Non-Gaussian likelihood of weak lensing power spectra

Alex Hall[*] and Andy Taylor

Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill,
Edinburgh, EH9 3HJ, United Kingdom

The power spectrum of weak lensing fluctuations has a non-Gaussian distribution due to its quadratic nature. On small scales the central limit theorem acts to Gaussianize this distribution but non-Gaussianity in the signal due to gravitational collapse is increasing and the functional form of the likelihood is unclear. Analyses have traditionally assumed a Gaussian likelihood with nonlinearity incorporated into the covariance matrix; here we provide the theory underpinning this assumption. We calculate, for the first time, the leading-order correction to the distribution of angular power spectra from non-Gaussianity in the underlying signal and study the transition to Gaussianity. Our expressions are valid for an arbitrary number of correlated maps and correct the Wishart distribution in the presence of weak (but otherwise arbitrary) non-Gaussianity in the signal. Surprisingly, the resulting distribution is not equivalent to an Edgeworth expansion. The leading-order effect is to broaden the covariance matrix by the usual trispectrum term, with residual skewness sourced by the trispectrum and the square of the bispectrum. Using log-normal lensing maps we demonstrate that our likelihood is uniquely able to model both large and mildly nonlinear scales. We provide easy-to-compute statistics to quantify the size of the non-Gaussian corrections. We show that the full non-Gaussian likelihood can be accurately modeled as a Gaussian on small, nonlinear scales. On large angular scales nonlinearity in the lensing signal imparts a negligible correction to the likelihood, which takes the Wishart form in the full-sky case. Our formalism is equally applicable to any kind of projected field.

I. INTRODUCTION

In the coming decade several large astronomy projects aim to place percent-level constraints on cosmic acceleration, testing the Λ cold dark matter (ΛCDM) cosmological model and its extensions. These include the European Space Agency's Euclid mission,[1] the Vera C. Rubin Observatory,[2] and the Nancy Grace Roman Space Telescope.[3] These "stage-IV" facilities will each image roughly a billion galaxies, with redshifts derived from photometry calibrated against deep spectroscopic training samples.

Upcoming surveys will use these galaxies to measure the weak gravitational lensing signal (see Refs. [1,2] for recent reviews), which directly probes the underlying matter distribution and provides the opportunity to derive cosmological constraints from its statistical properties. Current lensing surveys are now achieving precision on certain ΛCDM cosmological parameter combinations at the level of a few percent and providing useful constraints on the

properties of dark energy [3–5]. The influence of weak lensing in constraining cosmological parameters is likely to increase with the advent of stage-IV surveys, if systematic errors can be kept under control.

It is now fairly common practice to combine measurements of the cosmic shear signal with maps of the angular positions of galaxies, sometimes referred to as "3×2 pt" analysis [6]. The relatively coarse precision of photometric redshifts means that binning galaxies into broad redshift bins is reasonably lossless, a technique called "tomography" [7–10]. Combining tomographic galaxy clustering and weak lensing are examples of the use of *projected* maps of dark matter (or discrete tracers thereof) to constrain cosmological models.

The baseline parameter constraints from current and upcoming surveys using projected fields will be derived from two-point statistics, such as correlation functions or power spectra, combined with a likelihood function. Accurate cosmological inference using this approach requires the likelihood to be specified accurately. The functional form of the likelihood has traditionally been assumed to be Gaussian in the measured summary statistics, the rationale being that the central limit theorem (CLT) drives the distribution of these statistics toward Gaussianity. However, it has long been known that the

---

distribution of weak lensing or galaxy clustering two-point statistics is in fact *not* Gaussian, for several reasons. First, two-point statistics are nonlinear (quadratic) functions of the data. If the data are Gaussian distributed, power spectra and correlation functions are expected to be non-Gaussian as a result, which can easily be seen by noting that the variance of the field (the correlation function at zero lag) cannot be negative. On the full sky the distribution of power spectra of Gaussian fields can be written down analytically (e.g., Ref. [11]), but in the presence of a survey mask (required to mitigate foreground contamination) the naive estimator (i.e., the pseudo-C_l estimator) has a distribution that is complicated to compute and must typically be approximated [12,13].

A second source of non-Gaussianity in the likelihood arises from the fact that the fields themselves are not Gaussian due to nonlinear gravitational collapse. This issue is inextricably tied to the efficacy of the CLT to Gaussianize the estimator, because both the number of degrees of freedom being compressed by the estimator (e.g., the number of galaxy pairs of a given angular separation) and the strength of the signal nonlinearity increase as the minimum angular scale is lowered. Non-Gaussianity of the signal (the assumption of Gaussian noise is typically safe) is usually incorporated by modifying the covariance matrix of the two-point statistic with a term derived from the connected four-point function (trispectrum) of the signal, while keeping the Gaussian functional form (e.g., [14–16]).

Additionally, non-Gaussianity in the likelihood can also arise due to the presence of non-Gaussian systematic residuals or when the covariance has been estimated from simulations [17] or combined with an analytic model [18].

Recognizing that the approximation of Gaussianity may contribute bias or imprecision to parameter constraints, several works have successfully measured non-Gaussianity in the distribution of two-point statistics from simulations. Reference [19] was an early work to find evidence for non-Gaussian moments of power spectrum estimates, but Refs. [20,21] represent the first attempts to incorporate non-Gaussianity into the likelihood of a weak lensing survey. The approach taken in Ref. [20] was to model the data vector (correlation functions in their case) as the sum of independent non-Gaussian components, using the fastICA independent component analysis method to construct the likelihood. Subsequent analytic work by Refs. [22,23] further established the non-Gaussianity of the weak lensing correlation functions. Reference [24] introduced diagnostic statistics to identify non-Gaussianity in cosmological likelihood functions, and evidence for non-Gaussianity in a weak lensing likelihood was found in Ref. [25]. Reference [26] studied non-Gaussianity in galaxy clustering power spectra, building on and extending the work of Ref. [20], and found non-negligible (albeit sub-1σ) biases in parameters. In the context of weak lensing correlation functions from

stage-IV surveys, Ref. [27] identified significant skewness and kurtosis in the shear correlation functions and used a principal component analysis to model the resulting non-Gaussian likelihood, but found that a Gaussian approximation was sufficient for ΛCDM parameters and a Rubin-like survey. This conclusion is supported by the work of Ref. [28] who used likelihood-free methods to establish the validity of a Gaussian approximation for stage-IV surveys and ΛCDM parameters. Recently, Ref. [29] demonstrated clear non-Gaussianity in the likelihood of weak lensing power spectra and presented a method to model this using normalizing flows. Reference [30] has also studied the impact of non-Gaussian likelihoods on parameter constraints, finding a negligible impact for the Dark Energy Survey data vector.

What these works have established is that, while non-Gaussianity is clearly present in the weak lensing likelihood, the impact of the Gaussian assumption on the posterior of parameters appears to be very modest, at least in ΛCDM models. However, there has been relatively little attempt to distinguish between the non-Gaussianity that should arise on survey scales due to the finite number of modes (that makes a Gamma distribution more accurate than a Gaussian) from the non-Gaussianity that could arise on small scales due to the CLT being inefficient at Gaussianizing the summary statistics in the face of increasing nonlinearity in the signal. The use of the power spectrum as the statistic of choice aids in distinguishing between these two effects, as well as allowing the modeling to be simplified due to the diagonal covariance that arises for Gaussian fields. As well as being unable to distinguish between various sources of non-Gaussianity, most of the works modeling non-Gaussian likelihoods have used numerical techniques or *ad hoc* approximations to capture the corrections, often with little physical justification. This is partly because the full non-Gaussian distribution of the shear or galaxy clustering signal cannot be written down.

In this work, we incorporate the effects of a non-Gaussian signal into the likelihood of the angular power spectrum and study how this impacts various posterior distributions of interest. To make analytic progress we assume that the signal non-Gaussianity is sufficiently weak that it can be treated perturbatively. While this is a constraining assumption, it will allow us to study in detail how the various sources of non-Gaussianity interact and make sense of the results of the numerical studies referenced above. Our approach is more in the spirit of Ref. [31], who studied similar kinds of effects in the power spectra of the weakly lensed cosmic microwave background, than the numerical works that have sought to model the full non-perturbative likelihood. We argue that an analytic approach building from the well-understood case of Gaussian fields and full-sky coverage offers a valuable way of understanding the Gaussianity , or otherwise, of the likelihood function. Having a tractable model that can be quickly

computed (and sampled from) also allows the effects of non-Gaussianity to be studied for any cosmological model, rather than being limited to ΛCDM.

This paper is structured as follows. In Sec. II we present the main expression used in this work, Eq. (22), and study its properties. In Sec. III we detail the simulations used to validate our likelihood and quantify the corrections to the posterior distributions. We present our conclusions in Sec. IV. In a series of appendices we present the derivation of the likelihood function, study the impact of alternative angular binning choices, and test the impact of the source redshift.

II. THE LIKELIHOOD OF ANGULAR POWER SPECTRUM ESTIMATES

Consider a survey that has made noisy maps $d^i(\hat{\mathbf{n}})$ of the large-scale scale structure across the full sky in p tomographic redshift bins labeled by the index i. Taking the spherical harmonic coefficients a^i_{lm} of these maps and arranging them into a vector of length p, the angular power spectra including all cross-bin pairs can be packaged into a $p \times p$ matrix $\hat{\mathbf{C}}_l$ given by

$$\nu\hat{\mathbf{C}}_l \equiv \sum_{m=-l}^{l} \mathbf{a}_{lm}\mathbf{a}_{lm}^\dagger, \tag{1}$$

where $\nu \equiv 2l + 1$ is the number of angular degrees of freedom at each l, and the spherical multipoles of the data consist of signal plus noise as $a^i_{lm} = s^i_{lm} + n^i_{lm}$. We will assume the maps are spin-0 (i.e., scalar), although this could easily be relaxed; in the case of shear data for example we can define the vector $\mathbf{a}_{lm}$ to include two entries per redshift bin corresponding to the shear E and B modes. Similarly, projected galaxy number counts may be included in the data vector for a full 3×2 pt analysis. The hatted power spectrum in Eq. (1) denotes a measured, noisy quantity. We will assume that the maps are purely real, and hence all multipoles obey

$$\mathbf{a}_{l-m} = (-1)^m \mathbf{a}_{lm}^*. \tag{2}$$

Our goal is to compute the sampling distribution of these power spectrum estimates $p(\{\hat{\mathbf{C}}_l\})$ with braces indicating the full set of angular multipoles l, given knowledge of the underlying statistical properties of the noise and the signal. The sampling distribution considered as a function of these properties is the likelihood function. The power spectra are quadratic in the data and so are expected to obey Gaussian statistics only when the central limit theorem is effective. The other source of non-Gaussianity is from the signal itself, which when measured in the late Universe on small scales has higher-order moments due to nonlinear structure formation.

As discussed in Appendix A we cannot proceed by performing an Edgeworth expansion of the power spectrum distribution around its zero-order Wishart form, so instead we develop a new approach and begin by considering the statistics of the underlying data. Assuming the signal and noise are independent, the joint distribution of the power spectra (multiplied by ν) is

$$p(\{\nu\hat{\mathbf{C}}_l\}) = \int d\{\mathbf{s}_{lm}\} \int d\{\mathbf{n}_{lm}\} p_s(\{\mathbf{s}_{lm}\}) p_n(\{\mathbf{n}_{lm}\})$$
$$\times \left[\prod_{l=l_{\min}}^{l_{\max}} \delta_D\left(\nu\hat{\mathbf{C}}_l - \sum_m \mathbf{a}_{lm}\mathbf{a}_{lm}^\dagger\right)\right], \tag{3}$$

where the probability density $p(\{\nu\hat{\mathbf{C}}_l\})$ is defined with a measure over *real* and *symmetric* matrices, which have $n \equiv p(p+1)/2$ independent elements. We will expand the Dirac delta in terms of real and symmetric p-dimensional matrices $\mathbf{J}_l$ as

$$\delta_D\left(\nu\hat{\mathbf{C}}_l - \sum_m \mathbf{a}_{lm}\mathbf{a}_{lm}^\dagger\right) = 2^{\frac{p(p-1)}{2}} \int \frac{d\mathbf{J}_l}{(2\pi)^n} e^{i\mathrm{Tr}[\mathbf{J}_l(\nu\hat{\mathbf{C}}_l - \sum_m \mathbf{a}_{lm}\mathbf{a}_{lm}^\dagger)]}, \tag{4}$$

where we integrate over all real symmetric matrices and the volume element is

$$d\mathbf{J}_l = dJ_l^{1,1} dJ_l^{1,2} \ldots dJ_l^{1,p} dJ_l^{2,3} \ldots dJ_l^{p,p}. \tag{5}$$

Note that the $2^{\frac{p(p-1)}{2}}$ factor is needed since the off-diagonal elements of $\mathbf{J}_l$ are double counted in the summation in the exponential, and hence we need to redefine each of the $p(p-1)/2$ off-diagonal terms in the integration with a factor of 2.

We expand the signal distribution in Fourier modes as

$$p(\{\mathbf{s}_{lm}\}) = \int \frac{d\{\mathbf{k}_{lm}\}}{(2\pi)^{pN}} \phi_s(\{\mathbf{k}_{lm}\}) e^{i\sum_{lm} \mathbf{k}_{lm}^\dagger \mathbf{s}_{lm}}, \tag{6}$$

where $N \equiv \sum_{l=l_{\min}}^{l_{\max}} 2l + 1$ is the total number of modes per tomographic bin, and ϕ_s is the characteristic function of the signal.[4] We can, and will, define the "wave numbers" $\mathbf{k}_{lm}$ to obey reality conditions as in Eq. (2). The corresponding characteristic function of the noise is ϕ_n.

Inserting Eqs. (4) and (6) into Eq. (3), changing integration variables, and doing an integral over $\mathbf{a}_{lm}$ assuming $\mathbf{J}_l$ is nonsingular[5] gives

---

[4]We will define characteristic functions this way throughout, but we caution the reader that other authors sometimes define it with a complex conjugate.

[5]The set of singular $\mathbf{J}_l$ has zero measure and, as the integrand is finite for this set, the contribution from singular modes is vanishingly small in the limit.

$$p(\{\nu\hat{\mathbf{C}}_l\}) = 2^{\frac{\lambda p(p-1)}{2}} \int \frac{\mathrm{d}\{\mathbf{J}_l\}}{(2\pi)^{n\lambda}} e^{i\sum_l \mathrm{Tr}(\nu\mathbf{J}_l\hat{\mathbf{C}}_l)} \prod_{l=l_{\min}}^{l_{\max}} |2i\mathbf{J}_l|^{-\frac{\nu}{2}} \int \frac{\mathrm{d}\{\mathbf{k}_{lm}\}}{(2\pi)^{\frac{pN}{2}}} \phi_s(\{\mathbf{k}_{lm}\}) e^{-\frac{1}{2}\sum_{lm}\mathbf{k}_{lm}^{\dagger}(2i\mathbf{J}_l)^{-1}\mathbf{k}_{lm}} \phi_n(\{\mathbf{k}_{lm}\}), \qquad (7)$$

where $\lambda \equiv l_{\max} - l_{\min} + 1$.

Now, we define the characteristic function of $\{\nu\hat{\mathbf{C}}_l\}$ over its independent elements as

$$\phi_{\{\nu\hat{\mathbf{C}}_l\}}(\{\mathbf{J}_l\}) = \int \mathrm{d}\{\nu\hat{\mathbf{C}}_l\} e^{-i\sum_l \mathrm{Tr}(\nu\mathbf{J}_l\hat{\mathbf{C}}_l)} p(\{\nu\hat{\mathbf{C}}_l\}). \qquad (8)$$

This gives, finally,

$$\phi_{\{\nu\hat{\mathbf{C}}_l\}}(\{\mathbf{J}_l\}) = \left(\prod_{l=l_{\min}}^{l_{\max}} |2i\mathbf{J}_l|^{-\frac{\nu}{2}}\right) \int \frac{\mathrm{d}\{\mathbf{k}_{lm}\}}{(2\pi)^{\frac{pN}{2}}} \phi_s(\{\mathbf{k}_{lm}\}) e^{-\frac{1}{2}\sum_{lm}\mathbf{k}_{lm}^{\dagger}(2i\mathbf{J}_l)^{-1}\mathbf{k}_{lm}} \phi_n(\{\mathbf{k}_{lm}\}), \qquad (9)$$

i.e., a mode coupling integral between the characteristic functions of the signal, the noise, and a Gaussian in the wave number of the field $\mathbf{k}_{lm}$ having covariance $2i\mathbf{J}$, where $\mathbf{J}$ is the wave number of the power spectra.

A. The known case of a Gaussian signal and Gaussian noise

So far, we have not assumed anything about the statistics of the signal or the noise, only that they are independent. In the case of Gaussian signal and noise, the characteristic functions ϕ_s and ϕ_n also take a Gaussian form, and the integrals in Eq. (9) all decouple. Each integration has a Gaussian form that can be done analytically, with the result being a product of Wishart characteristic functions

$$\phi_{\{\nu\hat{\mathbf{C}}_l\}}(\{\mathbf{J}_l\}) = \prod_{l=l_{\min}}^{l_{\max}} |\mathbf{I} + 2i\mathbf{J}_l\mathbf{C}_l|^{-\frac{\nu}{2}}, \qquad (10)$$

where $\mathbf{C}_l$ is the expected total (signal plus noise) covariance matrix of the map. The inverse Fourier transform of this can be computed analytically (see Appendix B) and gives the well-known distribution of the power spectra as a product of Wishart distributions (e.g., Ref. [11])

$$p(\{\hat{\mathbf{C}}_l\}) = \prod_{l=l_{\min}}^{l_{\max}} \frac{|\hat{\mathbf{C}}_l|^{(\nu-p-1)/2}}{2^{\frac{p\nu}{2}}\Gamma_p(\frac{\nu}{2})|\mathbf{C}_l/\nu|^{\frac{\nu}{2}}} e^{-\frac{\nu}{2}\mathrm{Tr}(\mathbf{C}_l^{-1}\hat{\mathbf{C}}_l)}$$

$$\equiv \prod_{l=l_{\min}}^{l_{\max}} W_p(\hat{\mathbf{C}}_l; \mathbf{C}_l/\nu, \nu), \qquad (11)$$

where W_p is the p-dimensional Wishart density and Γ_p is the multivariate Gamma function.

Each $\hat{\mathbf{C}}_l$ estimate thus has a Wishart distribution with scale matrix $\mathbf{C}_l/\nu$ and degrees of freedom ν with $\nu > p - 1$.

As expected, the mean of each $\hat{\mathbf{C}}_l$ is $\mathbf{C}_l$ (the estimator is unbiased), and the variance of a given map power spectrum is $2C_l^2/\nu$. The latter also follows immediately from squaring the definition (1) and using Wick's theorem.

When $p = 1$ this distribution is equivalent to a Gamma distribution. The Gamma distribution has nonzero skewness, kurtosis, and higher-order cumulants. For example, the dimensionless skewness is equal to $S = 2\sqrt{2/\nu}$, and the dimensionless excess kurtosis is $\kappa = 12/\nu$. It is easy to see that higher-order cumulants come with successive factors of $\nu^{-1/2}$; this is how the central limit theorem manifests itself when $\nu \gg 1$, since these higher-order cumulants are all formally zero for a Gaussian field. From Eq. (1) we see that ν counts the azimuthal modes contributing to a given wave number, such that the power spectrum estimator is the average of an increasingly large number of independent variables as ν increases.

The convergence of any probability distribution to a Gaussian can be established formally by showing that its characteristic function tends pointwise to that of a Gaussian (Lévy's continuity theorem, e.g., Ref. [32]). Inspection of Eq. (10) makes clear how this happens, where the convergence to a Gaussian can be established by Taylor expanding around $\nu = \infty$. One can also establish Gaussianity from the density directly. In the $p = 1$ case one first has to change variables to $\hat{C}_l = C_l(1 + n\sqrt{2/\nu})$ then take the limit $\nu \to \infty$ at fixed n, using Stirling's approximation to deal with the Gamma function in the denominator. Whichever way one chooses to go, the resulting limiting distribution is a product of Gaussians, each with mean C_l and variance $2C_l^2/\nu$.

We close this section by noting that in general one has to be careful about assuming a Gaussian sampling distribution for power spectrum estimates, as this does not necessarily guarantee a Gaussian *posterior* for parameters at the same

level of accuracy. For example, if one has an informative and non-Gaussian prior on a parameter, the amount of constraining data required for the likelihood to dominate the posterior can be well above that needed to Gaussianize the likelihood. Second, and more important, posterior standard errors on parameters can shrink as fast as non-Gaussian contributions to the likelihood, meaning relative corrections from non-Gaussianity can remain important for posteriors even on small angular scales. In general these effects are most robustly studied with simulations. Reference [12] provides a thorough discussion of residual non-Gaussian effects from failure of the central limit theorem for Gaussian fields in the cosmic microwave background context, finding that sufficiently broad multipole bins can reduce the error from approximating the likelihood as Gaussian to acceptable levels. We will return to this question in the cosmic shear context in Sec. III.

B. Corrections to the tomographic shear power distribution from a non-Gaussian signal

In the previous section we derived the distribution of angular power spectrum estimates assuming both signal and noise were Gaussian. In this section we will consider the more realistic scenario of a non-Gaussian signal, but will keep the assumption of Gaussian noise.

