution and discarding multipoles higher than 2000) according to the lensing inverse deflection field  $\alpha^{-1}$ , defined through the condition that deflected points  $\mathbf{x} + \alpha(\mathbf{x})$  get remapped back to themselves,

$$\mathbf{x} + \alpha^{-1}(\mathbf{x} + \alpha(\mathbf{x})) \equiv \mathbf{x}. \quad (4.1)$$

This inversion is performed by exactly following Ref. [16], using a fast-converging Newton-Raphson solver. This delensing technique differs operationally and conceptually from the template method used for instance by the SPTpol team on their polarization data [40], which uses a Wiener-filtered  $E$ -mode map to build a  $B$  template, which is then subtracted from the data. The filters of the template method are optimized to minimize the resulting  $B$  power at linear order [9]. While the remapping method naturally contains higher-order terms, the noise is also remapped, which can result in higher total  $B$  power if this is high such as for Planck data [10,38]. However, we found that there is basically no difference between the two methods for the low levels of noise here. By default, we include the noise remapping into our delensed  $B$ -power predictions, but this is very small. Finally, there are no  $E/B$  separation complications [41,42] since we use periodic patches throughout this paper and the spectra estimator  $\hat{C}_\ell^{BB}$  are trivial to build from the delensed or nominal Stokes maps.

For the covariance matrix, we use the simple approximation [43,44] combining the Gaussian part with corrections according to the lensing kernels,

$$\begin{aligned} \Sigma_{\ell_1 \ell_2} = & \delta_{\ell_1 \ell_2} \sigma_{\ell_1}^2(C^{BB}) + \sum_{\ell'} \frac{\partial C_{\ell_1}^{BB}}{\partial C_{\ell'}^{EE, \text{unl}}} \sigma_{\ell'}^2(C^{EE, \text{unl}}) \frac{\partial C_{\ell_2}^{BB}}{\partial C_{\ell'}^{EE, \text{unl}}} \\ & + \sum_L \frac{\partial C_{\ell_1}^{BB}}{\partial C_L^{\phi\phi}} \sigma_L^2(C^{\phi\phi}) \frac{\partial C_{\ell_2}^{BB}}{\partial C_L^{\phi\phi}}, \end{aligned} \quad (4.2)$$

with Gaussian spectrum variance

$$\sigma_{\ell'}^2(C^{XX}) \equiv \frac{2(C^{\ell' XX})^2}{2\ell' + 1}, \quad (4.3)$$

where  $2\ell' + 1$  really stands for the number of multipoles in our flat-sky patch. The derivatives are evaluated to first order in  $C_L^{\phi\phi, \text{fid}}$ , or in  $C_L^{\phi\phi, \text{fid}}(1 - \rho_L^2)$  for the delensed band powers. On the relevant scales  $\ell \leq 200$ , the covariance correlation coefficients are small and of minor importance, and the perturbative expansion is accurate enough. Collecting reconstructions from simulations, we checked

that the diagonal of the covariance matrices matches the prediction to within a couple of percent at least.

## V. RESULTS

We have performed our baseline posterior reconstruction, including realization-dependent determinant calculation,

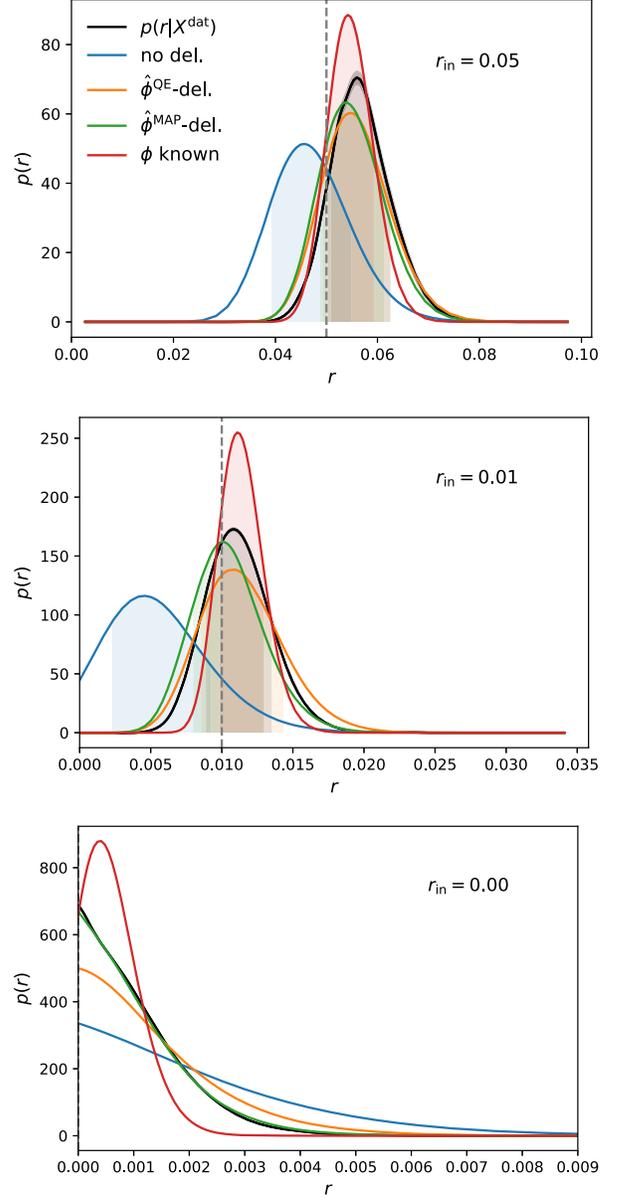


FIG. 4. Posterior constraints on the tensor-to-scalar ratio  $r$  for different reconstruction methods on one data realization with an input  $r_{\text{in}} = 0.05, 0.01,$  and  $0$  from top to bottom. The black line shows our new estimator. The green, orange, and blue lines show the constraints from nominal, quadratic estimator delensed, and iteratively delensed  $B$ -mode band powers, respectively, with likelihood built as described in the text. For comparison, the red line shows the case of perfect knowledge of the input lensing map, with posterior given by Eq. (5.1). In the upper two panels, the dashed bands contain 68% of the probability.