In the Gaussian case, the mode coupling integral in Eq. (9) can be performed analytically due to the Gaussian form of the signal characteristic function ϕ_s. The fully general non-Gaussian characteristic function of the signal has no analytic expression that can be written down, so to make progress we must restrict to *weak* non-Gaussianity, in the sense that the higher-order cumulants of the signal are progressively suppressed. This is the Edgeworth expansion [33–38], and truncating at leading order gives

$$
\begin{aligned}
&\phi_s(\{\mathbf{k}_{lm}\}) \\
&= \left(1 - i\frac{1}{6}\kappa_{\underline{lm}}^{ijk} k_{l_1m_1}^i k_{l_2m_2}^j k_{l_3m_3}^k + \frac{1}{24}\kappa_{\underline{lm}}^{ijkm} k_{l_1m_1}^i k_{l_2m_2}^j k_{l_3m_3}^k k_{l_4m_4}^m \right. \\
&\left. - \frac{1}{72}\kappa_{\underline{lm}}^{ijk}\kappa_{\underline{l'm'}}^{lmn} k_{l_1m_1}^i k_{l_2m_2}^j k_{l_3m_3}^k k_{l'_1m'_1}^l k_{l'_2m'_2}^m k_{l'_3m'_3}^n \right)\phi_s^G(\{\mathbf{k}_{lm}\}),
\end{aligned}
$$

$$(12)$$

where all repeated indices are summed over and ϕ_s^G is the Gaussian characteristic function. The compact notation $\underline{lm}$ denotes l_1m_1, l_2m_2, l_3m_3 when attached to a three-point cumulant, and $l_1m_1, l_2m_2, l_3m_3, l_4m_4$ when attached to a four-point cumulant. These are defined as

$$
\kappa_{\underline{lm}}^{ijk} \equiv \langle s_{l_1m_1}^i s_{l_2m_2}^j s_{l_3m_3}^k \rangle, \tag{13}
$$

$$
\kappa_{\underline{lm}}^{ijkm} \equiv \langle s_{l_1m_1}^i s_{l_2m_2}^j s_{l_3m_3}^k s_{l_4m_4}^m \rangle_c, \tag{14}
$$

where the subscript c denotes the fully connected part.

In Appendix A we discuss how the signal cumulants are related to those of the power spectrum estimates themselves. The covariance and leading-order dimensionless skewness of $\hat{C}_l$ in the $p = 1$ case are given by

$$
\langle \Delta\hat{C}_l \Delta\hat{C}_{l'} \rangle = \frac{2}{\nu} C_l^2 \delta_{ll'} + T_{ll'}, \tag{15}
$$

$$
\frac{\langle \Delta\hat{C}_l^3 \rangle}{\langle \Delta C_l^2 \rangle^{3/2}} \approx \sqrt{\frac{8}{\nu}} + \sqrt{2\nu^3}\frac{\tilde{B}_{lll}}{C_l^3} + \frac{3\sqrt{2\nu}}{2}\frac{T_{ll}}{C_l^2}. \tag{16}
$$

The skewness has a term coming from the Gaussian part of the signal ($\sqrt{8/\nu}$), plus a term proportional to the squared bispectrum of the signal $\tilde{B}_{lll}$, defined later in Eq. (21), and a term proportional to its trispectrum $T_{ll'}$. These latter two quantities are exactly those present in Eq. (12) and are also the leading-order non-Gaussian corrections to the dimensionless kurtosis of $\hat{C}_l$. Indeed, as discussed in Appendix A, *all* of the dimensionless cumulants of $\hat{C}_l$ have leading-order corrections from the trispectrum and squared bispectrum. Since these two quantities are of the same perturbative order this means that the cumulants of $\hat{C}_l$ do not get progressively smaller as their order increases, and an Edgeworth expansion of the distribution of $\hat{C}_l$ around a Wishart will not work. This justifies our approach of expanding the underlying signal as an Edgeworth expansion and then propagating to get the distribution of the power spectrum.

Substituting Eq. (12) into Eq. (9), we immediately see that the leading-order bispectrum term vanishes due to this term carrying an odd power of k_{lm}. The leading-order contribution to the distribution is therefore from terms of order the trispectrum or bispectrum squared, whose computation formally requires going to third order in perturbation theory. We will therefore truncate the expansion in Eq. (12) at the order shown. Since the probability density is linear in the characteristic function, we can simply add the trispectrum correction to that coming from bispectrum squared. We present the derivation of the characteristic function and the probability density of the power spectrum in Appendix B, with the final result given in Eqs. (B35) and (B56). We will simply quote the results here.

1. Non-Gaussian power distribution from a signal with nonzero trispectrum

In the $p = 1$ case the distribution of $\hat{C}_l$ corrected by a signal trispectrum given in Eq. (B35) reduces to

$$p(\{\hat{C}_l\}) = \left\{ \prod_{l=l_{\min}}^{l_{\max}} \Gamma[\hat{C}_l; \nu/2, \nu/(2C_l)] \right\} \left[1 + \frac{1}{8} \sum_{l_1, l_2} \nu_1 \nu_2 \frac{T_{l_1 l_2}}{C_{l_1} C_{l_2}} \left(\frac{\Delta \hat{C}_{l_1} \Delta \hat{C}_{l_2}}{C_{l_1} C_{l_2}} - \frac{2}{\nu_1 + 2} \delta_{l_1 l_2} \frac{\hat{C}_{l_1}^2}{C_{l_1}^2} \right) \right], \tag{17}$$

where $T_{ll'}$ is the m-averaged version of the signal trispectrum that contributes to the power spectrum covariance, see Eq. (15). One can verify that the distribution (17) is correctly normalized, has a mean given by $\langle \hat{C}_l \rangle = C_l$ (unchanged from the Gamma case), and a covariance matrix given by $\langle \Delta \hat{C}_{l_1} \Delta \hat{C}_{l_2} \rangle = 2 \delta_{l_1 l_2} C_{l_1}^2 / \nu_1 + T_{l_1 l_2}$, as expected.

Despite possessing the correct normalization, mean, and covariance, Eq. (17) is not a valid probability density since it is not everywhere positive. This arises for sufficiently large $\hat{C}_l$. This is an ubiquitous feature of densities based on the Edgeworth expansion and can be remedied by writing the correction as an exponential, i.e., $1 + x \approx e^x$. This is valid since the correction is assumed to perturbatively small.

The first correcting term in Eq. (17), proportional to $\Delta \hat{C}_{l_1} \Delta \hat{C}_{l_2}$, can be written in a suggestive way. It is the leading-order term in the expansion of

$$-\frac{1}{2} \sum_{l_1, l_2} \Delta \hat{C}_{l_1} \left(\frac{2C_{l_1}^2}{\nu_1} \delta_{l_1 l_2} + T_{l_1 l_2} \right)^{-1} \Delta \hat{C}_{l_2}, \tag{18}$$

where the notation $(A_{l_1 l_2})^{-1}$ refers to the l_1, l_2 element of the inverse of the matrix with elements $A_{l_1 l_2}$, as should be clear from context. Equation (18) is a χ^2-like functional having the inverse of the *total* covariance matrix. Once the Gaussian part of this expansion is "available" from the zero-order Gamma term when the $\nu \to \infty$ limit is taken, we will be left a Gaussian possessing the correct inverse covariance matrix in its exponent. Similarly, the second term in Eq. (17) is the leading-order correction to the determinant prefactor of a Gaussian due to the non-Gaussian covariance $T_{l_1 l_2}$. Thus, in the $\nu \to \infty$ limit, our distribution is the leading-order correction to a Gaussian due to $T_{l_1 l_2}$. This can also be seen directly from the characteristic function, Eq. (B25).

Considered as a function of the model power spectrum C_l and the trispectrum $T_{l_1 l_2}$, Eq. (17) gives the likelihood function for fixed data $\hat{C}_l$. In the next section we will compute the correction to the likelihood numerically, but there are a few analytic results we can derive from this likelihood. Most significantly, we can compute its expected curvature around a fiducial power spectrum $C_{f,l}$ and non-Gaussian covariance $T_{f,l_1 l_2}$, i.e., the Fisher matrix. Assuming a model with parameters θ_α, this is given by

$$\begin{aligned} F_{\alpha\beta} &= \frac{1}{2} \sum_{l=l_{\min}}^{l_{\max}} \nu \frac{1}{C_{f,l}^2} \frac{\partial C_l}{\partial \theta_\alpha} \frac{\partial C_l}{\partial \theta_\beta} \\ &\quad - \frac{1}{4} \sum_{l_1, l_2 = l_{\min}}^{l_{\max}} \nu_1 \nu_2 \frac{T_{f,l_1 l_2}}{C_{f,l_1}^2 C_{f,l_2}^2} \frac{\partial C_{l_1}}{\partial \theta_\alpha} \frac{\partial C_{l_2}}{\partial \theta_\beta} \\ &\approx \sum_{l_1, l_2 = l_{\min}}^{l_{\max}} \frac{\partial C_{l_1}}{\partial \theta_\alpha} \left(\frac{2C_{f,l_1}^2}{\nu} \delta_{l_1 l_2} + T_{f,l_1 l_2} \right)^{-1} \frac{\partial C_{l_2}}{\partial \theta_\beta}, \end{aligned} \tag{19}$$

where the derivatives are evaluated at the fiducial model. Equation (19) is the usual expression for the power spectrum Fisher matrix but with the proper non-Gaussian covariance matrix. This matches the expression for the Fisher matrix derived from a Gaussian likelihood with the covariance fixed to a fiducial model and including the non-Gaussian term and tells us that to recover the full-sky Fisher matrix from a likelihood approximated by its Gaussian limit we have to *fix* the Gaussian covariance to a fiducial model. This is the main result of Ref. [39], and here we recover it in the presence of signal non-Gaussianity. It is interesting to note that in principle one can allow the non-Gaussian part of the covariance matrix $T_{l_1 l_2}$ to depend on parameters in a Gaussian likelihood even when fixing the Gaussian part, as required to keep the approximation consistent. In our perturbative framework the additional information contributed by $T_{l_1 l_2}(\theta)$ is higher order, but whether this is consistent with the likelihood approximation for general $T_{l_1 l_2}$ is still an open question and covered by neither our analysis nor that of Ref. [39].

We note finally that we do not recover the field-level Fisher matrix from our power spectrum Fisher matrix, which now contains new terms due to non-Gaussianity. This is an expression of the fact that the power spectrum does not now contain all the information.

2. Non-Gaussian power distribution from a signal with nonzero bispectrum

We have argued that a consistent expansion of distribution of the power spectrum requires including terms of order the signal bispectrum squared. In the $p = 1$ case, the correction given in Eq. (B56) reduces to

$$\begin{aligned} p(\{\hat{C}_l\}) &= \left\{ \prod_{l=l_{\min}}^{l_{\max}} \Gamma[\hat{C}_l; \nu/2, \nu/(2C_l)] \right\} \\ &\times \left\{ 1 + \frac{1}{12} \sum_{l_1, l_2, l_3} \nu_1 \nu_2 \nu_3 \frac{\tilde{B}_{l_1 l_2 l_3}}{C_{l_1}^2 C_{l_2}^2 C_{l_3}^2} \left[\Delta \hat{C}_{l_1} \Delta \hat{C}_{l_2} \Delta \hat{C}_{l_3} - [3] \frac{2 \delta_{l_1 l_2}}{(\nu_1 + 2)} \hat{C}_{l_1}^2 \Delta \hat{C}_{l_3} + \frac{16 \delta_{l_1 l_2} \delta_{l_2 l_3}}{(\nu_1 + 4)(\nu_1 + 2)} \hat{C}_{l_1}^3 \right] \right\}, \end{aligned} \tag{20}$$

where $\tilde{B}$ is proportional to the square of the reduced bispectrum $b_{l_1 l_2 l_3}$ of the signal and is given by

$$\tilde{B}_{l_1 l_2 l_3} = \frac{1}{4\pi} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}^2 b_{l_1 l_2 l_3}^2, \qquad (21)$$

with the quantity in parentheses a Wigner $3j$ symbol. The notation $[n]$ in Eq. (20) refers to the n terms that follow from symmetry considerations. It is straightforward to verify that the correction leaves the normalization, mean, and covariance matrix unchanged, but corrects the three-point function by a term equal to $4\tilde{B}_{l,l',l''}$, as expected from Eq. (A2).

One can verify that the Fisher matrix is unaffected by the squared-bispectrum correction at leading order. In other words, the effect of $\tilde{B}_{l_1 l_2 l_3}$ is to modify the skewness of the likelihood, while leaving its mean and variance approximately unaffected.

As claimed in Appendix A 1, the correction to the density is $\mathcal{O}(\tilde{B}/C_l^3)$, the same order as the trispectrum correction. This justifies our decision to include terms of order the bispectrum squared.

In summary, the leading-order correction to the distribution of power spectrum estimates from signal non-Gaussianity is

$$p(\{\hat{C}_l\}|\{C_l, T_{l,l'}, \tilde{B}_{l,l',l''}\}) = \left\{ \prod_{l=l_{\min}}^{l_{\max}} \Gamma[\hat{C}_l; \nu/2, \nu/(2C_l)] \right\} \left\{ 1 + \frac{1}{8} \sum_{l_1,l_2} \nu_1 \nu_2 \frac{T_{l_1 l_2}}{C_{l_1} C_{l_2}} \left(\frac{\Delta \hat{C}_{l_1} \Delta \hat{C}_{l_2}}{C_{l_1} C_{l_2}} - \frac{2}{\nu_1+2} \delta_{l_1 l_2} \frac{\hat{C}_{l_1}^2}{C_{l_1}^2} \right) \right.$$
$$\left. + \frac{1}{12} \sum_{l_1,l_2,l_3} \nu_1 \nu_2 \nu_3 \frac{\tilde{B}_{l_1 l_2 l_3}}{C_{l_1}^2 C_{l_2}^2 C_{l_3}^2} \left[\Delta \hat{C}_{l_1} \Delta \hat{C}_{l_2} \Delta \hat{C}_{l_3} - [3] \frac{2\delta_{l_1 l_2}}{(\nu_1+2)} \hat{C}_{l_1}^2 \Delta \hat{C}_{l_3} + \frac{16 \delta_{l_1 l_2} \delta_{l_2 l_3}}{(\nu_1+4)(\nu_1+2)} \hat{C}_{l_1}^3 \right] \right\}.$$
$$(22)$$

Equation (22) is the first major result of this work. It is the leading-order likelihood function for the power spectrum that incorporates a non-Gaussian signal. Unlike the commonly assumed Gaussian approximation, this distribution applies equally well to the largest angular scales. This distribution has the correct normalization, mean, covariance, and three-point function. In Appendix B we provide the analogous expression in the case of multiple redshift bins, in which case the distribution is a perturbation around a Wishart distribution for positive-definite matrices. A practical application of the distribution may require ensuring positivity by taking the logarithm, but in practice we have not found this to be necessary. In Sec. III we present numerical results for the transition of the power spectrum distribution from Wishart to post-Wishart [our Eq. (22)] and finally to Gaussian as the number of degrees of freedom increases.

We close this section by noting that, while we have restricted ourselves to spin-0 fields (such as lensing convergence or projected galaxy number counts) for simplicity, our formalism may be easily generalized to the spin-2 case (i.e., weak lensing shear), as mentioned at the beginning of Sec. II. Since we work in harmonic space on the full sky, one can simply work with the E and B multipoles derived in the usual way from the shear field. This augments the data vector to have length $2 \times N_z$, but all the matrix expressions presented here can be used after making this modification. Redshift bin indices must now be

interpreted as also running over E and B indices, and the appropriate cross-bispectra and trispectra must be used (many of which will be zero due to parity invariance).

On the cut sky our expression must be modified to account for the mask, which we have not attempted to include. This induces additional mode mixing (as well as the usual E/B ambiguity), which could possibly be included using the techniques of either Ref. [12] or Ref. [13]. We defer this to a future work.

C. The transition to Gaussianity

A key goal of this paper is to understand and quantify how efficiently the central limit theorem Gaussianizes the distribution of $\hat{C}_l$ in the presence of signal non-Gaussianity. One could well imagine a situation for example where the strength of non-Gaussianity is increasing with $l_{\max}$ faster than the CLT is able to Gaussianize, in which the Gaussian limit may *never* be achieved. In this section we use our new distribution to study how the CLT operates in detail. We specialize to the single redshift bin case, $p = 1$.

Convergence of a probability density to a limiting distribution can be established by studying the pointwise convergence of its characteristic function. We have already computed this, and we will additionally change variables from $\hat{C}_l$ to the "standardized" variable $\hat{Z}_l \equiv \Delta \hat{C}_l / \sqrt{2 C_l^2 / \nu}$. The characteristic function of the set of $\hat{Z}_l$ variables is

$$\phi_{\{\hat{Z}_l\}}(\{J_l\}) = \left[\prod_{l=l_{\min}}^{l_{\max}} e^{iJ_l\sqrt{\nu/2}}(1 + iJ_l\sqrt{2/\nu})^{-\nu/2}\right]\left[1 - \frac{1}{2}\sum_{l_1,l_2}\frac{\varepsilon_{l_1l_2}J_{l_1}J_{l_2}}{(1 + iJ_{l_1}\sqrt{2/\nu_1})(1 + iJ_{l_2}\sqrt{2/\nu_2})}\right.$$

$$\left.+ \frac{i}{6}\sum_{l_1,l_2,l_3}\frac{\tilde{\varepsilon}_{l_1l_2l_3}J_{l_1}J_{l_2}J_{l_3}}{(1 + iJ_{l_1}\sqrt{2/\nu_1})(1 + iJ_{l_2}\sqrt{2/\nu_2})(1 + iJ_{l_3}\sqrt{2/\nu_3})}\right], \tag{23}$$

where we have defined the quantities

$$\varepsilon_{l_1l_2} \equiv \frac{T_{l_1l_2}}{\sqrt{2C_{l_1}^2/\nu_1}\sqrt{2C_{l_2}^2/\nu_2}}, \tag{24}$$

$$\tilde{\varepsilon}_{l_1l_2l_3} \equiv \frac{4\tilde{B}_{l_1l_2l_3}}{\sqrt{2C_{l_1}^2/\nu_1}\sqrt{2C_{l_2}^2/\nu_2}\sqrt{2C_{l_3}^2/\nu_3}}. \tag{25}$$

The quantity $\varepsilon_{l_1l_2}$ has the simple interpretation that it is the relative correction to the power spectrum covariance matrix from non-Gaussianity. Its diagonal elements can also be interpreted as the correction to the "effective degrees of freedom" through $\nu_{\rm eff} = \nu/(1 + \varepsilon_{ll})$ such that the variance of the power spectrum is $2C_l^2/\nu_{\rm eff}$, although simply making this correction in the likelihood is not sufficient as it does not capture, among other things, mode coupling.

Consider now taking every $\nu \gg 1$, i.e., consider the limit $l_{\min} \gg 1$. The zero-order prefactor in Eq. (23) converges to a Gaussian, with the "linear" part of the exponent getting canceled by the complex exponential already present. Likewise, all the denominators in the non-Gaussian correction terms tend to unity at the same rate, leaving a quadratic correction from the trispectrum and a cubic correction from the bispectrum. In this limit, the correction is of the Edgeworth form. At the order to which we work, we may just as well replace the quadratic term by a Gaussian using $e^{-\lambda^2} \approx 1 - \lambda^2$, in which case the trispectrum term is absorbed into the covariance matrix of the Gaussian prefactor. This leaves the correction in the $l_{\min} \gg 1$ limit as

$$\phi_{\{\hat{Z}_l\}}(\{J_l\}) \approx \exp\left[-\frac{1}{2}\sum_{l_1l_2}J_{l_1}(\delta_{l_1l_2} + \varepsilon_{l_1l_2})J_{l_2}\right]$$

$$\times\left(1 + \frac{i}{6}\sum_{l_1,l_2,l_3}\tilde{\varepsilon}_{l_1l_2l_3}J_{l_1}J_{l_2}J_{l_3}\right). \tag{26}$$

For this to be valid, we clearly need $\sum_{l_1l_2}J_{l_1}J_{l_2}\varepsilon_{l_1l_2} \ll 1$. A necessary condition for this is that every element of ε is "small" in a sense that depends on the quantity of interest. For example, if one was only interested in the marginal distribution of a single $\hat{C}_l$, it would be sufficient to demand that only a single $\varepsilon_{ll} \ll 1$.

The dominant effect of the trispectrum in this limit is clearly a change to the covariance matrix. Residual "post-Gaussian" effects come from the squared bispectrum through the quantity $\tilde{\varepsilon}_{l_1l_2l_3}$ affecting the three-point function of the power spectrum. These are negligible when $\tilde{\varepsilon}_{l_1l_2l_3} \ll 1$ at all multipoles.

One guaranteed way to make both ε and $\tilde{\varepsilon}$ small is to increase C_l at fixed $T_{l_1l_2}$ or fixed $\tilde{B}_{l_1l_2l_3}$. This may be achieved by simply increasing the (assumed Gaussian) noise power, since C_l here is the total power spectrum. This is equivalent to saying that the noise "Gaussianizes" the distribution. One must be careful however to check this Gaussianization also operates on not just the likelihood but the posterior, as we will do shortly.

The quantity $\varepsilon_{l_1l_2}$ is known to be significant on nonlinear scales and is routinely measured in simulations (usually in a binned form, see Fig. 1 for a related quantity). In contrast, the size of $\tilde{\varepsilon}_{l_1l_2l_3}$ is less obvious. We can make an order-of-magnitude estimate of its size using the asymptotic limit of the Wigner $3j$ symbol [40,41]

$$\lim_{\nu_1,\nu_2,\nu_3\to\infty}\begin{pmatrix}l_1 & l_2 & l_3\\0 & 0 & 0\end{pmatrix}^2 = \frac{1}{2\pi A(L_1,L_2,L_3)}, \tag{27}$$

where $L \equiv l + 1/2$ and $A(L_1,L_2,L_3)$ is the area of the triangle having side lengths L_1, L_2, L_3. The squared $3j$
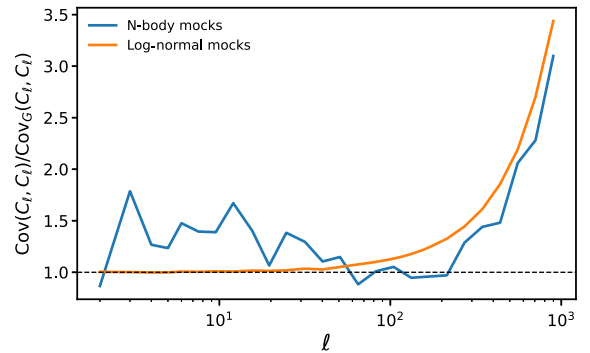

FIG. 1. Ratio of the diagonal elements of the C_l covariance matrix to that of Gaussian fields for our log-normal mocks (orange) and the ray-traced N-body simulations of Ref. [43] (blue). We have binned the power spectra with $\Delta\log_{10}l = 0.1$. Reference [43] speculates that the excess variance at large angular scales in the N-body mocks is due to the finite thickness of their lensing shells.

symbol thus decays as $1/\nu^2$ for large $\nu = \min(\nu_1, \nu_2, \nu_3)$. A necessary condition for negligible non-Gaussianity is therefore

$$\left[\frac{\sqrt{\nu_1 \nu_2 \nu_3 / 2}}{4\pi^2 A(L_1, L_2, L_3)} \right]^{1/2} \frac{|b_{l_1 l_2 l_3}|}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}} \ll 1 \qquad (28)$$

when $l_{\min} \gg 1$. The prefactor on the left-hand side of Eq. (28) is a weak function of ν, and is typically $\sim 10^{-2}$ for the relevant scales and for most triangle configurations. The reduced bispectrum for weak lensing has been measured for a few triangle configurations from N-body simulations in Ref. [42], who find that it *decays* with l for most configurations, with the slowest decay being a roughly l^{-1} dropoff in their "squeezed" configuration. This suggests that, for the lensing convergence, the bispectrum-squared term should become rapidly negligible with l. At $l \gtrsim 2000$ where shape noise starts to dominate the power spectrum [42] finds that the reduced bispectrum for an equilateral configuration is $\sim 10^{-17}$, and therefore the left-hand side of Eq. (28) is $\sim 10^{-5}$. At $l = 100$ we have $C_l \sim 10^{-9}$ and $b_{lll} \sim 10^{-17}$, so the left-hand side of Eq. (28) is again $\sim 10^{-5}$. For a squeezed shape with $l_1 = 50$, $l_2 = l_3 = 2000$ Ref. [42] finds $b \sim 10^{-16}$. We find that $C_{l_1} \sim 10^{-8}$, and the left-hand side of Eq. (28) is $\sim 10^{-4}$, i.e., slightly more significant. A typical element of $\tilde{\varepsilon}_{l_1 l_2 l_3}$ is therefore around 10^{-9}.