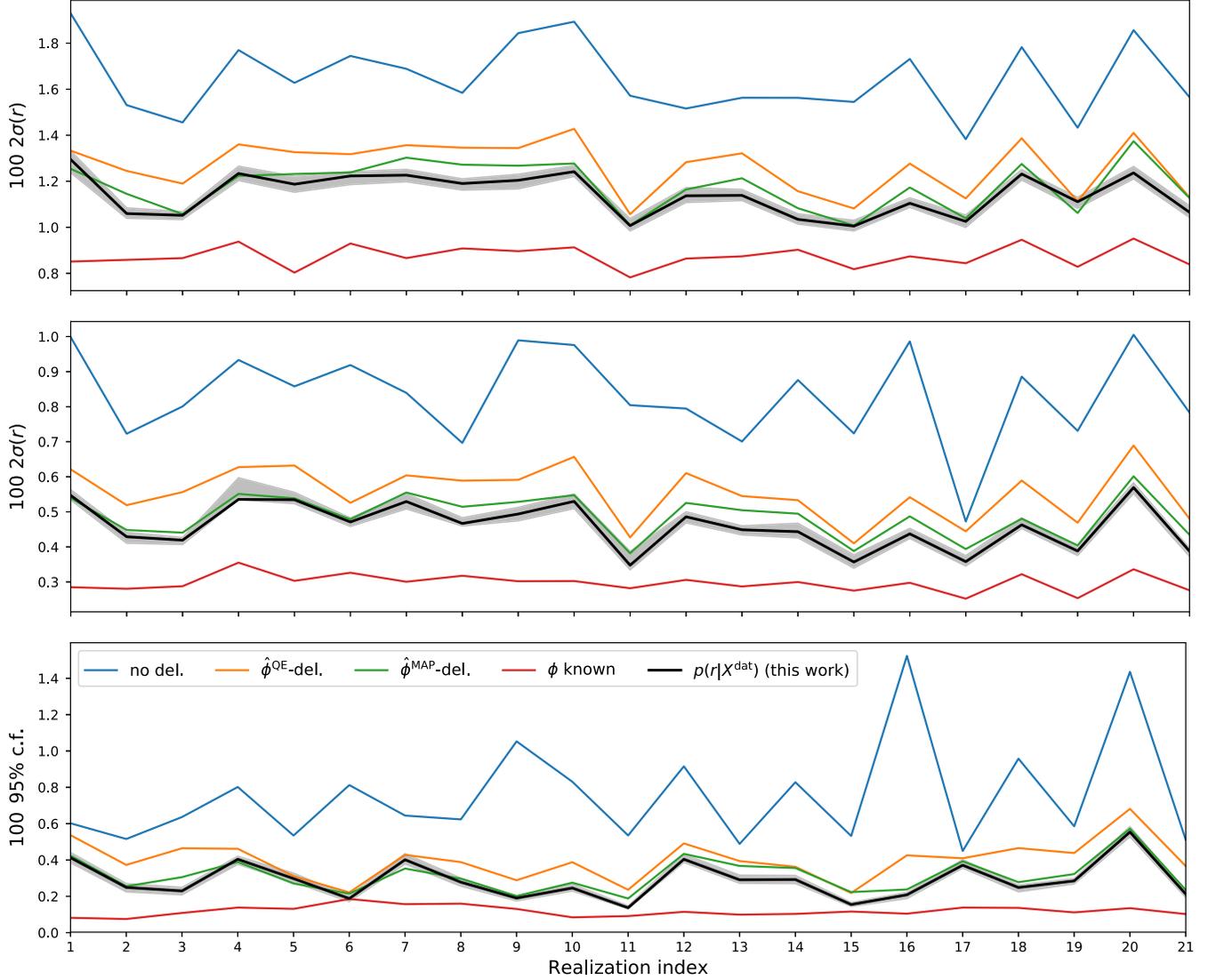


FIG. 5. Comparison of constraints on the tensor-to-scalar ratio  $r$  obtained from band-power analysis and the  $r$  posterior obtained in this work (black), for 21 simulated data maps. Shown are constraints from nominal  $B$ -mode band powers (blue), quadratic estimator delensed band powers (orange), and iteratively delensed band powers (green). The red curves show the constraints when the lensing deflection potential is perfectly known for comparison. The grey shaded area shows the  $2\sigma$  uncertainty on the constraint due to the Monte Carlo evaluation of one of the deflection Hessian curvature determinants in the posterior. The three panels use maps with different input tensor amplitudes  $r_{\text{in}} = 0.05, 0.01, \text{ and } 0$  (top to bottom). We quote two standard deviations (multiplied by 100) in the first two panels, where the tensor modes are well constrained, and the 95% confidence limit (also multiplied by 100) in the lowest panel with vanishing input tensor amplitude. All reconstructions are performed on  $645 \text{ deg}^2$ , with a white polarization noise level of  $2.12 \mu\text{K-arc min}$ .

on 21 CMB simulations for each of the three different levels  $r_{\text{in}} = 0.05, 0.01, 0$  of tensor modes advertised in the Introduction. For each realization, we also performed likelihood analyses from the quadratic-estimator delensed, iteratively delensed  $B$ -mode band powers and with no delensing. Figure 5 shows the following main results. For each simulation, the constraints obtained from either our posterior  $p(r|X^{\text{dat}})$  approximation are displayed (black), together with those obtained from the different band-power analyses (colored lines). Blue shows the results from the nominal (no delensing) band-power likelihood, orange is after

delensing with the Wiener-filtered quadratic estimator ( $\phi^{\text{QE}}$ ), and green is after delensing with the iterative [formally, the maximum *a posteriori* (MAP) solution  $\phi^{\text{MAP}}$ ] lensing solution. Finally, for comparison, the red line shows the constraints achievable if the lensing deflections were known perfectly. Up to irrelevant constants, this latter case is given by

$$\ln p(r, |X^{\text{dat}}, \phi^{\text{in}}) \propto -\frac{1}{2} X^{\text{dat}, \dagger} \text{Cov}_{\phi^{\text{in}}}^{-1} X^{\text{dat}} - \frac{1}{2} \ln \det \text{Cov}_{\phi^{\text{in}}}, \quad (5.1)$$

where  $\phi^{\text{in}}$  is the lensing potential input to data realization  $X^{\text{dat}}$ . The three panels show  $r_{\text{in}} = 0.05, 0.01, \text{ and } 0$  from top to bottom. In the first two panels, the quoted numbers are the  $2\sigma$  width of the PDF, and the lowest panel shows the 95% confidence upper limit. The grey area around the black line shows the  $2\sigma$  uncertainty originating from the Monte Carlo measurement of the Hessian determinant. We evaluate these errors simply by sampling an ensemble of posteriors from the Monte Carlo log-determinant's errors. Figure 4 shows explicitly the  $r$ -posterior PDF for one of the realizations using the same color scheme. As is visible there, all reconstructions show skewness and other non-Gaussian features; hence, these summary statistics are only an imperfect representation of the constraints on  $r$ .

All of our exact posterior estimates mean values lie within  $1.8\sigma$  ( $r^{\text{in}} = 0.05$ ) of the true value. The series of  $r^{\text{in}} = 0.01$  simulations shows a single outlier, with recovered posterior mean  $100 \langle r \rangle = 0.53$  formally lower than the input by  $2.5\sigma$ , but with a PDF showing substantial skew with no strong evidence of an anomaly. We can try to isolate a possible systematic bias in our posterior approximation by combining all posteriors as if they were independent measurements of  $r$ . Doing so, we obtain the constraints

$$100 \hat{r} = 5.15 \pm 0.12 \quad (100 r^{\text{in}} = 5, \text{ combined}) \quad (5.2)$$

$$100 \hat{r} = 1.04 \pm 0.05 \quad (100 r^{\text{in}} = 1, \text{ combined}), \quad (5.3)$$

both perfectly consistent with the input values. This constrains a systematic bias to a small fraction of a standard deviation. The same test on all the band-power likelihoods or the known  $\phi$  likelihood [Eq. (5.1)] also shows consistency.

It is apparent that there is significant spread in the summary statistics displayed on Fig. 5. As expected, this is more pronounced for the nominal band-power analysis which has the weakest constraints. While the different methods constraining power do rank as expected on average, the results using the iteratively delensed band powers in green and our full posterior reconstruction in black are very close, with a few reconstructions showing nominally slightly worse summary statistics from the posterior reconstruction. These cases seem compatible with the errors on the posterior. Also, since the exact band-power likelihood is intractable, it is very difficult to assess precisely the impact of our band-power likelihood approximation on each realization. In all cases, iterative delensing does perform better than quadratic estimator delensing, but the improvement shows substantial realization dependence as well and in a couple of cases can be absolutely minimal. Table I shows the summary statistics averaged across all these simulations. We find that the exact posterior outperforms the iteratively delensed band powers by 3%, 6%, and 8% for  $r^{\text{in}} = 0.05, 0.01, \text{ and } 0$ , respectively.

TABLE I. Summary statistics comparison on tensor-to-scalar ratio constraints averaged over our data realizations, for the different methods tested in this work and three input values of tensor modes as indicated in the first row. The reconstructions are performed on maps of  $645 \text{ deg}^2$  with  $2.12 \mu\text{K-arc min}$  polarization noise. The first three rows show the results of  $B$ -mode band-powers likelihood analysis, without delensing, with quadratic estimator delensing, and with iterative delensing. The next row shows the constraints from the exact posterior density function on  $r$ , obtained as described in this work. The last row shows the case of a perfectly known lensing map for comparison. We quote twice the standard deviation for the first two columns where  $r$  is well constrained and the 95% confidence limit in the last column.