In practice the quantity that should be small is the *sum* of ε across the scales most relevant to the quantity of interest. The order-of-magnitude analysis above suggests that for a small subset of the C_l the correction from non-Gaussianity is negligible, but if the number of relevant scales is large (as might be the case for massive data compression in order to infer cosmological parameters) then the conditions on ε are more strict. Assuming an approximately constant $\tilde{\varepsilon}_{l_1 l_2 l_3}$, using all available scales up to $l_{\max}$ would require $\tilde{\varepsilon}_{l_1 l_2 l_3} \ll l_{\max}^{-3}$ for example. For $\tilde{\varepsilon}_{l_1 l_2 l_3} \sim 10^{-9}$ corrections become large around $l_{\max} \approx 1000$. This is within the range of scale usable angular scales for stage-IV lensing surveys, which

motivates the more quantitative study with simulations in Sec. III.

1. Non-Gaussian effects on posterior distributions

The sampling distribution of $\hat{C}_l$ itself is usually not of interest. What matters more is the posterior distribution on the C_l or on model parameters that generate the C_l. In this section we study analytically how non-Gaussianity affects the posterior, using our corrected likelihood Eq. (22).

First we study the joint posterior distribution of the set of theory C_l. This is simply given by $p(\{C_l\}|\{\hat{C}_l\}) \propto p(\{\hat{C}_l\}|\{C_l\})\pi(\{C_l\})$, where $\pi(\{C_l\})$ is the prior. Natural priors for the C_l are the uniform prior, which isolates the dependence on the likelihood, and the Jeffreys prior, which is uninformative. For a Gaussian likelihood the two coincide, but this is not the case for either the Gamma likelihood nor our Eq. (22).

The Jeffreys prior is given by the square root of the determinant of the Fisher matrix. We have computed this for our likelihood, and at leading order we have

$$\pi(\{C_l\}) \propto \left(\prod_{l=l_{\min}}^{l_{\max}} C_l^{-1} \right) \left(1 - \frac{1}{4} \sum_l \nu \frac{T_{ll}}{C_l^2} \right). \qquad (29)$$

The zero-order posterior for each C_l is thus an inverse-Gamma distribution, denoted by Γ^{-1} given by

$$p_{\Gamma^{-1}}(C_l | \hat{C}_l) = \frac{(\nu \hat{C}_l / 2)^{\nu/2}}{\Gamma(\nu/2)} C_l^{-\nu/2-1} \exp\left(-\frac{\nu \hat{C}_l}{2 C_l} \right). \qquad (30)$$

Weak signal non-Gaussianity perturbs away from this zero-order distribution.

In what follows we will only include trispectrum terms, which we have found to be the dominant source of non-Gaussianity in the power spectrum distribution. Substituting in the Jeffreys prior and normalizing, the correction to the inverse-Gamma posterior is

$$p(\{C_l\}|\{\hat{C}_l\}) = \left[\prod_{l=l_{\min}}^{l_{\max}} \Gamma^{-1}(C_l; \nu/2, \nu C_l/2) \right]$$
$$\times \left\{ 1 + \frac{1}{4} \sum_{l_1 l_2} \hat{\varepsilon}_{l_1 l_2} \left[\frac{\hat{C}_{l_1} \hat{C}_{l_2}}{C_{l_1} C_{l_2}} \left(\frac{\Delta\hat{C}_{l_1} \Delta\hat{C}_{l_2}}{C_{l_1} C_{l_2}} - \frac{2\delta_{l_1 l_2}}{\nu_1 + 2} \frac{\hat{C}_{l_1}^2}{C_{l_1}^2} - \frac{2\tilde{\alpha}\delta_{l_1 l_2}}{\nu_1} \right) + \frac{\delta_{l_1 l_2}[2\tilde{\alpha}(\nu_1 + 2) - 4] - 4}{\nu_1 \nu_2} \right] \right\} \qquad (31)$$

where $\hat{\varepsilon}$ is defined as in Eq. (24) but with $\hat{C}_l$ in place of C_l, and $\tilde{\alpha}$ is a bookkeeping parameter that keeps track of the influence of the correction to the Jeffreys prior; $\tilde{\alpha} = 1$ represents the first-order-corrected Jeffreys prior, and $\tilde{\alpha} = 0$ uses only the zero-order Jeffreys prior.

Corrections to the posterior moments from $T_{l_1 l_2}$ can now be computed easily using the properties of the inverse-Gamma distribution. For example, the mean of a single C_L after marginalizing over all other C_l receives a fractional correction of order $\mathcal{O}(\hat{\varepsilon}/\nu)$, i.e., suppressed at large ν.

The leading-order fractional correction to the marginal variance of a single C_L is $\hat{\varepsilon}_{LL}$, as expected, and is entirely captured by a Gaussian having the full non-Gaussian covariance matrix. The leading-order post-Gaussian correction from signal non-Gaussianity is $\mathcal{O}(\hat{\varepsilon}/\nu)$. Note the variance under a Gaussian approximation also differs from that of the inverse-Gamma by a fractional amount $\mathcal{O}(1/\nu)$ when ν is finite. Our formalism is valid at all orders in $1/\nu$, but does assume small ε. Corrections to the mean and variance not captured by a Gaussian with non-Gaussian covariance matrix are thus $\mathcal{O}(\hat{\varepsilon}/\nu)$. It turns out that these are also the corrections to the *conditional* mean and variance when all but one C_l are fixed to their measured values, with the mean correction further suppressed if $\tilde{\alpha} = 1$.

To study the Gaussian limit in the C_l posterior, we write $C_l = \hat{C}_l(1 + n\sqrt{2/\nu})$ and take the $\nu \to \infty$ limit. The leading-order term, quadratic in n, is precisely that specified by a Gaussian with non-Gaussian covariance. Leading-order post-Gaussian terms at fixed n are $\mathcal{O}(\varepsilon/\sqrt{\nu})$, i.e., larger than the correction to the mean and variance. This correction is a cubic polynomial in n, suggesting that it corresponds to a change in the skewness of the posterior C_l, not captured by a Gaussian. It turns out that this correction can be accounted for by a slight modification to the Gaussian where we replace $T_{l_1 l_2}$ in the covariance matrix with $T_{l_1 l_2}(\hat{C}_{l_1}^2/C_{l_1}^2)(\hat{C}_{l_2}^2/C_{l_2}^2)$, but in practice this post-Gaussian term is small and the correction does not lead to noticeable improvement. Note that we have neglected the bispectrum-squared term in this derivation, which will also modify the C_l skewness.

We can also study the distribution of an amplitude parameter by setting $C_l = AC_{0,l}$ for some fiducial power spectrum $C_{0,l}$ (we neglect noise here for simplicity). We find a fractional correction to the mean of order $\mathcal{O}(\langle\varepsilon\rangle/\sum\nu)$, where we define the l-averaged quantity

$$\langle\varepsilon\rangle \equiv \frac{\sum_{l_1 l_2}\sqrt{\nu_1\nu_2}\,\varepsilon_{l_1 l_2}}{\sum_l\nu}$$
$$= \frac{\sum_{l_1 l_2}\nu_1\nu_2\frac{T_{l_1 l_2}}{2C_{l_1}C_{l_2}}}{\sum_l\nu}, \qquad (32)$$

with ε evaluated at the fiducial model. $\langle\varepsilon\rangle$ has the simple interpretation that it is the leading-order fractional correction to the posterior parameter variance from the non-Gaussian covariance term. Similarly, the leading-order post-Gaussian fractional correction to the variance of A is $\mathcal{O}(\langle\varepsilon\rangle/\sum\nu)$. Corrections to the dimensionless skew of A are $\mathcal{O}(\langle\varepsilon\rangle/\sqrt{\sum\nu})$, as are corrections to the pointwise posterior distribution of A.

To summarize this section, the biggest effect of a signal trispectrum is to correct the posterior variance of parameters through the non-Gaussian covariance term. This is

typically of order $\langle\varepsilon\rangle$, where the average runs over the scales of interest. We require this correction to be small in order for our perturbative series to converge, i.e., the validity of our distributions require $\langle\varepsilon\rangle \ll 1$. A good rule of thumb is that we require the correction to the Fisher information from signal non-Gaussianity to be much less than unity. The broadening of parameter errors is mostly accounted for by a Gaussian likelihood with non-Gaussian covariance—the genuinely new terms we have computed here are post-Gaussian and give a skewness and fractional pointwise correction to the probability density which are both of order $\langle\varepsilon\rangle/\sqrt{\sum\nu}$ at large ν, which by Eq. (A8) is also the leading correction to the dimensionless skewness of the measured power spectrum. At large ν this is strongly suppressed due to the central limit theorem if $\langle\varepsilon\rangle$ does not grow faster than $\sqrt{\sum\nu}$. In Sec. III we quantify this with simulations. Corrections due to the squared bispectrum also affect this skewness. Corrections to the posterior mean and variance are smaller, of order $\langle\varepsilon\rangle/\sum\nu$ at large ν, which is of order the correction to the dimensionless kurtosis of the measured power spectrum.

D. Band powers

It is usually advantageous to bin the power spectrum in angular scale in order to reduce dimensionality of the required covariance matrix, account for mode mixing due to the survey mask, or to speed up likelihood calculations. Binning has a strong effect on the apparent impact of non-Gaussianity in the covariance matrix, since the Gaussian term will be suppressed due to the increasing mode count while the non-Gaussian term will stay roughly constant.

Our binning scheme is defined by

$$\hat{C}_b = \frac{\sum_{l\in b}W_l\hat{C}_l}{\sum_{l\in b}W_l}, \qquad (33)$$

for a window function W_l, where b labels the bin. For narrow bins, an optimal binning scheme would weight with the inverse covariance of $\hat{C}_l$. To avoid dependence on the model we weight by the mode count and take $W_l = 2l + 1$. We explore different binning choices in Appendix C.

Assuming C_l is roughly constant over the extent of the bin, the zero-order characteristic function of the $\hat{C}_b$ is

$$\phi_{\{\nu_b\hat{C}_b\}}(\{J_b\}) = \prod_b\prod_{l\in b}(1 + 2iJ_bC_l)^{-\nu/2}$$
$$\approx \prod_b(1 + 2iJ_bC_b)^{-\nu_b/2}, \qquad (34)$$

where $\nu_b \equiv \sum_{l\in b}\nu$. In other words, the zero-order distribution of the binned power spectra is approximately a Gamma distribution. In practice we can eliminate the need to approximate a constant C_l over the bin by working with

$\hat{C}_l/C_l$, but for sufficiently narrow bins the above approximation is accurate.

In the presence of signal non-Gaussianity, we can make a similar approximation that C_l is flat across each bin and simply replace $l_{1,2,3}$ with $b_{1,2,3}$ and $\nu_{1,2,3}$ with ν_{b_1,b_2,b_3} everywhere in the likelihood. Quantities such as $T_{l_1 l_2}$ and $\tilde{B}_{l_1 l_2 l_3}$ can be replaced with their binned versions.

Since the number of modes in a bin scales with its width Δl_b, the fractional correction to the covariance scales as $\varepsilon_{bb} \propto \Delta l_b$. The number of terms in the sum over $\varepsilon_{b_1 b_2}$ in the denominator of $\langle \varepsilon \rangle$ scales between $1/\Delta l_b^2$ for nonsparse $\varepsilon_{b_1 b_2}$ and $1/\Delta l_b$ for diagonal $\varepsilon_{b_1 b_2}$. Therefore, corrections from the signal trispectrum scale somewhere between $\langle \varepsilon \rangle \propto \Delta l_b$ for diagonal ε and $\langle \varepsilon \rangle \approx$ constant for nonsparse ε. In practice the ε is not diagonal, and so we can expect our likelihood corrections to be only mildly sensitive to the choice of binning, with the generic feature that the corrections are larger for broader bins. We confirm this in Appendix C.

III. QUANTIFYING CORRECTIONS FROM NON-GAUSSIANITY

In the previous section we derived the leading-order correction to the probability density of power spectrum measurements in the presence of signal non-Gaussianity. Considered as a function of the unknown true power spectrum C_l, the signal trispectrum $T_{ll'}$, and the squared-bispectrum $\tilde{B}_{ll'l''}$, this distribution gives the likelihood function. It is the impact of non-Gaussianity on the likelihood (or posterior) that is of most interest to us here.

We have seen that the Fisher information contributed by the parameter dependence of $T_{ll'}$ and $\tilde{B}_{ll'l''}$ is formally of higher order in the (assumed weak) non-Gaussianity, so in this section we will fix these to a fiducial model. Our approach is to use approximate simulations to compute these quantities in a fiducial cosmology, but we could just as well compute them using the halo model or fitting functions. Once we have computed these non-Gaussian signal cumulants, we can use Eq. (22) to compute the correction to the likelihood of the C_l or any parameters on which C_l depends.

A. Simulations

To compute $T_{ll'}$ and $\tilde{B}_{ll'l''}$ we create a large suite of log-normal lensing convergence maps using the FLASK software package [44]. Log-normal fields are known to give a reasonable approximation for the covariance of lensing two-point functions [45,46], although their ability to predict configurations of the bispectrum is less well explored [47]. We will cross-validate our results with a small sample of lensing maps from ray-traced N-body simulations.

We generated 10^5 log-normal convergence maps with a fixed power spectrum generated with CAMB [48,49]. Our fiducial galaxy sample is modeled with a Euclid-like lensing sample in mind. We assume a single broad redshift distribution for lensing sources, peaked around $z \approx 0.9$, having the same functional form as Ref. [50], with cosmological parameters set to the best-fit values of Ref. [51]. The alternative choice of a narrow source redshift bin peaked at $z_s = 0.325$, corresponding to the lowest redshift bin of a Euclid-like survey with ten tomographic bins, is explored in Appendix D. We compute the power spectrum to $l_{\max} = 6143$, applying nonlinear and baryon feedback corrections according to the default specifications of Ref. [52]. For simplicity we do not include intrinsic alignments in the model; this will not affect our results. The "shift parameter" corresponding to the (negative) minimum value of the convergence fed to FLASK is set to 0.01214, which was chosen by hand to get a good match to N-body simulations, and is comparable to the value found in Ref. [30] by fitting the one-point kurtosis of the convergence. Convergence fields are rendered on a HEALPIX[6] [53] grid with $N_{\text{side}} = 2048$. In our actual likelihood computations we only use scales down to $l_{\max} = 1000$, which is sufficiently small to ensure the input power spectrum is recovered with negligible bias,[7] but sufficiently large to cover a realistic range of scales for lensing surveys. We apply the HEALPIX pixel window function to the map to mimic the effect of coarse binning of a higher resolution galaxy catalog.

To check that the log-normal fields are producing sensible results for the signal cumulants, we compare the convergence statistics with those from a suite of ray-traced N-body simulations from Ref. [43]. These simulations assume a thin source redshift plane at $z_s = 1.0334$, but the resulting lensing power spectrum only differs from a fiducial C_l by a few percent across most scales.

In Figs. 1 and 2 we compare the covariance matrix of the measured $\hat{C}_l$ from the log-normal mocks with those of the N-body mocks. Since the latter contains much fewer realizations (~ 100) the recovered statistics are much noisier, but a meaningful comparison is still possible. Figure 1 shows the ratio of the diagonal elements of the covariance matrix compared with a Gaussian prediction based on the input power spectrum (we have binned the $\hat{C}_l$ to emphasize the differences). The agreement between the two is reasonable, with the N-body mocks having excess variance on large scales, as discussed in Ref. [43]. The log-normal mocks overestimate the variance on small scales, but since we are only seeking approximations to the trispectrum here the difference is sufficiently small. The non-Gaussian contribution to the covariance dramatically kicks up when $l > 200$, due to the signal trispectrum. Note

---

[6]https://healpix.sourceforge.io.

[7]The exponentiation of the Gaussian field needed to produce a log-normal field transfers power to subgrid scales where it cannot be recovered from the discrete HEALPIX grid. This causes a negative bias in the power spectrum on all scales that increases in severity as the Nyquist frequency $l_{\text{Ny}} \approx 3N_{\text{side}} - 1$ is approached.
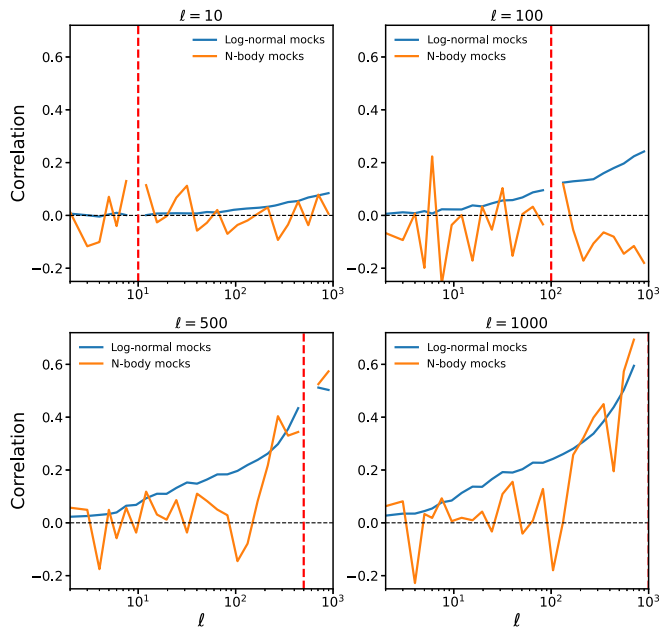
FIG. 2. Slices through the correlation matrix for a few multipole bins containing the l values indicated above each panel. The diagonal elements (unity) has been omitted for clarity and are indicated by the red dashed vertical lines.
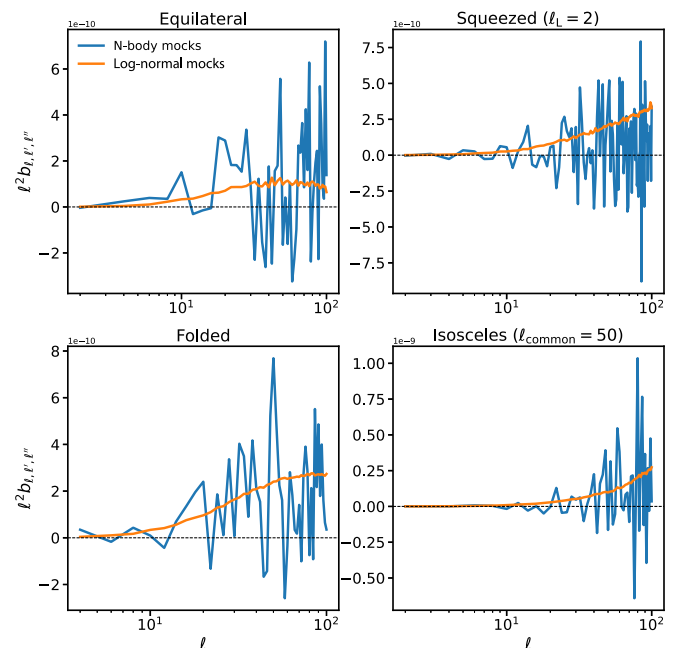

FIG. 3. Bispectra of our log-normal mocks (orange) and the N-body mocks of Ref. [43] (blue). We show an equilateral configuration (top left), a squeezed configuration with long wavelength mode $l_\mathrm{L} = 2$ (top right), a folded configuration with $l = 2l' = 2l''$ (bottom left), and an isosceles configuration with common side length $l_\mathrm{common} = 50$ (bottom right). The bispectra have not been binned in multipole.

that our log-normal maps are on the full sky and are produced by a local nonlinear transform of the linear field, and so do not capture the supersample covariance, but the agreement with the N-body mocks is nonetheless acceptable.

In Fig. 2 we plot slices through the correlation matrix. The agreement of the log-normal mocks with the N-body mocks is not perfect, but acceptable. Above $l \approx 200$ mode coupling in the N-body mocks kicks up dramatically, much in the same way as in the diagonal elements. The log-normal mocks do not capture this well, instead predicting a smooth increase, but the agreement is within an order of magnitude and acceptable.

To measure the bispectrum from the simulations we use the estimator

$$
\hat{b}_{l_1 l_2 l_3} = \frac{1}{N_{l_1 l_2 l_3}} \sum_{m_1=-l_1}^{l_1} \sum_{m_2=-l_2}^{l_2} \sum_{m_3=-l_3}^{l_3}
$$
$$
\times \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3}, \quad (35)
$$

where

$$
N_{l_1 l_2 l_3} = \sqrt{\frac{\nu_1 \nu_2 \nu_3}{4\pi}} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}, \quad (36)
$$

which is optimal for Gaussian fields. We use the PYTHON package SPHERICAL[8] to precompute the Wigner $3j$ symbols

---

[8]https://pypi.org/project/spherical/.

to allow for rapid summation over the azimuthal wave numbers in Eq. (35), but the implementation is slow for $l > 100$ so we limit our bispectrum measurements to this maximum multipole.