$100 r^{\text{in}}$	5.0	1.0	0.0
Band powers, no delensing	1.65 ( $2\sigma$ )	0.83 ( $2\sigma$ )	0.75 (95% cf.)
Band powers, $\phi^{\text{QE}}$ delensing	1.27	0.56	0.40
Band powers, $\phi^{\text{MAP}}$ delensing	1.18	0.49	0.31
Exact posterior	1.14	0.46	0.29
$\phi$ known posterior	0.87	0.30	0.12

## VI. SUMMARY

Unless the primordial  $B$ -mode power produced during inflation is very large, sophisticated analysis techniques such as delensing will be essential to provide the best constraints on primordial gravitational waves. We have presented a new estimator, based on a close approximation to the exact posterior of the tensor-mode amplitude. By careful Monte Carlo investigations of corrections to the approximation, we have demonstrated that it is unbiased and very close to optimal, providing the tightest possible constraints on primordial gravitational waves from CMB data. The estimator uses fast, joint estimation of the best lensing deflection map and of the unlensed CMB.

This first investigation used a simplified setting, including periodic sky patches, no analysis mask, and use of homogeneous noise, facilitating both the iterative lens reconstruction and the unlensed  $E$ - $B$  recovery from the observed lensed Stokes polarization data. These assumptions did not play a key role in obtaining our results, since the presence of the lensing deflections and data realization dependence break isotropy and prevent the existence of trivial basis to work with. Use of Monte Carlo simulations and inversion methods akin to conjugate-gradient seem unavoidable. Lens reconstruction and Wiener-filtering on masked data have already been demonstrated successfully [16] with the same methods, at the cost of a manageable increase in execution time.

The posterior reconstruction has dependency on two aspects of the cosmological model used to define the CMB likelihood: the unlensed CMB scalar perturbations spectra and the lensing potential power spectrum. Changes in the lensing spectrum (or lensing deflection map prior) impact

slightly the optimal lens reconstruction, and changes in the scalar spectra impact the unlensed  $E$  and  $B$  polarization field reconstruction. Hence, formally, use of a slightly different fiducial model might lead to a slightly different result. However, all these power spectra are extremely well constrained in practice from observations, including the lensing spectrum, so this is unlikely to bring significant biases. Furthermore, if necessary, it is possible to include the uncertainty in the power spectra in the posterior, by obtaining the linear response to the spectra and extending in this way the posterior density, in analogy to the way state-of-art lensing spectrum reconstruction likelihoods are built [45,46]. One can also go a step further and marginalize directly over these using the empirical spectra, as demonstrated by Ref. [10]. We also restricted our analysis to the extraction of the tensor-to-scalar ratio assuming a fixed template shape of the tensor spectrum. However, this is not a limitation of this approach; using high-quality observations, the obvious generalization of this framework will be able to distinguish features as well.

We have compared the performance of our estimator to that of more traditionally planned  $B$ -mode band-powers extraction. We found standard, well-demonstrated analytical likelihood models are able to describe meaningfully the delensed band powers, and we have used these likelihoods on nominal, quadratic estimator delensed, and iteratively delensed band powers. The performances of these delensed band powers do match naive expectations. For the configuration studied here, targeting the recombination peak of the  $B$ -mode spectrum with noise levels in line with expectations from a CMB Stage IV experiment, our new estimator does outperform the iteratively delensed band powers by a realization-dependent amount, also depending on the exact value of  $r$ , reaching 8% on average for small values. Producing the posterior PDF for  $r$  as we did is more expensive numerically than producing band powers. Nevertheless, we demonstrated in this paper that the analysis was possible, providing constraints optimal by construction and improving prospects of detecting a tantalizing component of modern cosmology.

### ACKNOWLEDGMENTS

I thank Antony Lewis for many discussions and useful comments on the manuscript, Anthony Challinor for early discussions that eventually lead to the material presented this paper, and Jesús Torrado for discussions on Gaussian vectors with large dense covariance matrices. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant No. 616170. This research used resources of the

National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

### APPENDIX: LENSING MAP POSTERIOR CURVATURE

This Appendix describes in more details the lensing map posterior curvature matrix  $H = H_r[\phi^*(r)]$ , defined as the matrix of second variation of the lensing map log-posterior,

$$H_{LL'} \equiv \left. \frac{\delta^2 S[\phi]}{\delta\phi_L \delta\bar{\phi}_{L'}} \right|_{\phi=\phi^*(r)}, \quad \text{with}$$

$$S[\phi] \equiv \frac{1}{2} X^{\text{dat}} \text{Cov}_\phi^{-1} X^{\text{dat}} + \frac{1}{2} \ln \det \text{Cov}_\phi + \frac{1}{2} \sum_L \frac{|\phi_L|^2}{C_L^{\phi\phi, \text{fid}}}, \quad (\text{A1})$$

where the first two terms on the right-hand side come from the lensing map likelihood and the last term comes from the Gaussian prior. In particular, in this Appendix, we obtain the result that the data-dependent part of the matrix can be applied in linear time to a vector (at the cost of one CMB Wiener filtering). This makes the calculation of its log-determinant and inverse possible. The data-independent part poses no particular problems. The Gaussian prior curvature is trivial, and for the second term in Eq. (A1), we adopt the same very accurate approximation as in Sec. III A,

$$\frac{1}{2} \ln \det \text{Cov}_\phi$$

$$\simeq \frac{1}{2} \ln \det \text{Cov}_{\phi=0} + \frac{1}{2} \sum_{LL'} \phi_L \bar{\phi}_{L'} \left. \frac{\delta^2 \frac{1}{2} \ln \det \text{Cov}_\phi}{\delta\phi_L \delta\bar{\phi}_{L'}} \right|_{\phi=0}$$

$$\equiv \frac{1}{2} \ln \det \text{Cov}_{\phi=0} + \frac{1}{2} \sum_L R_L |\phi_L|^2, \quad (\text{A2})$$

where  $R_L$  is the  $\phi$ -induced mean-field linear response (hence also the desired curvature term), obtained analytically following Ref. [16] and reproduced briefly at the end of this Appendix.

To obtain the data-dependent part of the curvature, we calculate for convenience the curvature  $H_{\text{dat}}^{ab}(\mathbf{x}, \mathbf{y})$  with respect to the two position-space deflection components  $\alpha_a(\mathbf{x})$ ,  $\alpha_b(\mathbf{y})$  of the flat sky, where  $\alpha = \nabla\phi$ . Once this is done, it is straightforward to apply the matrix to a lensing potential vector by expanding it into these two components, applying  $H^{ab}$ , and reprojecting eventually onto the potential Fourier harmonics. The data-dependent curvature splits into two terms, which we describe next:

$$\begin{aligned}
H_{\text{dat}}^{ab}(\mathbf{x}, \mathbf{y}) &= \frac{1}{2} X^{\text{dat}} \cdot \text{Cov}_{\phi}^{-1} \frac{\delta \text{Cov}_{\phi}}{\delta \alpha_a(\mathbf{x})} \text{Cov}_{\phi}^{-1} \frac{\delta \text{Cov}_{\phi}}{\delta \alpha_b(\mathbf{y})} \text{Cov}_{\phi}^{-1} X^{\text{dat}} + (a, \mathbf{x} \leftrightarrow b, \mathbf{y}) \quad (\equiv 2H_F^{ab}(\mathbf{x}, \mathbf{y})) \\
&\quad - \frac{1}{2} X^{\text{dat}} \cdot \text{Cov}_{\phi}^{-1} \frac{\delta^2 \text{Cov}_{\phi}}{\delta \alpha_a(\mathbf{x}) \delta \alpha_b(\mathbf{y})} \text{Cov}_{\phi}^{-1} X^{\text{dat}} \quad (\equiv H_{\text{det}}^{ab}(\mathbf{x}, \mathbf{y})). \quad (\text{A3})
\end{aligned}$$

When the data covariance precisely matches  $\text{Cov}_{\phi}$  (that is, for our simulations, for fiducial  $r$  values close to the true value in  $X^{\text{dat}}$ ), the term  $H_F$  coincides, on average, with the Fisher information matrix on the displacement field. Under these conditions, and setting  $\phi \equiv 0$ , it is precisely the isotropic inverse lensing reconstruction noise  $N_L^{(0)}$  for the deflection angle calculated with the unlensed CMB spectra in the weights [13]. The second term  $H_{\text{det}}$  is subdominant; on average, and again when the data covariance matches  $\text{Cov}_{\phi}$ , this term would be canceled by a contribution from  $R_L$ . This term  $R_L$  would also cancel one of the two  $H_F$  factors in Eq. (A3). The complete relation Eq. (A1) for the lensing potential Fourier modes under these conditions is

$$\begin{aligned}
\langle H_{LL'} \rangle_{X^{\text{dat}}} &= \delta_{LL'} \left( \frac{1}{N_L^{(0)}} + \frac{1}{C_L^{\phi \phi, \text{fid}}} \right) \\
&\quad (\text{when } \phi \equiv 0, \text{ and } \langle X^{\text{dat}} X^{\text{dat}, \dagger} \rangle = \text{Cov}_{\phi=0}), \quad (\text{A4})
\end{aligned}$$

which can serve as a useful consistency check of the data-dependent curvature calculation.

To proceed, we need the first and second derivatives of the covariance. To simplify notation, we introduce

$$\begin{aligned}
\xi^{\phi} &\equiv D_{\phi} C^{\text{unl}} D_{\phi}^{\dagger}, & \xi_{,a}^{\phi} &\equiv D_{\phi} \nabla_a C^{\text{unl}} D_{\phi}^{\dagger}, & \text{and} \\
\xi_{,ab}^{\phi} &\equiv D_{\phi} \nabla_a \nabla_b C^{\text{unl}} D_{\phi}^{\dagger}, \quad (\text{A5})
\end{aligned}$$

where the operator  $\nabla_a C^{\text{unl}}$  is defined by the block-diagonal matrix Fourier representation  $\delta_{ll'} i l_a \begin{pmatrix} C_{\ell}^{EE, \text{unl}} & 0 \\ 0 & r C_{\ell}^{BB, \text{tens}} \end{pmatrix}$ , and similarly for  $\nabla_a \nabla_b C^{\text{unl}}$  with an additional factor  $i l_b$ . The following relations hold:

$$\frac{\delta \text{Cov}_{\phi}^{ij}}{\delta \alpha_a(\mathbf{x})} = \mathcal{B}(\mathbf{x}_i, \mathbf{x})(\xi_{,a}^{\phi} \mathcal{B}^{\dagger})(\mathbf{x}, \mathbf{x}_j) - (\mathcal{B}_{\xi, a}^{\phi})(\mathbf{x}_i, \mathbf{x})(\mathcal{B}^{\dagger})(\mathbf{x}, \mathbf{x}_j), \quad (\text{A6})$$