In Fig. 3 we show the reduced bispectrum measured from our log-normal convergence maps and from the N-body mocks of Ref. [43]. The agreement is reasonable, although the noise in the N-body measurement is large. Note that the strength of log-normality was chosen to give a good match to the non-Gaussian part of the power spectrum covariance, so the agreement in the bispectrum is not guaranteed by construction. We also find qualitative agreement with the measurements of Ref. [42].

To mitigate the impact of noise in our likelihood we fit the non-Gaussian signal polyspectra with smooth multivariate fifth-order polynomials. To make these low-order polynomials good fits we first bin the polyspectra, as described in Sec. II D, in equally spaced bins in log l. We separately fit a univariate polynomial to the ratio of the diagonal elements of the covariance to the Gaussian prediction and a bivariate polynomial to the correlation matrix (excluding the diagonal elements). We fit a trivariate polynomial to the reduced bispectrum before squaring and forming the quantity $\tilde{B}_{l_1 l_2 l_3}$. We take care to include only the nonzero elements of the bispectrum in the fit.
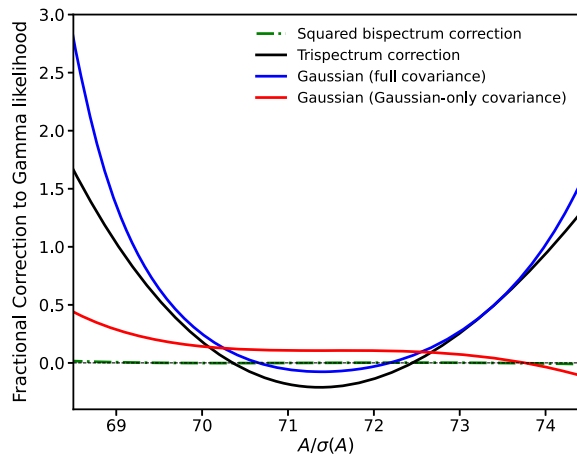
FIG. 4.   Fractional correction to the Gamma likelihood from a non-Gaussian signal. We assume a fixed measurement of the power spectrum $\hat{C}_l$ and look at the correction to the likelihood as function of the model C_l. Additionally all values of the model C_l have been scaled relative to $\hat{C}_l$ by an amplitude parameter, $C_l = A\hat{C}_l$. The corrections are shown as a function of A scaled by its Gamma 1σ uncertainty σ, i.e., its signal-to-noise ratio. The non-Gaussian covariance (trispectrum) term (black solid) and squared-bispectrum terms (dot-dashed green) are shown separately, along with the fractional difference of a Gaussian with covariance containing only a Gaussian part (red) or additionally the non-Gaussian part (blue).

With $T_{ll'}$ and $\tilde{B}_{ll'l''}$ in hand we can now compute the size of the correction to weak lensing likelihoods and posteriors from signal non-Gaussianity, using Eq. (22).

B. Results

The zero-order likelihood function for fixed $\hat{C}_l$ is proportional to an inverse-Gamma distribution. In Fig. 4 we show the fractional correction away from this due to signal non-Gaussianity, using $\hat{C}_l$ that have been binned in multipole with $\Delta \log l = 0.1$.[9] In this figure we assume a fixed measured $\hat{C}_l$ and show the correction to the likelihood when the model C_l are uniformly scaled relative to $\hat{C}_l$, i.e., we take $C_l = A\hat{C}_l$ for an amplitude parameter A. The width of the likelihood function is typically $\sigma \equiv 1/\sqrt{\sum \nu/2}$, the zero-order Gamma (or Wishart when $p \neq 1$) uncertainty, so we normalize the horizontal scale in Fig. 4 by this quantity. The best-fit value of A is close to unity, so the actual values plotted on the horizontal axis reflect the $l_{\max} = 100$ that we have assumed for this plot. We plot the contributions from the trispectrum and the squared-bispectrum separately. We have verified that the "binned" version of our model for the sampling distribution provides a close match to the histogram of measured power spectra from the log-normal mocks across the range of scales contributing to the all

the figures in the section, although we note that the mean and covariance are both matched by construction.

Figure 4 shows that the contribution from the squared-bispectrum $\tilde{B}_{ll'l''}$ is subdominant to that of the trispectrum in the vicinity of a few σ around the best-fit model, which is the range of interest. As expected due to the high mode count contributing to information on A the zero-order distribution is close to Gaussian, as demonstrated by the red curve—note that uniform contributions to the fractional difference will get normalized away when we compute posteriors, as shown shortly. In blue we show the fractional correction of a Gaussian having the correct covariance matrix including the trispectrum contribution. In agreement with the discussion in Sec. II C, this provides an accurate approximation to our distribution, with the main effect being a quadratic correction corresponding to the broadening of parameter error bars.

In practice we are not so much interested in the likelihood as the *posterior*. The easiest posterior to compute from our likelihood is the conditional posterior of a single C_l assuming all others are fixed to their measured values. This is shown in Fig. 5 for $l_{\max} = 100$ and again using binned power spectra with $\Delta \log l = 0.1$, which is the same binning choice used to fit the polyspectra from the simulations.[10] The bin barycenters are shown above each panel in Fig. 5. On large angular scales (top left panel), non-Gaussianity in the signal is weak and our distribution matches the Gamma distribution very closely (black and yellow curves are indistinguishable). A Gaussian approximation, in contrast, is poor on these scales. As the angular scale of the multipole bin is decreased, the Gaussian and Gamma approximations match ever more closely, with very little difference between any of the distributions at $l \approx 30$. As the angular scale decreases further, the effects of signal non-Gaussianity start to become important and broaden the distribution through the trispectrum contribution to the covariance matrix. Thus, our distribution starts to deviate from the Gamma and Gaussian-with-Gaussian-covariance distributions (themselves now indistinguishable) visibly at $l \approx 100$. Most of this deviation is captured by a Gaussian with a non-Gaussian covariance (blue curve), with the residual post-Gaussian effect being a small amount of skewness in the distribution. In contrast to the Gamma and Gaussian approximations, our distribution is the only model that correctly matches the large-scale and weakly nonlinear-scale limits of the true C_l distribution. Similar statements can be made of the marginal distribution of the C_l, whose leading-order properties were discussed in Sec. II C 1.

It is interesting to note from Fig. 5 that the regime where Gamma and Gaussian differ ($l \lesssim 30$ in this case) is distinct

---

[9]The sensitivity to binning choices is mild and is discussed in Sec. II D and Appendix C.

[10]Note that the binning implementation can differ between that used in the likelihood and that used to fit the polyspectra, but we found using different choices gave negligible effect on the results.
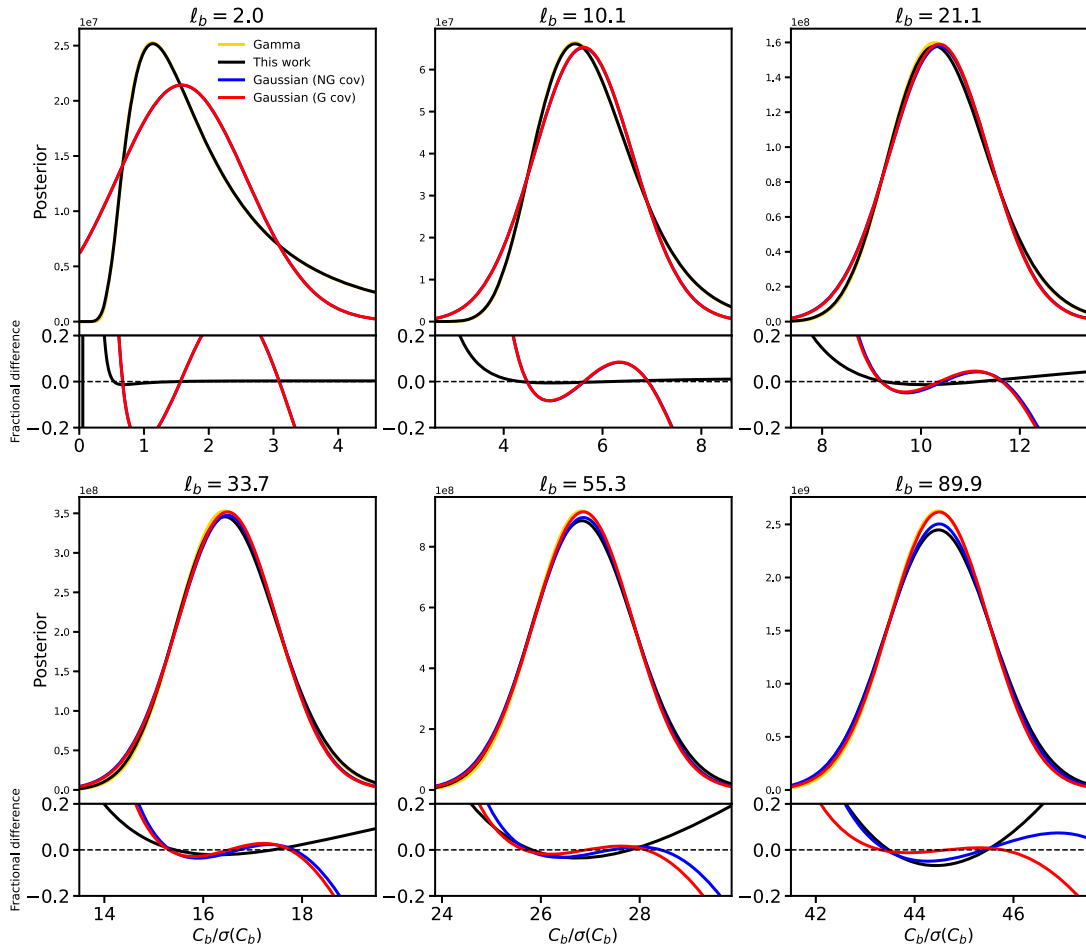
FIG. 5. Posteriors of individual power spectrum band powers, assuming all others are fixed to their measured values, labeled by their weighted central values above each panel. A Jeffreys prior has been assumed for all distributions, and we use scales up to $l_{max} = 100$. Our model (black) is indistinguishable from the Gamma likelihood (yellow) in the upper panels. We also show a Gaussian with Gaussian covariance (red, G cov) and non-Gaussian covariance (blue, NG cov). Differences between these are indistinguishable in the upper panels, and the red curve is very close to the yellow curve in the lower panels. Corrections are shown as a function of the model band power scaled by its Gamma 1σ uncertainty $\sigma(C_b)$, i.e., its signal-to-noise ratio. Our model transitions from Gamma at large angular scales to a Gaussian with non-Gaussian covariance at small angular scales. Below each panel we show the fractional differences with respect to the Gamma likelihood scenario.

from the regime where signal non-Gaussianity starts to become important ($l \gtrsim 100$). That these regimes are well separated has traditionally been the motivation for considering large-scale likelihood non-Gaussianity from the finite number of modes as a separate effect to the small-scale modification of the covariance from signal non-Gaussianity. The distinction is of course redshift dependent; as the redshift of a source bin is lowered, the angular scale at which signal non-Gaussianity becomes important will increase. We therefore expect that our new likelihood expression will be of unique value when these two sources of likelihood non-Gaussianity must be treated simultaneously, i.e., for low-redshift sources. The precise impact will additionally depend on the parameters of interest.

We explore a realistic low-z_s scenario in Appendix D, finding that skewness corrections to the posterior can become sizable (although still <10%) for $l_{max} = 1000$, but are subpercent once shape noise is included.

It is also straightforward to compute the posterior of an amplitude parameter scaling the signal power spectrum (we again neglect noise for simplicity) as $C_l = A\hat{C}_l$, shown in Fig. 6. Most of the constraining power on the amplitude comes from small scales, and hence the Gamma and Gaussian-with-Gaussian covariance (yellow and red curves respectively) overlap closely. Our distribution (black curve) correctly captures the significant increase in parameter variance from the trispectrum. As we have seen in Fig. 5, a Gaussian with appropriately modified covariance (blue
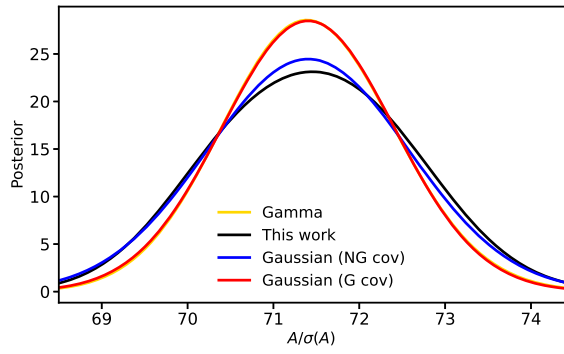
FIG. 6.   Posterior of an amplitude parameter A linearly scaling the model power spectrum, as a function of A scaled by its Gamma 1σ uncertainty. A Jeffreys prior has been assumed, and we have included scales up to $l_{\max} = 100$. From top to bottom in the center of the figure we plot the Gamma likelihood (yellow), a Gaussian likelihood with Gaussian covariance matrix (yellow, overlapping with red), a Gaussian likelihood with non-Gaussian covariance matrix (blue), and our model (black).

curve) captures most of the effect of signal non-Gaussianity, with our distribution imparting a small residual skewness into the posterior.

Since our formalism is based on a perturbative series expansion of the signal distribution, it is important to check that we are working in a regime where higher-order terms can be safely neglected. In Sec. II C 1 we showed that the leading-order correction to the posterior from signal non-Gaussianity is $\langle \varepsilon \rangle$, defined in Eq. (32). This quantity is the trispectrum contribution to the $\hat{C}_l$ covariance matrix divided by the Gaussian contribution $\varepsilon_{l_1 l_2}$, averaged over the scales contributing to the parameter of interest with a ν-dependent weight. It has a simple interpretation that it is equal to the perturbation to the Fisher information from signal non-Gaussianity. We require $\langle \varepsilon \rangle \ll 1$ for our perturbative series to be valid, with the leading-order post-Gaussian (i.e., beyond that predicted by a Gaussian functional form) skewness of order $\langle \varepsilon \rangle / \sqrt{\sum \nu}$. We have found bispectrum-squared terms to be subdominant, but they are potentially non-negligible when $l \gtrsim 1000$ by the arguments of Sec. II C.

In Fig. 7 we show a few convergence statistics that can be derived from $\varepsilon_{l_1 l_2}$. The left panel shows quantities relevant to the posterior of a single C_l, in which case $\langle \varepsilon \rangle$ is equivalent to ε_{ll}, the diagonal elements of the matrix $\varepsilon_{l_1 l_2}$. We also show the largest eigenvalue (i.e., the spectral radius) of this matrix. A spectral radius much less than unity is sufficient to ensure a valid perturbative expansion of the inverse covariance matrix around its Gaussian term, and must hold if our expression is to give a valid approximation to the pairwise distributions of C_l with different l.

As the blue curves in Fig. 7 show, the fractional correction from signal non-Gaussianity is much less than unity out to $l \approx 100$, which justifies our use of this maximum scale in the posteriors computed in this section
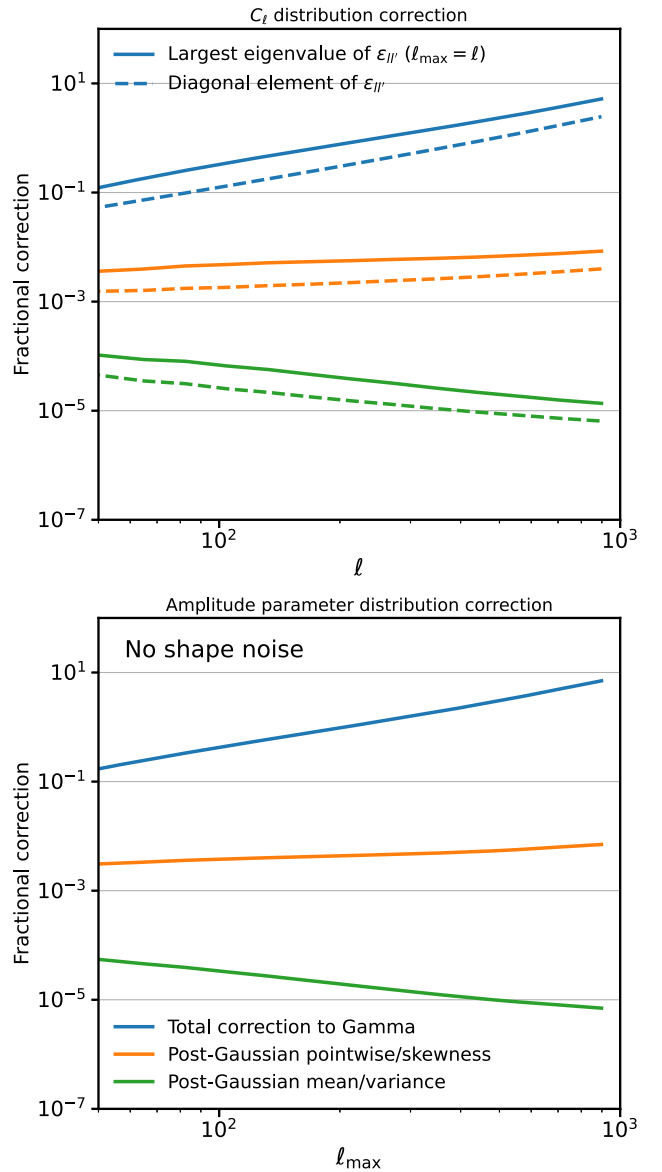

FIG. 7.   Leading-order fractional corrections to the posterior from a non-Gaussian signal as a function of angular wave number, valid in the $l \gg 1$ limit. Top: corrections to the Gamma posterior (conditional or marginal) of the model C_l. Shown are the spectral radius (i.e., largest eigenvalue) of the $\varepsilon_{ll'}$ matrix defined in Eqs. (24) and (32) (solid blue) and its diagonal elements (dashed blue). These quantities (blue top set of curves) are required to be small for higher-order non-Gaussian terms to be negligible. Leading-order corrections to the Gamma posterior not captured by a Gaussian functional form (post-Gaussian effects) are represented by the orange curves (middle set), and subdominant corrections to the mean and variance by the green curves (lower set). These are suppressed by factors of $\nu^{-1/2}$ and ν^{-1} respectively. Bottom: equivalent corrections to the posterior of an amplitude parameter, now showing the cumulative statistic $\langle \varepsilon \rangle$ (see text) up to $l_{\max}$ (blue, top curve) and leading post-Gaussian corrections (orange and green). The interpretation of these curves is the same as in the top panel, i.e., they represent the total correction to the Gamma posterior from a non-Gaussian signal (blue), the corrections not accounted for by a Gaussian at leading order in $\nu^{-1/2}$ (orange), and next-to-leading order (green).

and explains why the correction appears so small in Fig. 5. The leading-order corrections not captured by a Gaussian functional form are given by the orange curves and are never more than 10^{-2} even out to $l \approx 1000$, i.e., subpercent across all scales of interest for upcoming lensing surveys, an important result. Beyond $l \approx 100$ our perturbative series breaks down, but the quantity $\langle \varepsilon \rangle / \sqrt{\sum \nu}$ is still indicative of the relevance of residual non-Gaussianity in the likelihood, so it is reasonable to draw useful conclusions from Fig. 7 even beyond the regime where our likelihood expression is formally valid. We note that $\langle \varepsilon \rangle / \sqrt{\sum \nu}$ is the leading-order correction to the dimensionless skewness of the measured power spectrum from signal non-Gaussianity, valid at all (sufficiently large) l, so it is reasonable to expect that this quantity also quantifies non-Gaussian corrections to the likelihood function. Corrections to the posterior mean and variance not captured by a Gaussian are even more subdominant, suppressed by another factor of $\sqrt{\nu}$.

The right-hand panel of Fig. 7 shows $\langle \varepsilon \rangle$ for an amplitude parameter scaling the model power spectrum. The corrections are comparable in size with those of the individual C_l, with $l_{\max} = 100$ just about in the regime where our expression is valid. Non-Gaussian effects are still subpercent even out to $l_{\max} = 1000$. In Appendix C we study the (mild) impact of choosing a different binning scheme for the power spectrum band powers, shown in Fig. 9.

Taken together, Figs. 5 and 7 give the following picture for how the likelihood changes with angular multipole. On the largest scales a Gamma distribution is accurate as the signal is close to Gaussian. By $l \approx 30$ the Gamma distribution has Gaussianized, with corrections being percent level or small in the central 2σ region (albeit larger in the tails). As $l \approx 100$ is approached the non-Gaussianity in the signal starts to become important, and a Gaussian with appropriately broadened covariance is a good approximation. Corrections to the Gamma distribution not captured by the Gaussian approximation are subpercent across all angular scales, but can be accounted for by using our expression Eq. (22) on large scales and smoothly matching it to a Gaussian around $l \lesssim 100$.