and

$$\begin{aligned}
\frac{\delta^2 \text{Cov}_{\phi}^{ij}}{\delta \alpha_a(\mathbf{x}) \delta \alpha_b(\mathbf{y})} &= \mathcal{B}(\mathbf{x}_i, \mathbf{x})(\xi_{,ab}^{\phi} \mathcal{B}^{\dagger})(\mathbf{y}, \mathbf{x}_j) \delta^D(\mathbf{x} - \mathbf{y}) \\
&\quad - \mathcal{B}(\mathbf{x}_i, \mathbf{x})(\xi_{,ab}^{\phi})(\mathbf{x}, \mathbf{y})(\mathcal{B}^{\dagger})(\mathbf{y}, \mathbf{x}_j) \\
&\quad - (\mathcal{B})(\mathbf{x}_i, \mathbf{y})(\xi_{,ab}^{\phi})(\mathbf{y}, \mathbf{x})(\mathcal{B}^{\dagger})(\mathbf{x}, \mathbf{x}_i) \\
&\quad + (\mathcal{B}_{\xi, ab}^{\phi})(\mathbf{x}_i, \mathbf{y})(\mathcal{B}^{\dagger})(\mathbf{x}, \mathbf{x}_j) \delta^D(\mathbf{x} - \mathbf{y}). \quad (\text{A7})
\end{aligned}$$

There are implicit sums over Stokes indices in the above equations. We introduce further the following notation for the inverse-variance weighted CMB maps ( $\bar{X}$ ) and Wiener-filtered CMB ( $X_{\phi}^{\text{WF}}$ ):

$$\bar{X} \equiv \mathcal{B}^{\dagger} \text{Cov}_{\phi}^{-1} X^{\text{dat}} \quad \text{and} \quad X_{\phi}^{\text{WF}} \equiv \xi^{\phi} \mathcal{B}^{\dagger} \text{Cov}_{\phi}^{-1} X^{\text{dat}} = \xi^{\phi} \bar{X}. \quad (\text{A8})$$

The subscript  $\phi$  on the Wiener-filtered CMB is present to emphasize that these are the lensed Wiener-filtered de-lensed CMB maps, in contrast to Eq. (2.9). With this, we may write (full indices)

$$\begin{aligned}
\left[ \frac{\delta \text{Cov}_{\phi}}{\delta \alpha_b(\mathbf{y})} \text{Cov}_{\phi}^{-1} X^{\text{dat}} \right] (\mathbf{x}_j) \\
= \mathcal{B}(\mathbf{x}_j, \mathbf{y}) X_{,b}^{\text{WF}}(\mathbf{y}) - (\mathcal{B}_{\xi, b}^{\phi})(\mathbf{x}_j, \mathbf{y}) \bar{X}(\mathbf{y}). \quad (\text{A9})
\end{aligned}$$

The Fisher-like terms contract two such vectors, with an inverse covariance in the middle. When doing the contraction, the operator  $\mathcal{B}^{\dagger} \text{Cov}_{\phi}^{-1} \mathcal{B} \equiv K_{\phi}$  appears, and derivatives of  $\xi$  on the left-hand side get a minus sign because the first derivative is antisymmetric. The result is

$$\begin{aligned}
H_F^{ab}(\mathbf{x}, \mathbf{y}) &= \frac{1}{2} X_{\phi, a}^{\text{WF}}(\mathbf{x}) K_{\phi}(\mathbf{x}, \mathbf{y}) X_{\phi, b}^{\text{WF}}(\mathbf{y}) \\
&\quad - \frac{1}{2} \bar{X}(\mathbf{x})(\xi_{,a}^{\phi} K_{\phi} \xi_{,b}^{\phi})(\mathbf{x}, \mathbf{y}) \bar{X}(\mathbf{y}) \\
&\quad - \frac{1}{2} X_{\phi, a}^{\text{WF}}(\mathbf{x})(K_{\phi} \xi_{,b}^{\phi})(\mathbf{x}, \mathbf{y}) \bar{X}(\mathbf{y}) \\
&\quad + \frac{1}{2} \bar{X}(\mathbf{x})(\xi_{,a}^{\phi} K_{\phi})(\mathbf{x}, \mathbf{y}) X_{,b}^{\text{WF}}(\mathbf{y}). \quad (\text{A10})
\end{aligned}$$

This matrix is positive definite for any model. To apply this a vector  $v$ , we can do as follows:

$$\int d^2y H_F^{ab}(\mathbf{x}, \mathbf{y}) v_b(\mathbf{y}) = X_{\phi, a}^{\text{WF}}(\mathbf{x})(V - W)(\mathbf{x}) + \bar{X}(\mathbf{x})(\xi_{,a}(\mathbf{x})(V - W))(\mathbf{x}) \quad (\text{A11})$$

with

$$V = K_{\phi} X_{\phi, b}^{\text{WF}} v^b \quad W = K_{\phi} \xi_{, b}^{\phi} \bar{X} v^b. \quad (\text{A12})$$

There is thus only one set of maps to inverse filter, the difference between  $(\xi_{,b} \bar{X})(\mathbf{x}) v^b(\mathbf{x})$  and  $(\xi_{,b}(\bar{X} v^b))(\mathbf{x})$ . The term  $H_{\text{det}}$ , which comes from the second derivative of the covariance in Eq. (A3), remains. It is also a contraction of these types of maps. Explicitly,

$$H_{\text{det}}^{ab}(\mathbf{x}, \mathbf{y}) = \bar{X}(\mathbf{x}) \xi_{, ab}^{\phi}(\mathbf{x}, \mathbf{y}) \bar{X}(\mathbf{y}) - \delta^D(\mathbf{x} - \mathbf{y}) \bar{X}(\mathbf{x}) X_{, ab}^{\text{WF}}(\mathbf{x}), \quad (\text{A13})$$

which is easily applied to input vectors. The full  $H$  matrix, after rescaling by a simple isotropic approximation, is close enough to unity that its inverse can be applied to a vector with conjugate gradient inversion without much difficulty.

Finally, we reproduce for completeness the expression for the mean-response  $R_L$  used in several places in this

work, defined above in Eq. (A2). Again, we state the result  $R^{ab}(\mathbf{x}, \mathbf{y})$  for the two components of the displacement field, defined as

$$R^{ab}(\mathbf{x}, \mathbf{y}) \equiv \frac{\delta^2 \frac{1}{2} \ln \det \text{Cov}_{\phi}}{\delta \alpha_a(\mathbf{x}) \delta \alpha_b(\mathbf{x})} \Big|_{\phi=0}. \quad (\text{A14})$$

Since this is evaluated for vanishing deflection, the matrix is only a function of  $\mathbf{r} = \mathbf{x} - \mathbf{y}$ . The result follows straightforwardly from the covariance first and second variations, Eqs. (A5) and (A7):

$$\begin{aligned} -R^{ab}(\mathbf{r}) &= (\xi_{,a}^0 K_0)(\mathbf{r})(\xi_{,b}^0 K_0)(\mathbf{r}) \\ &+ K_0(\mathbf{r})(\xi_{,a}^0 K_0 \xi_{,b}^0)(\mathbf{r}) - K_0(\mathbf{r}) \xi_{,ab}^0(\mathbf{r}) \\ &+ \delta^D(\mathbf{r})(K_0 \xi_{,ab}^0)(\mathbf{r}). \end{aligned} \quad (\text{A15})$$

A contraction on Stokes indices is implicit in this equation. Contracting  $R^{ab}$  after Fourier transformation with  $L_a L_b$  gives the lensing potential response  $R_L$ . All terms are easily calculated with two-dimensional fast Fourier transform methods.

- 
- [1] Y. Akrami *et al.* (Planck Collaboration), Planck 2018 results. I. Overview and the cosmological legacy of Planck, [arXiv:1807.06205](https://arxiv.org/abs/1807.06205).
- [2] M. Kamionkowski, A. Kosowsky, and A. Stebbins, Statistics of cosmic microwave background polarization, *Phys. Rev. D* **55**, 7368 (1997).
- [3] M. Zaldarriaga and U. Seljak, An all sky analysis of polarization in the microwave background, *Phys. Rev. D* **55**, 1830 (1997).
- [4] U. Seljak and M. Zaldarriaga, Signature of gravity waves in polarization of the microwave background, *Phys. Rev. Lett.* **78**, 2054 (1997).
- [5] M. Kamionkowski and E. D. Kovetz, The quest for B modes from inflationary gravitational waves, *Annu. Rev. Astron. Astrophys.* **54**, 227 (2016).
- [6] K. N. Abazajian *et al.* (CMB-S4 Collaboration), CMB-S4 Science Book, First Edition, [arXiv:1610.02743](https://arxiv.org/abs/1610.02743).
- [7] M. Zaldarriaga and U. Seljak, Gravitational lensing effect on cosmic microwave background polarization, *Phys. Rev. D* **58**, 023003 (1998).
- [8] A. Lewis and A. Challinor, Weak gravitational lensing of the CMB, *Phys. Rep.* **429**, 1 (2006).
- [9] B. D. Sherwin and M. Schmittfull, Delensing the CMB with the cosmic infrared background, *Phys. Rev. D* **92**, 043005 (2015).
- [10] N. Aghanim *et al.* (Planck Collaboration), Planck 2018 results. VIII. Gravitational lensing, [arXiv:1807.06210](https://arxiv.org/abs/1807.06210).
- [11] A. Manzotti, Future cosmic microwave background delensing with galaxy surveys, *Phys. Rev. D* **97**, 043527 (2018).
- [12] B. Yu, J. C. Hill, and B. D. Sherwin, Multi-tracer CMB delensing maps from Planck and WISE data, *Phys. Rev. D* **96**, 123511 (2017).
- [13] T. Okamoto and W. Hu, CMB lensing reconstruction on the full sky, *Phys. Rev. D* **67**, 083002 (2003).
- [14] C. M. Hirata and U. Seljak, Reconstruction of lensing from the cosmic microwave background polarization, *Phys. Rev. D* **68**, 083002 (2003).
- [15] U. Seljak and C. M. Hirata, Gravitational lensing as a contaminant of the gravity wave signal in CMB, *Phys. Rev. D* **69**, 043005 (2004).
- [16] J. Carron and A. Lewis, Maximum *a posteriori* CMB lensing reconstruction, *Phys. Rev. D* **96**, 063510 (2017).
- [17] D. Hanson, A. Challinor, and A. Lewis, Weak lensing of the CMB, *Gen. Relativ. Gravit.* **42**, 2197 (2010).
- [18] P. J. E. Peebles, Tracing galaxy orbits back in time, *Astrophys. J.* **344**, L53 (1989).
- [19] P. A. R. Ade *et al.* (Planck Collaboration), Planck 2015 results. XIII. Cosmological parameters, *Astron. Astrophys.* **594**, A13 (2016).
- [20] A. Lewis, A. Challinor, and A. Lasenby, Efficient computation of CMB anisotropies in closed FRW models, *Astrophys. J.* **538**, 473 (2000).
- [21] P. A. R. Ade *et al.* (BICEP2 and Keck Array Collaborations), BICEP2/Keck Array x: Constraints on Primordial