One question we wish to answer in this work is whether the strength of signal non-Gaussianity grows faster than the CLT is able to Gaussianize the power spectrum distribution. From Fig. 7 it appears that the struggle between these two competing effects is delicately balanced, with a very slow growth in the post-Gaussian correction from signal non-Gaussianity (orange curve) visible as $l_{\max}$ increases. One might wonder whether this correction becomes non-negligible at very high $l_{\max}$. At some point as $l_{\max}$ is increased the effect of the noise will become important, which suppresses ε by boosting C_l while not adding any non-Gaussianity (we assume the noise is Gaussian here). In Fig. 8 we show the size of post-Gaussian effects (i.e.,
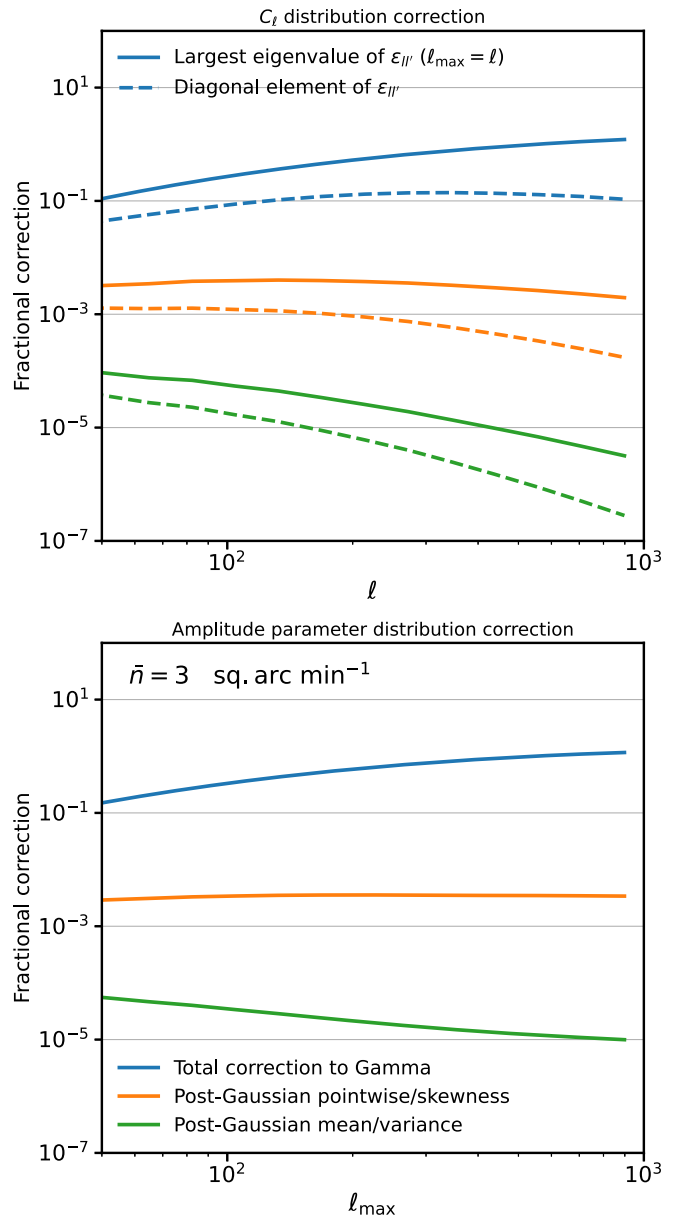

FIG. 8. Same as Fig. 7 but with shape noise corresponding to $\bar{n} = 3$ sq. arc min^{-1}.

corrections to the Gamma distribution not captured by a Gaussian likelihood) on the posterior with shape noise given by $\sigma_\gamma^2 = \sigma_e^2 / \bar{n}$, i.e., the model power spectrum is $C_l = A S_{0,l} + N_l$. We assume a per-component ellipticity rms of $\sigma_e = 0.21$ and set the source number density to $\bar{n} = 3$ arc min^{-2}, as might be expected in a single tomographic bin in a Euclid-like survey with ten redshift bins [50]. The Gaussianizing effects of noise start to become important at $l \gtrsim 100$ and ensure that post-Gaussian effects in the likelihood are never more than subpercent. For $\bar{n} = 30$ arc min^{-2} this turnover happens around $l_{\max} \approx 1000$, suggesting that post-Gaussian effects are safely negligible on small scales for stage-IV surveys.

As discussed above, it will likely be the case that post-Gaussian effects are most relevant when the largest angular scales in the survey are mildly nonlinear, i.e., for low-redshift tomographic bins. These carry little statistical weight in a Euclid-like survey where most of the sources are around $z \approx 1$, but could be relevant in other surveys. In Appendix D we explore the corrections for the lowest source redshift bin of a ten-bin lensing survey having intrinsic source redshifts comparable to that of a Euclid-like survey, showing that corrections are subpercent once the effects of shape noise are included. This is partly because the amplitude of the shear power spectrum is suppressed for nearby sources due to the lensing geometry, so it may be that effects are larger for projected galaxy clustering. The quantity $\langle \varepsilon \rangle / \sqrt{\sum \nu}$ or, alternatively, the dimensionless skewness of the measured power spectrum, provides an easy-to-compute diagnostic of whether one needs to worry about non-Gaussianity in the likelihood when low-redshift bins are important. If the effects are large, one will also need to account for effects of the mask in a full pseudo-C_l analysis to build an accurate likelihood model, as in Refs. [12,13,54,55], or resort to map-level or likelihood-free methods.

IV. CONCLUSIONS

We have computed, for the first time, the leading-order correction to the distribution of angular power spectrum estimates of a projected field from non-Gaussianity in the underlying signal. Our expressions are given in Eq. (22) in the case of a single map and by the sum of Eqs. (B35) and (B56) in the case of multiple correlated maps. These expressions correct the Wishart distribution for the matrix of power spectra and consist of a term proportional to the signal trispectrum and a term proportional to the square of its bispectrum. The expression can easily be generalized to include non-Gaussian noise.

Our expressions are correctly normalized and have the correct mean, covariance, and three-point function. We have shown that, in the limit of a large number of degrees of freedom and sufficiently well-behaved non-Gaussianity, the main effect of signal non-Gaussianity is to correct the covariance matrix of the power spectra by the standard trispectrum term. The residual post-Gaussian effect of non-Gaussianity is to impart some skewness to the likelihood, on top of that present due to the residual effects of finite degrees of freedom. On small scales, the size of the correction to the posterior of a parameter is given approximately by the fractional correction to the Fisher information from the trispectrum, scaled by the inverse square root of the total degrees of freedom that contribute. For a given C_l and $l \gg 1$ the relevant quantity is $\Delta_l \equiv \sqrt{l/2} T_{ll}/C_l^2$, where T_{ll} is the trispectrum contribution to the variance. When $\Delta_l \ll 1$, corrections to the likelihood of the C_l that

are not captured by a Gaussian with appropriately modified covariance are suppressed.

We have provided diagnostic statistics, given in Eqs. (24) and (32), that quantify corrections to likelihoods and posteriors. These can be computed straightforwardly from a model for the non-Gaussian part of the covariance matrix. Such a model is already a requirement when using a likelihood for two-point statistics, so these statistics should be easy to compute for a given observational setup.

We quantified corrections to the likelihood function using mock non-Gaussian weak lensing maps created assuming log-normal statistics, with input power spectra and noise levels representative of an upcoming stage-IV survey. The bispectrum-squared term is subdominant to the correction from the trispectrum at least to $l \approx 100$, but we do not rule out stronger effects on small angular scales. Corrections to the likelihood not captured by a Gaussian are strongly suppressed on small scales in all our tests, which provides evidence that a Gaussian likelihood is sufficient on small scales even in the presence of realistic signal non-Gaussianity.

On large angular scales our likelihood correctly tends to a Gamma/Wishart distribution. We have found that the angular scales where the Gamma transitions to a Gaussian are sufficiently well separated from those where the trispectrum becomes important that it is sufficient to approximate the functional form of the likelihood as Gaussian when incorporating signal non-Gaussianity. This is the approach that has been taken traditionally, albeit without rigorous justification. This property remains true for the lowest redshift bin of a ten-bin Euclid-like lensing survey, but may break down for galaxy clustering power spectra at low redshifts. The likelihood function we have derived is unique among commonly used likelihoods in that it has the correct behavior on both large and small angular scales, being nonperturbative in the degrees of freedom.

Our model relies on the signal non-Gaussianity being weak in a certain, well-defined, sense. We have taken care to identify the regime where this approximation is valid, but have argued that requiring $\Delta_l \ll 1$ is a good diagnostic for the post-Gaussian effects even beyond the limiting scale where our expansion is valid. Use of our likelihood in a practical situation will require ensuring that $(2l+1)\Delta_l \ll 1$ or, more generally, that the corrections from signal non-Gaussianity are small in the likelihood or posterior of interest. This can also be checked by ensuring that non-Gaussianity in the sampling distribution of the relevant combinations of the *measured* power spectrum is small, which can be checked with simulations (which are in any case necessary to validate or produce covariance matrix estimates).

We have elucidated how the central limit theorem competes with signal non-Gaussianity in suppressing

non-Gaussian corrections to the likelihood. The balance is surprisingly delicate, but our conclusion is that our Universe is not non-Gaussian enough on the relevant scales that post-Gaussian terms in the likelihood are necessary in the high-l regime—the central limit theorem wins. This conclusion is strengthened once Gaussian noise (e.g., from galaxy intrinsic shapes or shot noise) is present in the data. This essentially explains why previous works such as Refs. [27,28] have found a Gaussian likelihood to be sufficient for ΛCDM parameter inference in the presence of realistic signal non-Gaussianity. Our findings are also consistent with the results of Ref. [29], who found that by far the dominant source of non-Gaussianity in the power spectrum likelihood comes from large angular scales where signal non-Gaussianity is negligible. Our formalism has allowed us to cleanly separate the effects of non-Gaussianity in the signal with that from the quadratic nature of the power spectrum estimator, improving upon previous numerical works that have been unable to distinguish between these distinct effects.

An important limitation of our likelihood is that it is only valid for full-sky power spectrum estimates. For the standard pseudo-C_l estimator even the zero-order Gaussian-field likelihood is difficult to compute [13], and accounting for the mode coupling entailed by a survey mask in the non-Gaussian case seems intractable at first sight. The main effect of the mask is to reduce the number of degrees of freedom, increasing parameter variances. As the size of the survey is reduced, we can expect that at some point the largest observable scale is nonlinear, in which case a suitably modified version of our likelihood might be a good place from which to build an approximation. The effects of non-Gaussianity from the signal in the case of small surveys may be behind the fairly large posterior corrections found in Ref. [20]. In the case that the full-sky power spectra are estimated from the data directly (as in quadratic maximum likelihood approaches) our distribution is likely to be more accurate and bears some resemblance to the approximations developed in Ref. [56]. For small survey cuts the mode coupling induced by the mask may only be relevant in the regime where the signal is linear, in which case an $f_{\rm sky}$ correction to the number of degrees of freedom may be sufficient to incorporate the survey mask into our likelihood. However, it may be the case that a map-level likelihood such as that used in cosmic microwave background analyses [57], a likelihood-free approach [58], or suitable data transformations [59] may be the way forward for modeling these effects in the distribution of projected fields.

While this work has studied the power spectrum likelihood, most weak lensing surveys up to date have used the shear correlation functions to derive their cosmological constraints. An analytic expression for the correlation function likelihood is difficult to write down even in the case of linear shear fields, but may be inferred numerically by sampling power spectra and subsequently transforming to real space. We defer an analytic study of this to a future work, but note that the results of the numerical studies in Refs. [27,28] suggest that non-Gaussian corrections are likely negligible for ΛCDM models.

This work represents the first rigorous analytic study of the use of a Gaussian approximation for likelihood function of the power spectrum of non-Gaussian fields. While residual non-Gaussian effects appear small on small angular scales, we recommend that this be checked on a survey-by-survey basis using the diagnostic statistics defined in Eqs. (24) and (32). With the huge increase in statistical constraining power offered by near-future wide imaging surveys, it will be increasingly important to ensure that no biases arise from an incorrect statistical description of the data. Since we have found no evidence that signal non-Gaussianity imparts significant non-Gaussianity into the likelihood beyond that coming from the finite degrees of freedom, our recommendation for upcoming stage-IV lensing surveys is to model the power spectrum likelihood as Gaussian (with the correct nonlinear covariance matrix) on small scales $l \gtrsim 30$. On large scales in the case that mode coupling from the mask is irrelevant the Gamma (or Wishart in the case of multiple maps) distribution should be used, and when mode coupling is relevant one should resort to map-level, likelihood-free, or existing likelihood approximations. The likelihood we have derived in this work may be used as a bridge between these two regimes.

ACKNOWLEDGMENTS

APPENDIX A: POWER SPECTRUM CUMULANTS AND THE EDGEWORTH EXPANSION

When considering how non-Gaussianity in a signal affects the distribution of the power spectrum, it is useful to start by studying the first few cumulants. In this section we will keep the assumption of Gaussian noise and specialize to a single redshift bin ($p = 1$) for simplicity.

First, let us compute the covariance of the power spectrum estimates. Non-Gaussianity contributes via the trispectrum of the signal as

$$\langle \Delta \hat{C}_l \Delta \hat{C}_{l'} \rangle = \frac{2}{\nu} C_l^2 \delta_{ll'} + T_{ll'}, \qquad (A1)$$

where $T_{ll'}$ is an m-averaged version of the signal trispectrum; explicit forms for this are given in Ref. [61]. We remind the reader here that C_l is the sum of signal and noise power. The trispectrum term $T_{ll'}$ is in general difficult to compute, with current strategies in weak lensing focused around halo models [62,63] or N-body simulations [64,65].

Now consider the three-point function of $\hat{C}_l$. This consists of terms having six a_{lm} factors. We can compute this from the definition Eq. (1), using Wick's theorem to pick out the various contributions. This gives

$$\langle \Delta\hat{C}_{l_1} \Delta\hat{C}_{l_2} \Delta\hat{C}_{l_3} \rangle = 8 \frac{\delta_{l_1 l_2} \delta_{l_2 l_3}}{\nu_1^2} C_{l_1}^3 + 4\tilde{B}_{l_1 l_2 l_3} + 2\tilde{S}_{l_1 l_2 l_3}$$
$$+ 4[3]\frac{\delta_{l_1 l_2}}{\nu_1} C_{l_1} T_{l_1 l_3} + P_{l_1 l_2 l_3}, \qquad \text{(A2)}$$

where the notation [3] denotes the three terms resulting from symmetrization over wave numbers in terms to the right, $P_{l_1 l_2 l_3}$ is an m-averaged connected signal six-point function ("pentaspectrum"), and the bispectrum-squared quantities $\tilde{B}$ and $\tilde{S}$ are defined as

$$\tilde{B}_{l_1 l_2 l_3} \equiv \frac{1}{\nu_1 \nu_2 \nu_3} \sum_{m_1, m_2, m_3} \langle s_{l_1 m_1} s_{l_2 m_2} s_{l_3 m_3} \rangle \langle s_{l_1 m_1}^* s_{l_2 m_2}^* s_{l_3 m_3}^* \rangle,$$
$$\text{(A3)}$$

$$\tilde{S}_{l_1 l_2 l_3} \equiv \frac{1}{\nu_1 \nu_2 \nu_3} \sum_{m_1, m_2, m_3} \langle s_{l_1 m_1} s_{l_1 m_1}^* s_{l_2 m_2} \rangle \langle s_{l_2 m_2}^* s_{l_3 m_3} s_{l_3 m_3}^* \rangle$$
$$+ \text{sym.} \qquad \text{(A4)}$$

By statistical isotropy the signal bispectra must be invariant under rotations, which constrains the terms in angle brackets in Eqs. (A3) and (A4) to be proportional to $3j$ symbols that carry all the m dependence [66]. We will further impose that the signal statistics are invariant under a parity inversion, which introduces a second $3j$ symbol imposing that all the wave numbers sum to an even integer. Explicitly we have

$$\langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle$$
$$= \sqrt{\frac{\nu_1 \nu_2 \nu_3}{4\pi}} \begin{pmatrix} l_1 & l_1 & l_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_1 & l_1 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} b_{l_1 l_2 l_3},$$
$$\text{(A5)}$$

where $b_{l_1 l_2 l_3}$ is the reduced bispectrum. Substituting this expression into Eqs. (A3) and (A4), we see that $\tilde{S}$ is proportional to terms like

$$\sum_{m_1} (-1)^{m_1} \begin{pmatrix} l_1 & l_1 & l_3 \\ m_1 & -m_1 & m_3 \end{pmatrix} = \delta_{l_3 0} \delta_{m_3 0} (-1)^{l_1} \sqrt{2l_1 + 1},$$
$$\text{(A6)}$$

where the identity follows from the orthogonality of the $3j$ symbols. Since the signal $l = 0$ mode is unobservable, the overall contribution from terms of this form is zero, and hence $\tilde{S} = 0$. For $\tilde{B}$ we have

$$\tilde{B}_{l_1 l_2 l_3} = \frac{1}{4\pi} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}^2 b_{l_1 l_2 l_3}^2. \qquad \text{(A7)}$$

As mentioned above, the $3j$ symbol imposes that $l_1 + l_2 + l_3$ is an even integer, as well as the triangle conditions $|l_1 - l_2| \le l_3 \le l_1 + l_2$.

The three-point function of power spectrum estimates thus receives contributions from squared bispectra, products of trispectra and power spectra, and the connected six-point function. Similarly, the connected four-point function will receive contributions from squared trispectra, squared bispectra multiplied by power spectra, bispectra multiplied by the connected five-point function, pentaspectra multiplied by power spectra, and the connected eight-point function.

We want to keep our model for non-Gaussianity as general as possible, so in principle we would like to keep track of the whole cumulant hierarchy of the power spectrum estimates. This is clearly going to be computationally unfeasible however, so we will hereafter assume that the signal non-Gaussianity is *weak*.

1. Weak non-Gaussianity

We will assume that non-Gaussianity in the signal is weak in the sense that each term in its cumulant hierarchy is successively smaller than the last. Introducing the order-counting parameter λ, we will assume that $C_l \sim \mathcal{O}(\lambda)$, $b_{l_1 l_2 l_3} \sim \mathcal{O}(\lambda^2)$, $T_{l_1 l_2} \sim \mathcal{O}(\lambda^3)$, etc., with $\lambda \ll 1$. This is justified on sufficiently large scales where nonlinearity in the underlying field is weak, and perturbation theory is accurate. We will precisely define the regime in which this perturbative approach is valid later on.

In this scenario, bispectrum-squared terms like $\tilde{B}_{l_1 l_2 l_3}$ are the same perturbative order as power-trispectrum cross terms like $C_{l_1} T_{l_1 l_3}$, which can also be seen by expanding the fields to second order in the linear signal; the leading-order (tree-level) bispectrum features two factors of the linear power spectrum [67], meaning the squared-bispectrum contains four such factors. The trispectrum contains three power spectra at leading order, and hence $C_{l_1} T_{l_1 l_3}$ also contains four factors of the linear power spectrum. The connected six-point function is suppressed as $P_{l_1 l_2 l_3} \sim \mathcal{O}(\lambda^5)$. We will retain leading-order "non-Gamma" terms throughout. The three-point function of $\hat{C}_l$ is thus Eq. (A2) with $\tilde{S} = 0$ and the final $P_{l_1 l_2 l_3}$ term dropped.

The *dimensionless* one-point skewness of $\hat{C}_l$ can be derived from Eq. (A2) as

$$\frac{\langle\Delta\hat{C}_l^3\rangle}{\langle\Delta\hat{C}_l^2\rangle^{3/2}} = \frac{8C_l^3/\nu^2 + 4\tilde{B}_{lll} + 12C_lT_{ll}/\nu}{(2C_l^2/\nu + T_{ll})^{3/2}}$$

$$= \sqrt{\frac{8}{\nu}} + \sqrt{2\nu^3}\frac{\tilde{B}_{lll}}{C_l^3} + \frac{3\sqrt{2\nu}}{2}\frac{T_{ll}}{C_l^2} + \mathcal{O}(\lambda^2). \quad \text{(A8)}$$

It is important to note here that we could consistently neglect the $\mathcal{O}(\lambda^5)$ term $P_{l_1l_2l_3}$ in both the three-point function and its dimensionless version, since the variance itself is $\mathcal{O}(\lambda^2)$ meaning that connected six-point terms contribute at $\mathcal{O}(\lambda^2)$ in the dimensionless skew. This is in contrast to the connected four-point function of the power spectrum estimates. The one-point excess kurtosis for example is

$$\langle\Delta\hat{C}_l^4\rangle_c = 48\frac{C_l^4}{\nu^3} + 96\frac{C_l\tilde{B}_{lll}}{\nu} + 144\frac{C_l^2T_{ll}}{\nu^2}, \quad \text{(A9)}$$

where we have kept terms at $\mathcal{O}(\lambda^5)$. Working to a consistent $\mathcal{O}(\lambda^4)$ order would mean dropping all but the first term from this expression. In contrast, the dimensionless kurtosis of $\hat{C}_l$ is

$$\frac{\langle\Delta\hat{C}_l^4\rangle_c}{\langle\Delta\hat{C}_l^2\rangle^2} = \frac{48C_l^4/\nu^3 + 96C_l\tilde{B}_{lll}/\nu + 144C_l^2T_{ll}/\nu^2}{(2C_l^2/\nu + T_{ll})^2}$$

$$= \frac{12}{\nu} + 24\nu\frac{\tilde{B}_{lll}}{C_l^3} + 24\frac{T_{ll}}{C_l^2} + \mathcal{O}(\lambda^2). \quad \text{(A10)}$$

Therefore, while the raw non-Gaussian cumulant hierarchy of the power spectrum estimates can be truncated consistently in perturbation theory, *all* of the dimensionless cumulants, i.e., the cumulants of a normalized $\hat{C}_l$ estimator with unit variance, receive corrections from a non-Gaussian signal. The cumulant hierarchy of the normalized estimator cannot therefore be truncated consistently, since terms of order $\mathcal{O}(\lambda)$ always contribute.

The higher-order non-Gaussian cumulants of a normalized $\hat{C}_l$ estimator thus do not get successively smaller, in contrast to those of the signal. This observation has important consequences, since successively smaller cumulants are a requirement for the Edgeworth expansion to converge. One might be tempted to speculate that a good way of incorporating the effects of a non-Gaussian signal into the $\hat{C}_l$ distribution would be to treat the non-Gaussianity as weak and then Edgeworth expand around a Gamma distribution, but the argument above tells us this will not work; we have to instead expand the *signal* distribution around a Gaussian. Attempts to use an Edgeworth expansion directly on the distribution of $\hat{C}_l$ (or a decorrelated version of it, as in Ref. [27]) in order to incorporate signal non-Gaussianity should be treated with caution.

APPENDIX B: DERIVATION OF THE TOMOGRAPHIC REDSHIFT BIN LIKELIHOOD

In this section we derive the correction to the likelihood function of tomographic redshift bin power spectra. It is straightforward to write down the correct likelihood for full-sky spectra when the signal and noise are both Gaussian—this is the Wishart distribution for positive-definite matrices. Similarly, it is straightforward to write down a Gaussian approximation to the Wishart distribution. Here we present the derivation for the case of multiple redshift bins.

Our strategy will be to substitute the Edgeworth-expanded signal characteristic function (12) into the mode coupling integral (9) and perform the integrations to derive the characteristic function of the power spectrum. We will then inverse Fourier transform this to get the distribution of the power spectrum.

We will compute the trispectrum and bispectrum-squared corrections separately, but start with the simple case of Gaussian signal and noise.

1. Gaussian signal, Gaussian noise

As a warm-up, consider the situation where both signal and noise are Gaussian. The former should be a good approximation on large scales where the signal is a linearly evolved version of the (close to) Gaussian initial conditions, and the latter is practically guaranteed for maps made by compression from large numbers of underlying tracers (e.g., shear maps constructed from averaged galaxy shapes or density maps made by counting many galaxies).

We assume throughout that the underlying signal and noise obey statistical isotropy, such that modes with different wave vectors are uncorrelated. When the signal and noise are both Gaussian, the covariance of different modes captures all the statistical information of the field, implying that modes with different wave vectors are not only uncorrelated but statistically independent. The characteristic functions of signal and noise thus factorize, and since the Fourier transform of a Gaussian is also a Gaussian we have

$$\phi_s(\{\mathbf{k}_{lm}\}) = e^{-\frac{1}{2}\sum_{lm}\mathbf{k}_{lm}^\dagger\mathbf{S}_l\mathbf{k}_{lm}}, \quad \text{(B1)}$$

$$\phi_n(\{\mathbf{k}_{lm}\}) = e^{-\frac{1}{2}\sum_{lm}\mathbf{k}_{lm}^\dagger\mathbf{N}_l\mathbf{k}_{lm}}, \quad \text{(B2)}$$

where $\mathbf{S}_l$ and $\mathbf{N}_l$ are the signal and noise covariance matrices.

Equation (9) factorizes in each l and m, with each integral a Gaussian integral. Making these substitutions immediately gives

$$\phi_{\{\nu\hat{C}_l\}}(\{\mathbf{J}_l\}) = \prod_{l=l_{\min}}^{l_{\max}}|\mathbf{I} + 2i\mathbf{J}_l\mathbf{C}_l|^{-\frac{\nu}{2}}, \quad \text{(B3)}$$

where $\mathbf{C}_l$ is the total (signal plus noise) covariance matrix of the map. We recognize this as a product of characteristic functions of Wishart matrices.

The joint distribution of the power spectra is thus

$$p(\{\nu\hat{\mathbf{C}}_l\}) = \prod_{l=l_{\min}}^{l_{\max}} 2^{\frac{p(p-1)}{2}} \int \frac{\mathrm{d}\mathbf{J}_l}{(2\pi)^n} e^{i\mathrm{Tr}(\nu\mathbf{J}_l\hat{\mathbf{C}}_l)} |\mathbf{I} + 2i\mathbf{J}_l\mathbf{C}_l|^{-\frac{\nu}{2}},$$

(B4)

i.e., each $\hat{\mathbf{C}}_l$ is independent.

The integral in Eq. (B4) is almost the standard integral that defines the multivariate Gamma function, with the exception that it is over real symmetric matrices rather than positive-definite matrices. Now, let $\mathbf{T} \equiv 2\mathbf{C}_l^{\frac{1}{2}}\mathbf{J}_l\mathbf{C}_l^{\frac{1}{2}}$ and change integration variables from $\mathbf{J}_l$ to $\mathbf{T}$. We will assume the total covariance matrix is positive definite. The Jacobian of this transformation is

$$\mathrm{d}\mathbf{J}_l = 2^{-\frac{p(p+1)}{2}} |\mathbf{C}_l|^{-\frac{(p+1)}{2}} \mathrm{d}\mathbf{T},$$

(B5)

which follows from a standard result. Each term in the product is therefore

$$\frac{2^{-p}|\mathbf{C}_l|^{-\frac{(p+1)}{2}}}{(2\pi)^{\frac{p(p+1)}{2}}} \int \mathrm{d}\mathbf{T} e^{\frac{i}{2}\mathrm{Tr}(\mathbf{\Omega}\mathbf{T})} |\mathbf{I} + i\mathbf{T}|^{-\frac{\nu}{2}}$$

(B6)

where $\mathbf{\Omega} \equiv \nu\mathbf{C}_l^{-\frac{1}{2}}\hat{\mathbf{C}}_l\mathbf{C}_l^{-\frac{1}{2}}$ is symmetric and the integration is over all real symmetric $p \times p$ matrices $\mathbf{T}$. Our task is therefore to compute the integral

$$P_{p,\nu}(\mathbf{\Omega}) \equiv \int \mathrm{d}\mathbf{T} e^{\frac{i}{2}\mathrm{Tr}(\mathbf{\Omega}\mathbf{T})} |\mathbf{I} + i\mathbf{T}|^{-\frac{\nu}{2}}.$$

(B7)

There are several ways of solving this integral, as discussed in, e.g., Ref. [68]. Making the substitution $\mathbf{Z} = \mathbf{I} + i\mathbf{T}$ in the definition of $P_{p,\nu}(\mathbf{\Omega})$ gives

$$P_{p,\nu}(\mathbf{\Omega}) = i^{-\frac{p(p+1)}{2}} e^{-\frac{1}{2}\mathrm{Tr}(\mathbf{\Omega})} \int_{\mathrm{Re}(\mathbf{Z})=\mathbf{I}} \mathrm{d}\mathbf{Z} e^{\frac{1}{2}\mathrm{Tr}(\mathbf{\Omega}\mathbf{Z})} |\mathbf{Z}|^{-\frac{\nu}{2}},$$

(B8)

where the integral is over a fixed real part and a symmetric imaginary part. The integral is an inverse Laplace transform and can be derived from the result [69]

$$\frac{2^{\frac{1}{2}p(p-1)}}{(2\pi i)^{\frac{1}{2}p(p+1)}} \int_{\mathrm{Re}(Z)=X_0>0} \mathrm{d}\mathbf{Z} e^{\mathrm{Tr}(\mathbf{Z}\mathbf{\Lambda})} |\mathbf{Z}|^{-a-\frac{1}{2}(p+1)}$$

$$= \frac{|\mathbf{\Lambda}|^a}{\Gamma_p[a + \frac{1}{2}(p+1)]},$$

(B9)

where $\mathbf{\Lambda} > 0$ and the integral is over symmetric imaginary parts and a fixed positive-definite real part. We require

$\mathbf{\Omega} > 0$, i.e., $\nu \geq p$. Since $X_0 = I$ is positive definite, we get, after some rearranging,

$$P_{p,\nu}(\mathbf{\Omega}) = \frac{2^{\frac{p(p-\nu+3)}{2}} \pi^{\frac{p}{2}(p+1)}}{\Gamma_p(\frac{\nu}{2})} |\mathbf{\Omega}|^{\frac{\nu-p-1}{2}} e^{-\frac{1}{2}\mathrm{Tr}(\mathbf{\Omega})}.$$

(B10)

Substituting $P_{p,\nu}(\mathbf{\Omega})$ back in to Eq. (B4) gives

$$p(\{\nu\hat{\mathbf{C}}_l\}) = \prod_{l=l_{\min}}^{l_{\max}} \frac{|\nu\hat{\mathbf{C}}_l|^{(\nu-p-1)/2}}{2^{\frac{p\nu}{2}}\Gamma_p(\frac{\nu}{2})|\mathbf{C}_l|^{\frac{\nu}{2}}} e^{-\frac{\nu}{2}\mathrm{Tr}(\mathbf{C}_l^{-1}\hat{\mathbf{C}}_l)},$$

(B11)

i.e.,

$$p(\{\hat{\mathbf{C}}_l\}) = \prod_{l=l_{\min}}^{l_{\max}} \frac{|\hat{\mathbf{C}}_l|^{(\nu-p-1)/2}}{2^{\frac{p\nu}{2}}\Gamma_p(\frac{\nu}{2})|\mathbf{C}_l/\nu|^{\frac{\nu}{2}}} e^{-\frac{\nu}{2}\mathrm{Tr}(\mathbf{C}_l^{-1}\hat{\mathbf{C}}_l)}$$

$$\equiv \prod_{l=l_{\min}}^{l_{\max}} W_p(\hat{\mathbf{C}}_l; \mathbf{C}_l/\nu, \nu),$$

(B12)

where W_p is the p-dimensional Wishart density.

2. Convergence of the Wishart distribution to Gaussianity

Convergence of probability distributions are most easily studied at the level of the characteristic function. A single $\hat{\mathbf{C}}_l$ has the characteristic function

$$\phi_{\hat{\mathbf{C}}_l}(\mathbf{J}_l) = \left|\mathbf{I} + \frac{2i\mathbf{J}_l\mathbf{C}_l}{\nu}\right|^{-\frac{\nu}{2}}.$$

(B13)

Consider the characteristic function of $\mathbf{M}_l \equiv \mathbf{C}^{-1/2}\hat{\mathbf{C}}_l\mathbf{C}_l^{-1/2}$. This is

$$\phi_{\mathbf{M}_l}(\mathbf{J}_l) = \left|\mathbf{I} + \frac{2i\mathbf{J}_l}{\nu}\right|^{-\frac{\nu}{2}}.$$

(B14)

We can rewrite this as

$$\phi_{\mathbf{M}_l}(\mathbf{J}_l) = \left|\mathbf{I} + \frac{2i\mathbf{R}\mathbf{J}_l\mathbf{R}^T}{\nu}\right|^{-\frac{\nu}{2}},$$

(B15)

where $\mathbf{R}$ is any orthogonal matrix. Choosing $\mathbf{R}$ as the matrix that diagonalizes $\mathbf{J}_l$, this gives

$$\phi_{\mathbf{M}_l}(\mathbf{J}_l) = \prod_{i=1}^{p} \left[1 + \frac{2i\lambda_i}{\nu}\right]^{-\frac{\nu}{2}},$$

(B16)

where λ_i are the eigenvalues of $\mathbf{J}_l$.

Now fixing $\mathbf{J}_l$, taking $\nu \to \infty$, and keeping terms up to $\mathcal{O}(\nu^{-1})$ gives

$$\lim_{\nu\to\infty}\phi_{\mathbf{M}_l}(\mathbf{J}_l) = \exp\left(-i\sum_{i=1}^p \lambda_i - \sum_{i=1}^p \frac{\lambda_i^2}{\nu}\right)$$

$$= \exp\left[-i\mathrm{Tr}(\mathbf{J}_l) - \frac{1}{\nu}\mathrm{Tr}(\mathbf{J}_l\mathbf{J}_l)\right]. \quad \text{(B17)}$$

This implies that the characteristic function of our original variable is

$$\lim_{\nu\to\infty}\phi_{\hat{\mathbf{C}}_l}(\mathbf{J}_l) = \exp\left[-i\mathrm{Tr}(\mathbf{J}_l\mathbf{C}_l) - \frac{1}{\nu}\mathrm{Tr}(\mathbf{J}_l\mathbf{C}_l\mathbf{J}_l\mathbf{C}_l)\right]. \quad \text{(B18)}$$

The inverse Fourier transform is a Gaussian integral, giving

$$\lim_{\nu\to\infty}p(\hat{\mathbf{C}}_l) = 2^{\frac{p(p-1)}{2}}\int\frac{\mathrm{d}\mathbf{J}_l}{(2\pi)^{\frac{p(p+1)}{2}}}e^{i\mathrm{Tr}[\mathbf{J}_l(\hat{\mathbf{C}}_l-\mathbf{C}_l)]}e^{-\frac{1}{\nu}\mathrm{Tr}(\mathbf{J}_l\mathbf{C}_l\mathbf{J}_l\mathbf{C}_l)}. \quad \text{(B19)}$$

To do this integral, it is advantageous to move to vectorized notation [70]. For a general $p\times p$ matrix Y define $\mathrm{vec}(Y)$ as the p^2-dimensional vector whose elements are $(Y_{1,1},\ldots,Y_{p,1},Y_{1,2},\ldots,Y_{p,2},\ldots)$. For a symmetric $p\times p$ matrix X, define $\mathrm{vecp}(X)$ as the $\frac{1}{2}p(p+1)$-dimensional vector whose elements are $(X_{1,1},X_{1,2},\ldots,X_{1,p},X_{2,2},X_{2,3},\ldots)$. Then some standard matrix identities give

$$\mathrm{Tr}(\mathbf{XDXE}) = \mathrm{vec}^T(\mathbf{X})(\mathbf{E}\otimes\mathbf{D})\mathrm{vec}(\mathbf{X})$$
$$= \mathrm{vecp}^T(\mathbf{X})\mathbf{B}_p^+(\mathbf{E}\otimes\mathbf{D})\mathbf{B}_p^{+T}\mathrm{vecp}(\mathbf{X}), \quad \text{(B20)}$$

where $\otimes$ is the Kronecker product, and $\mathbf{B}_p^+ = (\mathbf{B}_p^T\mathbf{B}_p)^{-1}\mathbf{B}_p^T$ is the Moore-Penrose inverse of $\mathbf{B}_p$, the $p^2\times\frac{1}{2}p(p+1)$ transition matrix. This matrix satisfies $\mathrm{vec}(\mathbf{X}) = \mathbf{B}_p^{+T}\mathrm{vecp}(\mathbf{X})$ and $\mathrm{vec}(\mathbf{X}) = \mathbf{B}_p^T\mathrm{vecp}(\mathbf{X})$ and has typical elements given by [70]

$$(\mathbf{B}_p)_{ij,gh} = \frac{1}{2}(\delta_{ig}\delta_{jh} + \delta_{ih}\delta_{jg}), i\le p, j\le p, g\le h\le p. \quad \text{(B21)}$$

Note that $\mathbf{B}_p^+$ is a $\frac{1}{2}p(p+1)\times p^2$ matrix. We also have

$$\mathrm{Tr}(\mathbf{XA}) = \mathrm{vec}^T(\mathbf{X})\mathrm{vec}(\mathbf{A})$$
$$= \mathrm{vecp}^T(\mathbf{X})\mathbf{B}_p^+\mathbf{B}_p^{+T}\mathrm{vecp}(\mathbf{A}). \quad \text{(B22)}$$

Making these substitutions, defining $\Delta\hat{\mathbf{C}}_l = \hat{\mathbf{C}}_l - \mathbf{C}_l$ and using that $|\mathbf{B}_p^T\mathbf{B}_p| = 2^{-\frac{1}{2}p(p-1)}$ gives

$$\lim_{\nu\to\infty}p(\hat{\mathbf{C}}_l) = 2^{\frac{p(p-1)}{2}}\int\frac{\mathrm{dvecp}(\mathbf{J}_l)}{(2\pi)^{\frac{p(p+1)}{2}}}e^{i\mathrm{vecp}^T(\mathbf{J}_l)\mathbf{B}_p^+\mathbf{B}_p^{+T}\mathrm{vecp}(\Delta\hat{\mathbf{C}}_l)}e^{-\frac{1}{\nu}\mathrm{vecp}^T(\mathbf{J}_l)\mathbf{B}_p^+(\mathbf{C}_l\otimes\mathbf{C}_l)\mathbf{B}_p^{+T}\mathrm{vecp}(\mathbf{J}_l)}$$
$$= \frac{\exp\left\{-\frac{1}{2}\mathrm{vecp}^T(\Delta\hat{\mathbf{C}}_l)[\frac{2}{\nu}\mathbf{B}_p^T(\mathbf{C}_l\otimes\mathbf{C}_l)\mathbf{B}_p]^{-1}\mathrm{vecp}(\Delta\hat{\mathbf{C}}_l)\right\}}{\sqrt{(2\pi)^{\frac{1}{2}p(p+1)}|\frac{2}{\nu}\mathbf{B}_p^T(\mathbf{C}_l\otimes\mathbf{C}_l)\mathbf{B}_p|}}, \quad \text{(B23)}$$

i.e., $\mathrm{vecp}(\hat{\mathbf{C}}_l)$ is $\frac{1}{2}p(p+1)$-dimensional Gaussian with mean $\mathrm{vecp}(\mathbf{C}_l)$ and covariance matrix $\frac{2}{\nu}\mathbf{B}_p^T(\mathbf{C}_l\otimes\mathbf{C}_l)\mathbf{B}_p$. The matrix $\hat{\mathbf{C}}_l$ is thus symmetric matrix normal. Note that the mean and covariance matrix do not change compared with their Wishart forms when the limit is taken.

We have shown that the characteristic function of the Wishart distribution tends pointwise to that of a symmetric matrix Gaussian in the limit $\nu\to\infty$. This is enough to establish that Wishart matrices converge in distribution to symmetric Gaussian matrices, by Lévy's continuity theorem.

3. Signal with nonzero trispectrum

We write the signal characteristic function as

$$\phi_s(\{\mathbf{k}_{lm}\}) = \left(1 + \frac{1}{24}\kappa_{\underline{lm}}^{ijkm}k_{l_1m_1}^i k_{l_2m_2}^j k_{l_3m_3}^k k_{l_4m_4}^m\right)\phi_s^G(\{\mathbf{k}_{lm}\}). \quad \text{(B24)}$$

Substituting this into Eq. (9) gives a Gaussian integral, as in the Gaussian signal case, but now with four factors of the wave number in the integrand. This integral can be done analytically, which gives the power spectrum characteristic function as

$$\phi_{\{\nu\hat{\mathbf{C}}_l\}}(\{\mathbf{J}_l\}) = \left(\prod_{l=l_{\min}}^{l_{\max}}|\mathbf{I} + 2i\mathbf{J}_l\mathbf{C}_l|^{-\frac{\nu}{2}}\right)\left[1 + \frac{1}{8}\sum_{l_1,l_2}(2l_1+1)(2l_2+1)\langle\Delta\hat{C}_{l_1}^{ij}\Delta\hat{C}_{l_2}^{km}\rangle_{\mathrm{NG}}M_{l_1}^{ij}M_{l_2}^{km}\right]. \quad \text{(B25)}$$

A few consequences of the corrected distribution are immediately evident from Eq. (B25). First, the leading-order correction in the limit $J_l \to 0$ is quadratic in J_l, which tells us that both the normalization and the mean of $p(\hat{C}_l)$ are unaffected by our correction (the latter is true by

construction), but the covariance does receive a correction, given by $\langle \Delta \hat{C}_{l_1}^{ij} \Delta \hat{C}_{l_2}^{km} \rangle_{\mathrm{NG}}$. This is as expected, agreeing with the nonperturbative expression (A1).

Inverse Fourier transforming Eq. (B25) gives the correction to the probability density as

$$\Delta p(\{\nu \hat{\mathbf{C}}_l\}) = \frac{1}{8} 2^{\frac{\lambda p(p-1)}{2}} \int \frac{d\{\mathbf{J}_l\}}{(2\pi)^{n\lambda}} e^{i \sum_l \mathrm{Tr}(\nu \mathbf{J}_l \hat{\mathbf{C}}_l)} \left(\prod_{l=l_{\min}}^{l_{\max}} |\mathbf{I} + 2i\mathbf{J}_l \mathbf{C}_l|^{-\frac{k}{2}} \right) \sum_{l_1,l_2} \nu_1 \nu_2 \langle \Delta \hat{C}_{l_1}^{ij} \Delta \hat{C}_{l_2}^{km} \rangle_{\mathrm{NG}} (\mathbf{M}_{l_1})^{ij} (\mathbf{M}_{l_2})^{km}, \qquad (B26)$$

with $\mathbf{M}_l = [\mathbf{C}_l + (2i\mathbf{J}_l)^{-1}]^{-1}$ and $\lambda \equiv l_{\max} - l_{\min} + 1$. The quantity $\langle \Delta \hat{C}_{l_1}^{ij} \Delta \hat{C}_{l_2}^{km} \rangle_{\mathrm{NG}}$ is the non-Gaussian contribution to the covariance matrix. In the $p = 1$ case we call this $T_{l_1 l_2}$. The crucial effect of non-Gaussianity here is to couple $\hat{C}_l$ with different l, a result of nonzero off-diagonal terms in the non-Gaussian covariance matrix; the probability density no longer factorizes in l.

The integrals in Eq. (B26) are of the Wishart type when $l_1 \neq l_2$, and the only new term is that with $l_1 = l_2$. After some rearranging and relabeling, one can show that it is necessary to compute the integral

$$H_{p,\nu}^{ab,cd}(\mathbf{\Omega}) \equiv \int d\mathbf{T} e^{\frac{i}{2}\mathrm{Tr}(\mathbf{\Omega}\mathbf{T})} |\mathbf{I} + i\mathbf{T}|^{-\frac{k}{2}} [(\mathbf{I} + i\mathbf{T})^{-1}]_{ab} [(\mathbf{I} + i\mathbf{T})^{-1}]_{cd}. \qquad (B27)$$

We will evaluate this using derivatives, but first we need to establish a few rules for taking derivatives with respect to symmetric matrices. These can be found in Ref. [71].

For a symmetric matrix $\mathbf{X}$, we have

$$\frac{\partial \mathbf{X}}{\partial X_{ij}} = \mathbf{J}^{ij} + \mathbf{J}^{ji} - \mathbf{J}^{ij}\mathbf{J}^{ij}, \qquad (B28)$$

$$\frac{\partial \mathrm{Tr}(\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} = \mathbf{A} + \mathbf{A}^T - (\mathbf{A} \circ \mathbf{I}), \qquad (B29)$$

$$\frac{\partial \det \mathbf{X}}{\partial \mathbf{X}} = (\det \mathbf{X})[2\mathbf{X}^{-1} - (\mathbf{X}^{-1} \circ \mathbf{I})], \qquad (B30)$$

where $\circ$ denotes the Hadamard (elementwise) product, and $\mathbf{J}^{ij}$ is the matrix whose km element is $\delta_{ik}\delta_{jm}$, i.e., the matrix whose only nonzero values are in cell ij.