- Gravitational Waves using Planck, WMAP, and New BICEP2/Keck Observations through the 2015 Season, *Phys. Rev. Lett.* **121**, 221301 (2018).
- [22] Y. Akrami *et al.* (Planck Collaboration), Planck 2018 results. X. Constraints on inflation, [arXiv:1807.06211](https://arxiv.org/abs/1807.06211).
- [23] A. Lewis, A. Hall, and A. Challinor, Emission-angle and polarization-rotation effects in the lensed CMB, *J. Cosmol. Astropart. Phys.* **08** (2017) 023.
- [24] A. Lewis, A. Challinor, and D. Hanson, The shape of the CMB lensing bispectrum, *J. Cosmol. Astropart. Phys.* **03** (2011) 018.
- [25] D. Beck, G. Fabbian, and J. Errard, Lensing reconstruction in post-born cosmic microwave background weak lensing, *Phys. Rev. D* **98**, 043512 (2018).
- [26] V. Boehm, B. D. Sherwin, J. Liu, J. C. Hill, M. Schmittfull, and T. Namikawa, On the effect of non-Gaussian lensing deflections on CMB lensing measurements, *Phys. Rev. D* **98**, 123510 (2018).
- [27] J. Liu, J. C. Hill, B. D. Sherwin, A. Petri, V. Boehm, and Z. Haiman, CMB lensing beyond the power spectrum: Cosmological constraints from the one-point probability distribution function and peak counts, *Phys. Rev. D* **94**, 103501 (2016).
- [28] S. Dorn and T. A. Enßlin, Stochastic determination of matrix determinants, *Phys. Rev. E* **92**, 013302 (2015).
- [29] E. J. Allen, J. Baglama, and S. K. Boyd, Numerical approximation of the product of the square root of a matrix with a vector, *Linear Algebra Appl.* **310**, 167 (2000).
- [30] P. H. Van Cittert, Zum einfluss der spaltbreite auf die intensitätsverteilung in spektrallinien. ii, *Z. Phys.* **69**, 298 (1931).
- [31] J.-L. Starck, E. Pantin, and F. Murtagh, Deconvolution in astronomy: A review, *Publ. Astron. Soc. Pac.* **114**, 1051 (2002).
- [32] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 77: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, New York, 1996), Vol. 1.
- [33] G. Efstathiou, Myths and truths concerning estimation of power spectra, *Mon. Not. R. Astron. Soc.* **349**, 603 (2004).
- [34] S. Hamimeche and A. Lewis, Likelihood analysis of CMB temperature and polarization power spectra, *Phys. Rev. D* **77**, 103013 (2008).
- [35] T. Namikawa and R. Nagata, Non-Gaussian structure of B-mode polarization after delensing, *J. Cosmol. Astropart. Phys.* **10** (2015) 004.
- [36] S. Smith, A. Challinor, and G. Rocha, What can be learned from the lensed cosmic microwave background b-mode polarization power spectrum?, *Phys. Rev. D* **73**, 023517 (2006).
- [37] W.-H. Teng, C.-L. Kuo, and J.-H. Protty Wu, Cosmic microwave background delensing revisited: Residual biases and a simple fix, [arXiv:1102.5729](https://arxiv.org/abs/1102.5729).
- [38] J. Carron, A. Lewis, and A. Challinor, Internal delensing of Planck CMB temperature and polarization, *J. Cosmol. Astropart. Phys.* **05** (2017) 035.
- [39] T. Namikawa, CMB internal delensing with general optimal estimator for higher-order correlations, *Phys. Rev. D* **95**, 103514 (2017).
- [40] A. Manzotti *et al.* (Herschel and SPT Collaborations), CMB polarization B-mode delensing with SPTpol and Herschel, *Astrophys. J.* **846**, 45 (2017).
- [41] A. Lewis, A. Challinor, and N. Turok, Analysis of CMB polarization on an incomplete sky, *Phys. Rev. D* **65**, 023505 (2001).
- [42] K. M. Smith and M. Zaldarriaga, A general solution to the E-B mixing problem, *Phys. Rev. D* **76**, 043001 (2007).
- [43] K. M. Smith, W. Hu, and M. Kaplinghat, Weak lensing of the CMB: Sampling errors on B-modes, *Phys. Rev. D* **70**, 043002 (2004).
- [44] A. Benoit-Levy, K. M. Smith, and W. Hu, Non-Gaussian structure of the lensed CMB power spectra covariance matrix, *Phys. Rev. D* **86**, 123008 (2012).
- [45] P. A. R. Ade *et al.* (Planck Collaboration), Planck 2015 results. XV. Gravitational lensing, *Astron. Astrophys.* **594**, A15 (2016).
- [46] B. D. Sherwin *et al.* (ACT Collaboration), The Atacama cosmology telescope: Two-season ACTPol lensing power spectrum, *Phys. Rev. D* **95**, 123529 (2017).