We can therefore write the following identity holding for symmetric matrices $\mathbf{T}$:

$$|\mathbf{I} + i\mathbf{T}|^{-\frac{k}{2}} (\mathbf{I} + i\mathbf{T})_{ij}^{-1} = \frac{i}{\nu} \frac{\partial}{\partial T_{ij}} |\mathbf{I} + i\mathbf{T}|^{-\frac{k}{2}} + \frac{1}{2} (\mathbf{I} + i\mathbf{T})_{ij}^{-1} \delta_{ij}, \qquad (B31)$$

with no summation over repeated indices. Using this identity one can establish the identity

$$|\mathbf{I} + i\mathbf{T}|^{-\frac{k}{2}} (\mathbf{I} + i\mathbf{T})_{ij}^{-1} (\mathbf{I} + i\mathbf{T})_{km}^{-1} = +\frac{i}{\nu} \frac{\partial}{\partial T_{ij}} [|\mathbf{I} + i\mathbf{T}|^{-\frac{k}{2}} (\mathbf{I} + i\mathbf{T})_{km}^{-1}] + \frac{1}{2} |\mathbf{I} + i\mathbf{T}|^{-\frac{k}{2}} (\mathbf{I} + i\mathbf{T})_{ij}^{-1} \delta_{ij} (\mathbf{I} + i\mathbf{T})_{km}^{-1}$$
$$-\frac{1}{\nu} |\mathbf{I} + i\mathbf{T}|^{-\frac{k}{2}} [2(\mathbf{I} + i\mathbf{T})_{i(k}^{-1} (\mathbf{I} + i\mathbf{T})_{m)j}^{-1} - \delta_{ij} (\mathbf{I} + i\mathbf{T})_{kj}^{-1} (\mathbf{I} + i\mathbf{T})_{mi}^{-1}], \qquad (B32)$$

where again there is no summation over repeated indices, and parentheses denote symmetrization over enclosed indices. Substituting this into Eq. (B27), integrating by parts, and noting that the boundary term vanishes (which it must do for the Fourier transforms to converge) gives

$$H_{p,\nu}^{ijkm}(\mathbf{\Omega}) = \frac{1}{2\nu} (2\mathbf{\Omega} - \mathbf{\Omega} \circ \mathbf{I})_{ij} P_{p,\nu}(\mathbf{\Omega}) \frac{\Omega_{km}}{\nu} + \frac{1}{2} \delta_{ij} H_{p,\nu}^{ijkm}(\mathbf{\Omega}) - \frac{2}{\nu} H_{p,\nu}^{i(km)j}(\mathbf{\Omega}) + \frac{1}{\nu} \delta_{ij} H_{p,\nu}^{kjmi}(\mathbf{\Omega}). \qquad (B33)$$

Permuting indices we can get three simultaneous equations for $H_{p,\nu}^{ijkm}$. The solution is

$$H_{p,\nu}^{ijkm}(\mathbf{\Omega}) = \frac{(\nu+1)}{\nu(\nu+2)(\nu-1)}\left[\Omega_{ij}\Omega_{km} - \frac{2}{(\nu+1)}\Omega_{i(k}\Omega_{m)j}\right]P_{p,\nu}(\mathbf{\Omega}). \tag{B34}$$

Putting everything together gives the corrected density for the $\hat{\mathbf{C}}_l$ as

$$p(\{\hat{\mathbf{C}}_l\}) = \left[\prod_{l=l_{\min}}^{l_{\max}} W_p(\hat{\mathbf{C}}_l; \mathbf{C}_l/\nu, \nu)\right]\left\{1 + \frac{1}{8}\sum_{l_1,l_2}\nu_1\nu_2\langle\Delta\hat{C}_{l_1}^{ab}\Delta\hat{C}_{l_2}^{cd}\rangle_{\mathrm{NG}}C_{l_1,ai}^{-1}C_{l_1,bj}^{-1}C_{l_2,ck}^{-1}C_{l_2,dm}^{-1}\right.$$
$$\left. \times\left[\Delta\hat{C}_{l_1,ij}\Delta\hat{C}_{l_2,km} + \frac{2\delta_{l_1 l_2}}{(\nu+2)(\nu-1)}(\hat{C}_{l_1,ij}\hat{C}_{l_1,km} - \nu_1\hat{C}_{l_1,i(k}\hat{C}_{l_1,m)j})\right]\right\}, \tag{B35}$$

with implicit summation over repeated indices. One can verify that the distribution (B35) is correctly normalized and has a mean given by $\langle\hat{\mathbf{C}}_l\rangle = \mathbf{C}_l$, i.e., unchanged from the Wishart case.

We can write the correction in a more compact vectorial notation as follows. We define the Gaussian covariance matrix

$$\mathbf{\Sigma}_l \equiv \langle\mathrm{vecp}(\Delta\hat{\mathbf{C}}_l)\mathrm{vecp}^T(\Delta\hat{\mathbf{C}}_l)\rangle_G = \frac{2}{\nu}\mathbf{B}_p^T(\mathbf{C}_l \otimes \mathbf{C}_l)\mathbf{B}_p, \tag{B36}$$

which implies that $\mathbf{\Sigma}_l^{-1} = \frac{\nu}{2}\mathbf{B}_p^+(\mathbf{C}_l^{-1} \otimes \mathbf{C}_l^{-1})\mathbf{B}_l^{+T}$. After some manipulations, the term quadratic in $\Delta\hat{C}_l$ in Eq. (17) can be written as

$$\frac{1}{2}\sum_{l_1,l_2}\mathrm{vecp}^T(\Delta\hat{\mathbf{C}}_{l_1})\mathbf{\Sigma}_{l_1}^{-1}\langle\mathrm{vecp}(\Delta\hat{\mathbf{C}}_{l_1})\mathrm{vecp}(\Delta\hat{\mathbf{C}}_{l_2})\rangle_{\mathrm{NG}}\mathbf{\Sigma}_{l_2}^{-1}\mathrm{vecp}^T(\Delta\hat{\mathbf{C}}_{l_2}), \tag{B37}$$

which is the next-to-leading-order term in the expansion of

$$-\frac{1}{2}\sum_{l_1,l_2}\mathrm{vecp}^T(\Delta\hat{\mathbf{C}}_{l_1})[\mathbf{\Sigma}_{l_1}\delta_{l_1 l_2} + \langle\mathrm{vecp}(\Delta\hat{\mathbf{C}}_{l_1})\mathrm{vecp}^T(\Delta\hat{\mathbf{C}}_{l_2})\rangle_{\mathrm{NG}}]^{-1}\mathrm{vecp}^T(\Delta\hat{\mathbf{C}}_{l_2}), \tag{B38}$$

i.e., an expansion of a chi-squared having the inverse of the total covariance matrix. Once the Gaussian part of this expansion is available when the $\nu \to \infty$ limit is taken, we will be left a Gaussian possessing the correct inverse covariance matrix in its exponent.

The first of the two terms quadratic in $\hat{C}_l$ in Eq. (B35) has the same form as the term above and can be written as a quadratic in $\mathrm{vecp}(\hat{\mathbf{C}}_l)$. Note that this term is suppressed with respect to the second term quadratic in $\hat{C}_l$ by a factor of ν. This second term can be written as, with $\mathbf{A}_l \equiv \mathbf{C}_l^{-1}\hat{\mathbf{C}}_l\mathbf{C}_l^{-1}$,

$$-\frac{1}{4}\sum_{l_1}\frac{\nu_1^3}{(\nu_1+2)(\nu_1-1)}\mathrm{Tr}[(\mathbf{A}_{l_1} \otimes \mathbf{A}_{l_1})\langle\mathrm{vec}(\Delta\hat{\mathbf{C}}_{l_1})\mathrm{vec}^T(\Delta\hat{\mathbf{C}}_{l_1})\rangle_{\mathrm{NG}}]$$
$$= -\frac{1}{4}\sum_{l_1}\frac{\nu_1^3}{(\nu_1+2)(\nu_1-1)}\mathrm{Tr}[\mathbf{B}_p^+(\mathbf{A}_{l_1} \otimes \mathbf{A}_{l_1})\mathbf{B}_p^{+T}\langle\mathrm{vecp}(\Delta\hat{\mathbf{C}}_{l_1})\mathrm{vecp}^T(\Delta\hat{\mathbf{C}}_{l_1})\rangle_{\mathrm{NG}}]. \tag{B39}$$

This looks like the next-to-leading-order term in an expansion of the determinant of the total covariance, with the exception that it depends on the measured power. In the limit $\nu_1 \to \infty$ we can write $\hat{\mathbf{C}}_l = \mathbf{C}_l + \mathcal{O}(\nu^{-1})$, so we set $\hat{\mathbf{C}}_l = \mathbf{C}_l$ in this term and hence $\mathbf{A}_l = \mathbf{C}_l^{-1}$. In that limit, this term dominates over the first term quadratic in $\hat{C}_l$ and is precisely the expansion of the determinant of the total covariance, i.e., the prefactor in the asymptotic Gaussian density.

4. Signal with nonzero bispectrum

We now consider a signal having nonzero bispectrum. The signal characteristic function reads

$$\phi_s(\{\mathbf{k}_{lm}\}) = \left(1 - \frac{1}{72}\kappa_{\underline{lm}}^{ijk}\kappa_{\underline{l'm'}}^{lmn}k_{l_1 m_1}^i k_{l_2 m_2}^j k_{l_3 m_3}^k k_{l'_1 m'_1}^l k_{l'_2 m'_2}^m k_{l'_3 m'_3}^n\right)\phi_s^G(\{\mathbf{k}_{lm}\}). \tag{B40}$$

Following the same steps as the trispectrum case, the correction to the characteristic function is

$$\Delta\phi_{\{\nu\hat{\mathbf{C}}_l\}}(\{\mathbf{J}_l\}) = \left(\prod_{l=l_{\min}}^{l_{\max}} |\mathbf{I} + 2i\mathbf{J}_l\mathbf{C}_l|^{-\frac{\nu}{2}}\right)\left[1 - \frac{1}{12}\sum_{l_1,l_2,l_3} \nu_1\nu_2\nu_3 \tilde{B}_{l_1 l_2 l_3}^{ijk,lmn}\right.$$
$$\left.\times \left(\mathbf{C}_{l_1} - \frac{i}{2}\mathbf{J}_{l_1}^{-1}\right)_{il}^{-1}\left(\mathbf{C}_{l_2} - \frac{i}{2}\mathbf{J}_{l_2}^{-1}\right)_{jm}^{-1}\left(\mathbf{C}_{l_3} - \frac{i}{2}\mathbf{J}_{l_3}^{-1}\right)_{kn}^{-1}\right]. \tag{B41}$$

In the limit that $J_l \to 0$ this is a cubic correction to the Wishart characteristic function, telling us that the effect of $\tilde{B}$ is to modify the three-point function or skewness of $\hat{C}_l$, as expected. Just as in the trispectrum case, the correction is not a polynomial in J_l and hence not an Edgeworth expansion, even though it is perturbatively "close" to the zero-order Wishart distribution.[11]

Inverse Fourier transforming Eq. (B41) gives the correction to the density as

$$\Delta p(\{\nu\hat{\mathbf{C}}_l\}) = -\frac{1}{12} 2^{\frac{\lambda p(p-1)}{2}} \int \frac{\mathrm{d}\{\mathbf{J}_l\}}{(2\pi)^{n\lambda}} e^{i\sum_l \mathrm{Tr}(\nu\mathbf{J}_l\hat{\mathbf{C}}_l)} \left(\prod_{l=l_{\min}}^{l_{\max}} |\mathbf{I} + 2i\mathbf{J}_l\mathbf{C}_l|^{-\frac{\nu}{2}}\right) \sum_{l_1,l_2,l_3} \nu_1\nu_2\nu_3 \tilde{B}_{l_1 l_2 l_3}^{ijk,lmn}(\mathbf{M}_{l_1})^{il}(\mathbf{M}_{l_2})^{jm}(\mathbf{M}_{l_3})^{kn}, \tag{B42}$$

with $\mathbf{M}_l = [\mathbf{C}_l + (2i\mathbf{J}_l)^{-1}]^{-1}$.

Once again the only new term is the case $l_1 = l_2 = l_3$. The new integral we need is

$$K_{p,\nu}^{ab,cd,ef}(\mathbf{\Omega}) \equiv \int \mathrm{d}\mathbf{T} e^{\frac{i}{2}\mathrm{Tr}(\mathbf{\Omega}\mathbf{T})} |\mathbf{I} + i\mathbf{T}|^{-\frac{\nu}{2}}[(\mathbf{I} + i\mathbf{T})^{-1}]_{ab}[(\mathbf{I} + i\mathbf{T})^{-1}]_{cd}[(\mathbf{I} + i\mathbf{T})^{-1}]_{ef}. \tag{B43}$$

Our strategy will again be to take derivatives. Letting $\mathbf{A} \equiv (\mathbf{I} + i\mathbf{T})^{-1}$, we have the identity

$$|\mathbf{A}|^{\frac{\nu}{2}}A_{ab}A_{ij}A_{km} = \frac{i}{\nu}\frac{\partial}{\partial T_{ij}}(|\mathbf{A}|^{\frac{\nu}{2}}A_{km}A_{ab}) + \frac{1}{2}\delta_{ij}|\mathbf{A}|^{\frac{\nu}{2}}A_{ab}A_{ij}A_{km} - \frac{2}{\nu}|\mathbf{A}|^{\frac{\nu}{2}}A_{ab}A_{k(i}A_{j)m}$$
$$+ \frac{1}{\nu}\delta_{ij}|\mathbf{A}|^{\frac{\nu}{2}}A_{ab}A_{ki}A_{mj} - \frac{2}{\nu}|\mathbf{A}|^{\frac{\nu}{2}}A_{km}A_{a(i}A_{j)b} + \frac{1}{\nu}\delta_{ij}|\mathbf{A}|^{\frac{\nu}{2}}A_{km}A_{ai}A_{jb}, \tag{B44}$$

with no summation over repeated indices.

Integrating by parts, substituting Eq. (B34), and defining $K_{p,\nu}^{ab,ij,km} \equiv M^{ab,ij,km}P_{p,\nu}$ [where $P_{p,\nu}$ is defined in Eq. (B10)] gives

$$M^{ab,ij,km} - \frac{1}{2}\delta_{ij}M^{ab,ij,km} + \frac{2}{\nu}M^{ab,k(i,j)m} - \frac{1}{\nu}\delta_{ij}M^{ab,ki,mj} + \frac{2}{\nu}M^{km,a(i,j)b} - \frac{1}{\nu}\delta_{ij}M^{km,ai,jb}$$
$$= \frac{(\nu+1)}{\nu^2(\nu+2)(\nu-1)}\left(\Omega_{ij} - \frac{1}{2}\Omega_{ij}\delta_{ij}\right)\left(\Omega_{km}\Omega_{ab} - \frac{2}{\nu+1}\Omega_{k(a}\Omega_{b)m}\right). \tag{B45}$$

Multiplying by δ_{ij} and subtracting the result gives

$$M^{ab,ij,km} + \frac{2}{\nu}M^{ab,k(i,j)m} + \frac{2}{\nu}M^{km,a(i,j)b} = \frac{(\nu+1)}{\nu^2(\nu+2)(\nu-1)}\left(\Omega_{ij}\Omega_{km}\Omega_{ab} - \frac{2}{\nu+1}\Omega_{ij}\Omega_{k(a}\Omega_{b)m}\right). \tag{B46}$$

We can write two further equations like Eq. (B46) by taking derivatives with respect to T_{km} and T_{ab}. Adding all three equations gives an expression manifestly symmetric in its indices

---

[11]Note that to get the characteristic function of $\hat{C}_l$ rather than $\nu\hat{C}_l$ one has to replace J_l with J_l/ν.

$$3M^{ab,ij,km} + \frac{2}{\nu}(M^{ab,ki,jm} + M^{ab,kj,im} + M^{km,aj,ib} + M^{km,ai,jb} + M^{ij,ak,mb} + M^{ij,am,kb})$$

$$= \frac{(\nu+1)}{\nu^2(\nu+2)(\nu-1)} \left[3\Omega_{ij}\Omega_{km}\Omega_{ab} - \frac{1}{1+\nu}(\Omega_{ij}\Omega_{ka}\Omega_{bm} + \Omega_{ij}\Omega_{kb}\Omega_{am} + \Omega_{km}\Omega_{ia}\Omega_{bj} \right.$$

$$\left. + \Omega_{km}\Omega_{ib}\Omega_{aj} + \Omega_{ab}\Omega_{ki}\Omega_{jm} + \Omega_{ab}\Omega_{kj}\Omega_{im}) \right]. \tag{B47}$$

From the $p = 1$ case we know that the solution must be cubic in the elements of $\boldsymbol{\Omega}$. There are three kinds of terms that we can use to build $M^{ab,ij,km}$ dictated by the symmetries in play, given by

$$M^{ab,ij,km} = X\Omega_{ij}\Omega_{km}\Omega_{ab} + 2Y[\Omega_{ij}\Omega_{k(a}\Omega_{b)m} + \Omega_{km}\Omega_{i(a}\Omega_{b)j} + \Omega_{ab}\Omega_{k(i}\Omega_{j)m}] + 2Z[\Omega_{jm}\Omega_{k(a}\Omega_{b)i} + \Omega_{ik}\Omega_{j(a}\Omega_{b)m}$$

$$+ \Omega_{jk}\Omega_{i(a}\Omega_{b)m} + \Omega_{im}\Omega_{k(a}\Omega_{b)j}], \tag{B48}$$

for unknown coefficients X, Y, Z. The three terms here are closed under the symmetries of $M^{ab,ij,km}$. Substituting into Eq. (B47) and identifying coefficients of these terms gives the following equations for the unknown coefficients:

$$3X + \frac{12}{\nu}Y = \frac{3(\nu+1)}{\nu^2(\nu+2)(\nu-1)}, \tag{B49}$$

$$6Y + \frac{2}{\nu}(2X + 2Y + 8Z) = -\frac{2}{\nu^2(\nu+2)(\nu-1)}, \tag{B50}$$

$$6Z + \frac{2}{\nu}(6Y + 6Z) = 0. \tag{B51}$$

The solution is

$$X = \frac{\nu^2 + 3\nu - 2}{\nu(\nu-1)(\nu-2)(\nu+2)(\nu+4)}, \tag{B52}$$

$$Y = \frac{-1}{\nu(\nu-1)(\nu-2)(\nu+4)}, \tag{B53}$$

$$Z = \frac{2}{\nu(\nu-1)(\nu-2)(\nu+2)(\nu+4)}. \tag{B54}$$

The integral (B43) is thus

$$K_{p,nu}^{ab,cd,ef}(\boldsymbol{\Omega}) = \frac{P_{p,\nu}(\boldsymbol{\Omega})}{\nu(\nu+4)(\nu-2)(\nu-1)} \left\{ \frac{\nu^2 + 3\nu - 2}{\nu+2}\Omega_{ab}\Omega_{cd}\Omega_{ef} - 2[\Omega_{cd}\Omega_{e(a}\Omega_{b)f} \right.$$

$$+ \Omega_{ef}\Omega_{c(a}\Omega_{b)d} + \Omega_{ab}\Omega_{e(c}\Omega_{d)f}] + \frac{4}{\nu+2}[\Omega_{df}\Omega_{e(a}\Omega_{b)c}$$

$$\left. + \Omega_{ce}\Omega_{d(a}\Omega_{b)f} + \Omega_{de}\Omega_{c(a}\Omega_{b)f} + \Omega_{cf}\Omega_{e(a}\Omega_{b)d}] \right\}. \tag{B55}$$

Putting everything together gives the correction to the density as

$$
p(\{\hat{\mathbf{C}}_l\}) = \left[\prod_{l=l_{\min}}^{l_{\max}} W_p(\hat{\mathbf{C}}_l; \mathbf{C}_l/\nu, \nu)\right]\left\{1 + \frac{1}{12}\sum_{l_1,l_2,l_3}\nu_1\nu_2\nu_3 \tilde{B}_{l_1 l_2 l_3}^{abc,def} C_{l_1,ai}^{-1} C_{l_2,bj}^{-1} C_{l_3,ck}^{-1} C_{l_1,dl}^{-1} C_{l_2,em}^{-1} C_{l_3,fn}^{-1}\right.
$$

$$
\times \left[\Delta\hat{C}_{l_1,il}\Delta\hat{C}_{l_2,jm}\Delta\hat{C}_{l_3,kn} + [3]\frac{2\delta_{l_1 l_2}}{(\nu_1+2)(\nu_1-1)}\hat{C}_{l_1,il}\hat{C}_{l_1,jm}\Delta\hat{C}_{l_3,kn} - [3]\frac{2\delta_{l_1 l_2}\nu_1}{(\nu_1+2)(\nu_1-1)}\hat{C}_{l_1,i(j}\hat{C}_{l_1,m)l}\Delta\hat{C}_{l_3,kn}\right.
$$

$$
+ \frac{32\delta_{l_1 l_2}\delta_{l_2 l_3}}{(\nu_1+2)(\nu_1-1)(\nu_1+4)(\nu_1-2)}\hat{C}_{l_1,il}\hat{C}_{l_1,jm}\hat{C}_{l_1,kn} - [3]\frac{16\nu_1\delta_{l_1 l_2}\delta_{l_2 l_3}}{(\nu_1+2)(\nu_1-1)(\nu_1+4)(\nu_1-2)}\hat{C}_{l_1,jm}\hat{C}_{l_1,k(i}\hat{C}_{l_1,l)n}
$$

$$
\left.\left.+ [4]\frac{4\nu_1^2\delta_{l_1 l_2}\delta_{l_2 l_3}}{(\nu_1+2)(\nu_1-1)(\nu_1+4)(\nu_1-2)}\hat{C}_{l_1,mn}\hat{C}_{l_1,k(i}\hat{C}_{l_1,l)j}\right]\right\}, \tag{B56}
$$

with implicit summation over repeated indices. We have verified that this distribution is correctly normalized. Computing the corrections to the mean, covariance, and three-point function is tedious so we have only checked for the $p = 1$ case, confirming that these three cumulants are recovered correctly.

Equations (B35) and (B56) give the leading-order correction to the Wishart distribution from signal non-Gaussian. The expressions are clearly quite cumbersome when $p \neq 1$, but may be simplified using the symmetries of the bin indices. It may also be possible to write these expressions in a more covariant way, although we were not able to find a more compact expression.

APPENDIX C: ALTERNATIVE BINNING CHOICES

The default results of Sec. III B are made with evenly spaced logarithmically binned power spectra with $\Delta \log l = 0.1$ and a weight function $W_l = 2l + 1$. The mathematical details of the binning are described in Sec. II D.

The main use of binning for our purposes is to allow us to easily measure the covariance matrix of our log-normal maps from a finite number of simulations. Binning reduces the Monte Carlo noise in this measurement and increases the relative strength of the non-Gaussian part.

As discussed in Sec. II D, the correction to posteriors from signal non-Gaussianity in expected to depend only mildly on the bin width, as we verify numerically in Fig. 9 where we plot the corrections to C_l and amplitude posteriors for logarithmic bins with $\Delta \log l = 0.2$. This plot should be compared with Fig. 7. The impact of these broader bins is very mild. The broader bins roughly preserve the total mode count when $W_l = 2l + 1$, so the effect on the posterior is naturally expected to be small.

It is beyond the scope of this paper to provide an exhaustive investigation of the different binning choices for the power spectrum, but a few obvious schemes deserve discussion. Instead of logarithmic bins we can also choose linearly spaced bins. This choice compresses many of the large-scale modes into a single bin and hence does not allow the low-l behavior of the likelihood to be clearly
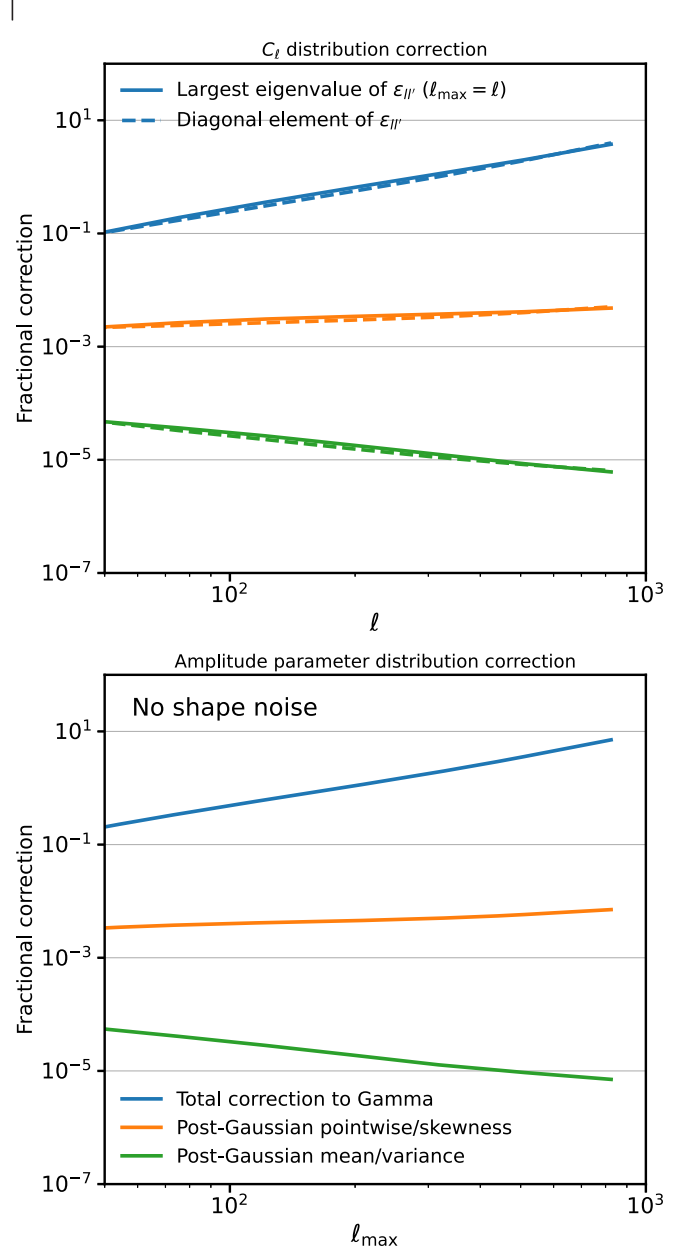

FIG. 9. The leading-order fractional corrections to the Gamma posterior with band powers binned with $\Delta \log l = 0.2$. Curves have the same meaning as in Fig. 7.

seen, so we chose not to adopt this choice in our baseline results. Another alternative is to use a weight function $W_l = l(l+1)$, which is a common choice for power spectrum band powers. This upweights the smallest scales in each bin and consequently causes our likelihood approximation to break down at smaller l_b (the bin barycenter) for logarithmic bins with $\Delta l = 0.1$. An advantage of our formalism is that the likelihood correction diagnostic $\langle \varepsilon \rangle$ can be computed for any binning choice, something which we advocate when analyzing power spectra binned with this weight function.

We close by noting that any binning choice may be adopted so long as the bins are sufficiently small that the power spectrum can be approximated as flat over the extent of the bin. This assumption can be avoided entirely by working with $\hat{C}_l/C_l$ rather than C_l, as discussed in Sec. II D.

APPENDIX D: DEPENDENCE ON SOURCE REDSHIFT

The baseline results in this paper have assumed a broad source redshift bin following a Euclid-like distribution, peaking at $z_s \approx 1$. Lensing surveys typically place galaxies
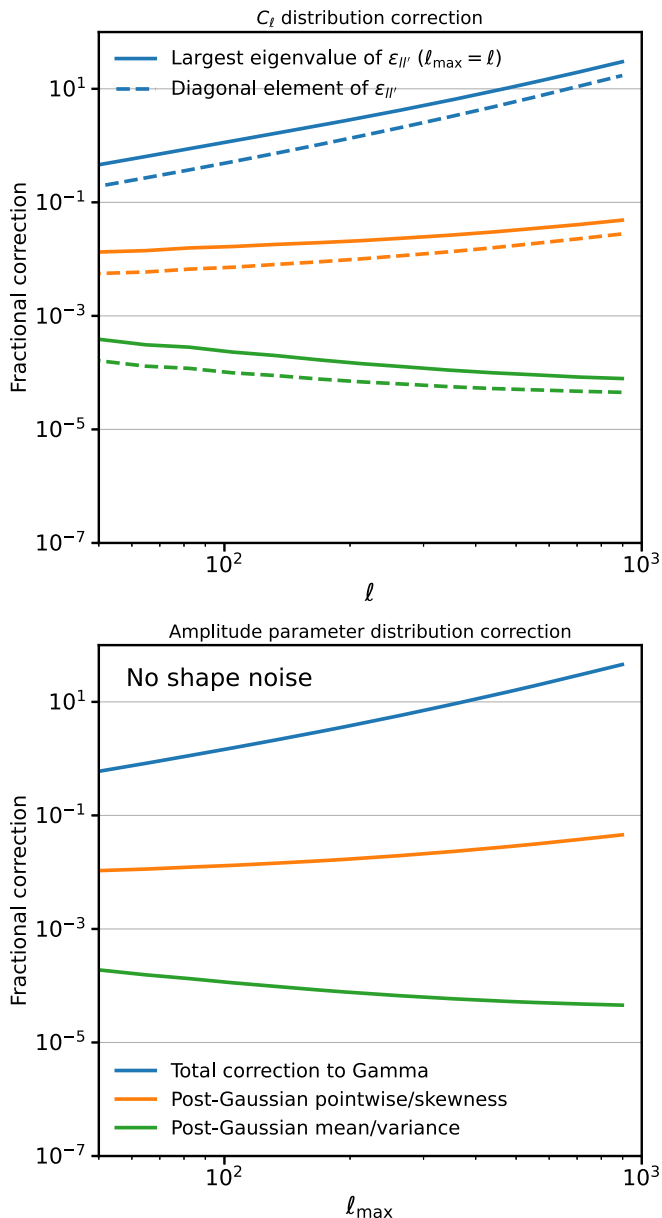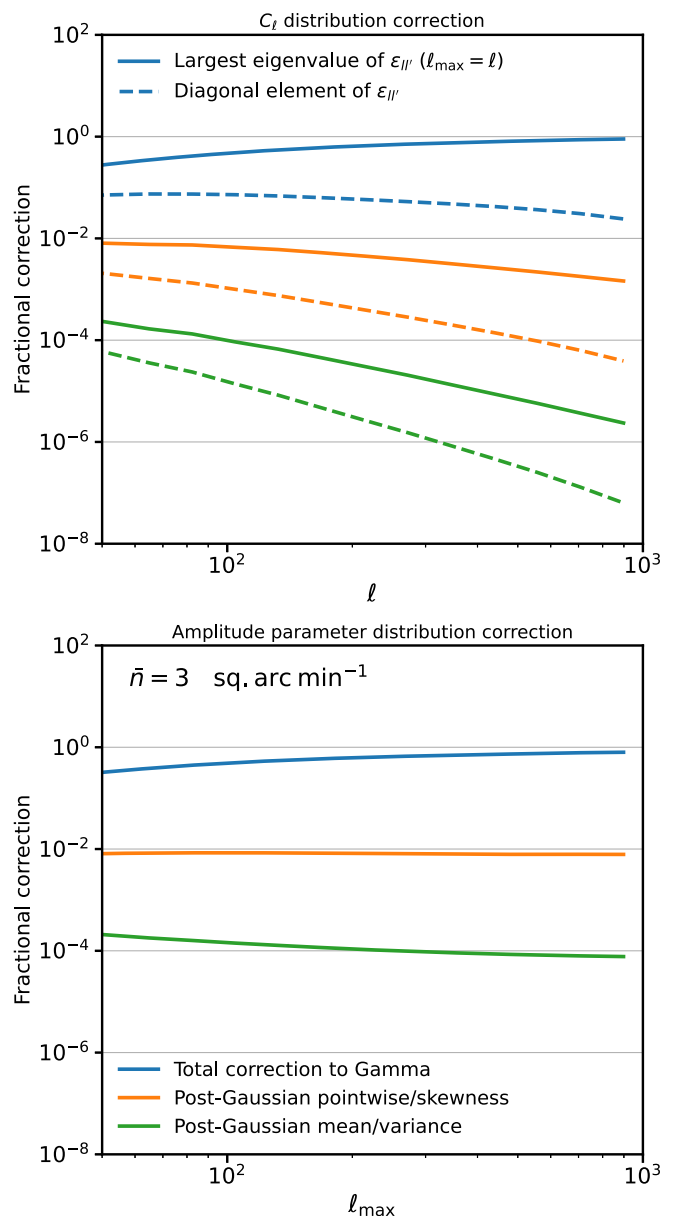

FIG. 10. Leading-order fractional corrections to the Gamma posterior for a Gaussian source redshift distribution centered at $z_s = 0.325$ with width $\sigma_z = 0.1$. Power spectra are assumed to be binned with $\Delta \log l = 0.1$. Curves have the same meaning as in Fig. 7.

FIG. 11. Same as Fig. 10 but with shape noise corresponding to $\bar{n} = 3$ sq. arc min^{-1}.

in tomographic redshift bins—the fiducial choice of Ref. [50] uses ten redshift bins, which is representative of upcoming stage-IV surveys. Assuming equipopulated bins, the redshift distribution of the lowest bin is approximately Gaussian (after accounting for photometric redshift errors) with a peak around $z_s = 0.325$ and width $\sigma_z = 0.1$. Non-Gaussianity in the shear field is relatively stronger at a fixed angular scale at lower redshift due to that scale subtending smaller, more nonlinear spatial scales and the evolution of structure formation toward lower redshift. This is offset to an extent by the overall signal strength being suppressed by the lensing kernel compared with a bin at higher source redshift. This also makes the relative effect of any Gaussianizing shape noise stronger.

To test our likelihood approximation in what is likely the most extreme case for a Euclid-like survey, we generate an ensemble of log-normal fields for this low-redshift source bin. We scale the shift parameter quantifying the strength of the log-normality according to the redshift-scaling formula proposed by Ref. [44], confirming this provides a reasonable match to the covariance matrix of the ray-traced N-body simulations of Ref. [43].

Figure 10 shows the leading-order corrections to the posterior and the convergence statistics for this redshift bin, which should be compared with Fig. 7. The stronger non-Gaussianity is clear to see, with the blue curves quantifying the correction to the Fisher matrix on an individual C_l (left panel) or an amplitude parameter (right panel) passing unity around $l \approx 100$. Our assumption of perturbative non-Gaussianity in the signal is valid only up to around $l \approx 30$. At this angular scale the Gamma distribution is reasonably well approximated by a Gaussian, and passing the full

non-Gaussian covariance matrix to this Gaussian absorbs most of the correction to the posterior that our approximation implies. Residual post-Gaussian effects are at most 1%, and extrapolating out to $l = 1000$ we see that corrections are likely to be no more than 10%. It therefore appears that even for this low source redshift bin the scale of non-Gaussianity ($l \approx 30$) and the scale where Gamma-type corrections are necessary ($l \approx 10$) are separated enough that the likelihood can be approximated as transitioning from a Gamma to a Gaussian with the correct covariance matrix.

When shape noise is added the effects of non-Gaussianity are strongly suppressed. Figure 11 shows leading-order corrections to the posterior for a model in which $C_l = A(S_{0,l} + N_l)$ when applying shape noise with a galaxy number density of $\bar{n} = 3$ sq. arc min^{-1}, appropriate for a single redshift bin in a Euclid-like survey. Our approximation is now convergent out to much higher l, and post-Gaussian corrections are negligible. It thus appears that although the strength of the signal non-Gaussianity may be stronger at low redshift, the relatively higher noise level suppresses non-Gaussianity in the likelihood.

We close this section by noting that these conclusions may need to be modified when considering galaxy clustering at low redshift. There the geometric suppression suffered by lensing is smaller, and hence the Gaussianizing effect of noise may be expected to be relatively weaker. We have also neglected intrinsic alignments, which may be expected to play a greater role in cross-bin combinations at low redshift. We advocate that the statistics plotted in Figs. 7 and 10 be computed whenever a likelihood approximation needs to be made.

[1] M. Kilbinger, Rep. Prog. Phys. **78**, 086901 (2015).

[2] R. Mandelbaum, Annu. Rev. Astron. Astrophys. **56**, 393 (2018).

[3] T. Hamana, M. Shirasaki, S. Miyazaki, C. Hikage, M. Oguri, S. More, R. Armstrong, A. Leauthaud, R. Mandelbaum, H. Miyatake et al., Publ. Astron. Soc. Jpn. **72**, 16 (2020).

[4] C. Heymans, T. Tröster, M. Asgari, C. Blake, H. Hildebrandt, B. Joachimi, K. Kuijken, C.-A. Lin, A. G. Sánchez, J. L. van den Busch et al., Astron. Astrophys. **646**, A140 (2021).

[5] DES Collaboration, T. M. C. Abbott, M. Aguena, A. Alarcon, S. Allam, O. Alves, A. Amon, F. Andrade-Oliveira, J. Annis, S. Avila et al., Phys. Rev. D **105**, 023520 (2022).

[6] T. M. C. Abbott, F. B. Abdalla, A. Alarcon, J. Aleksić, S. Allam, S. Allen, A. Amara, J. Annis, J. Asorey, S. Avila et al., Phys. Rev. D **98**, 043526 (2018).

[7] W. Hu, Astrophys. J. Lett. **522**, L21 (1999).

[8] T. Schrabback, J. Hartlap, B. Joachimi, M. Kilbinger, P. Simon, K. Benabed, M. Bradač, T. Eifler, T. Erben, C. D. Fassnacht et al., Astron. Astrophys. **516**, A63 (2010).

[9] C. Heymans, E. Grocutt, A. Heavens, M. Kilbinger, T. D. Kitching, F. Simpson, J. Benjamin, T. Erben, H. Hildebrandt, H. Hoekstra et al., Mon. Not. R. Astron. Soc. **432**, 2433 (2013).

[10] M. J. Jee, J. A. Tyson, S. Hilbert, M. D. Schneider, S. Schmidt, and D. Wittman, Astrophys. J. **824**, 77 (2016).

[11] W. J. Percival and M. L. Brown, Mon. Not. R. Astron. Soc. **372**, 1104 (2006).

[12] S. Hamimeche and A. Lewis, Phys. Rev. D **77**, 103013 (2008).

[13] R. E. Upham, L. Whittaker, and M. L. Brown, Mon. Not. R. Astron. Soc. **491**, 3165 (2020).

[14] C. Hikage, M. Oguri, T. Hamana, S. More, R. Mandelbaum, M. Takada, F. Köhlinger, H. Miyatake, A. J. Nishizawa, H. Aihara *et al.*, Publ. Astron. Soc. Jpn. **71,** 43 (2019).

[15] A. Nicola, C. García-García, D. Alonso, J. Dunkley, P. G. Ferreira, A. Slosar, and D. N. Spergel, J. Cosmol. Astropart. Phys. 03 (2021) 067.

[16] A. Loureiro, L. Whittaker, A. Spurio Mancini, B. Joachimi, A. Cuceu, M. Asgari, B. Stölzner, T. Tröster, A. H. Wright, M. Bilicki *et al.*, arXiv:2110.06947.

[17] E. Sellentin and A. F. Heavens, Mon. Not. R. Astron. Soc. **456,** L132 (2016).

[18] A. Hall and A. Taylor, Mon. Not. R. Astron. Soc. **483,** 189 (2019).

[19] R. Scoccimarro, Astrophys. J. **544,** 597 (2000).

[20] J. Hartlap, T. Schrabback, P. Simon, and P. Schneider, Astron. Astrophys. **504,** 689 (2009).

[21] P. Schneider and J. Hartlap, Astron. Astrophys. **504,** 705 (2009).

[22] D. Keitel and P. Schneider, Astron. Astrophys. **534,** A76 (2011).

[23] P. Wilking and P. Schneider, Astron. Astrophys. **556,** A70 (2013).

[24] E. Sellentin and A. F. Heavens, Mon. Not. R. Astron. Soc. **473,** 2355 (2018).

[25] E. Sellentin, C. Heymans, and J. Harnois-Déraps, Mon. Not. R. Astron. Soc. **477,** 4879 (2018).

[26] C. Hahn, F. Beutler, M. Sinha, A. Berlind, S. Ho, and D. W. Hogg, Mon. Not. R. Astron. Soc. **485,** 2956 (2019).

[27] C.-H. Lin, J. Harnois-Déraps, T. Eifler, T. Pospisil, R. Mandelbaum, A. B. Lee, S. Singh (LSST Dark Energy Science Collaboration), Mon. Not. R. Astron. Soc. **499,** 2977 (2020).

[28] P. L. Taylor, T. D. Kitching, J. Alsing, B. D. Wandelt, S. M. Feeney, and J. D. McEwen, Phys. Rev. D **100,** 023519 (2019).

[29] A. Diaz Rivero and C. Dvorkin, Phys. Rev. D **102,** 103507 (2020).

[30] O. Friedrich, F. Andrade-Oliveira, H. Camacho, O. Alves, R. Rosenfeld, J. Sanchez, X. Fang, T. F. Eifler, E. Krause, C. Chang *et al.*, Mon. Not. R. Astron. Soc. **508,** 3125 (2021).

[31] S. Smith, A. Challinor, and G. Rocha, Phys. Rev. D **73,** 023517 (2006).

[32] R. Durrett, *Probability: Theory and Examples. Volume 31 of Cambridge Series in Statistical and Probabilistic Mathematics* (Cambridge University Press, Cambridge, UK, 2010).

[33] F. R. Bouchet, R. Juszkiewicz, S. Colombi, and R. Pellat, Astrophys. J. Lett. **394,** L5 (1992).

[34] R. Juszkiewicz, D. H. Weinberg, P. Amsterdamski, M. Chodorowski, and F. Bouchet, Astrophys. J. **442,** 39 (1995).

[35] F. Bernardeau and L. Kofman, Astrophys. J. **443,** 479 (1995).

[36] L. Amendola, Mon. Not. R. Astron. Soc. **283,** 983 (1996).

[37] S. Blinnikov and R. Moessner, Astron. Astrophys. Suppl. Ser. **130,** 193 (1998).

[38] A. N. Taylor and P. I. R. Watts, Mon. Not. R. Astron. Soc. **328,** 1027 (2001).

[39] J. Carron, Astron. Astrophys. **551,** A88 (2013).

[40] G. Ponzano and T. Regge, in *Spectroscopy and Group Theoretical Methods in Physics*, edited by F. Bloch (North-Holland, Amsterdam, 1968).

[41] F. Bernardeau, C. Pitrou, and J.-P. Uzan, J. Cosmol. Astropart. Phys. 02 (2011) 015.

[42] D. Munshi, T. Namikawa, T. D. Kitching, J. D. McEwen, R. Takahashi, F. R. Bouchet, A. Taruya, and B. Bose, Mon. Not. R. Astron. Soc. **493,** 3985 (2020).

[43] R. Takahashi, T. Hamana, M. Shirasaki, T. Namikawa, T. Nishimichi, K. Osato, and K. Shiroyama, Astrophys. J. **850,** 24 (2017).

[44] H. S. Xavier, F. B. Abdalla, and B. Joachimi, Mon. Not. R. Astron. Soc. **459,** 3693 (2016).

[45] A. Taruya, M. Takada, T. Hamana, I. Kayo, and T. Futamase, Astrophys. J. **571,** 638 (2002).

[46] S. Hilbert, J. Hartlap, and P. Schneider, Astron. Astrophys. **536,** A85 (2011).

[47] S. Martin, P. Schneider, and P. Simon, Astron. Astrophys. **540,** A9 (2012).

[48] A. Lewis, A. Challinor, and A. Lasenby, Astrophys. J. **538,** 473 (2000).

[49] C. Howlett, A. Lewis, A. Hall, and A. Challinor, J. Cosmol. Astropart. Phys. 04 (2012) 027.

[50] Euclid Collaboration, A. Blanchard, S. Camera, C. Carbone, V. F. Cardone, S. Casas, S. Clesse, S. Ilić, M. Kilbinger, T. Kitching *et al.*, Astron. Astrophys. **642,** A191 (2020).

[51] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo *et al.*, Astron. Astrophys. **641,** A6 (2020).

[52] A. J. Mead, S. Brieden, T. Tröster, and C. Heymans, Mon. Not. R. Astron. Soc. **502,** 1401 (2021).

[53] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, Astrophys. J. **622,** 759 (2005).

[54] K. Benabed, J. F. Cardoso, S. Prunet, and E. Hivon, Mon. Not. R. Astron. Soc. **400,** 219 (2009).

[55] R. de Belsunce, S. Gratton, W. Coulton, and G. Efstathiou, Mon. Not. R. Astron. Soc. **507,** 1072 (2021).

[56] U. Seljak, G. Aslanyan, Y. Feng, and C. Modi, J. Cosmol. Astropart. Phys. 12 (2017) 009.

[57] B. D. Wandelt, D. L. Larson, and A. Lakshminarayanan, Phys. Rev. D **70,** 083511 (2004).

[58] J. Alsing, T. Charnock, S. Feeney, and B. Wandelt, Mon. Not. R. Astron. Soc. **488,** 4440 (2019).

[59] M. S. Wang, W. J. Percival, S. Avila, R. Crittenden, and D. Bianchi, Mon. Not. R. Astron. Soc. **486,** 951 (2019).

[60] http://cuillin.roe.ac.uk.

[61] T. Okamoto and W. Hu, Phys. Rev. D **66,** 063008 (2002).

[62] A. Cooray and R. Sheth, Phys. Rep. **372,** 1 (2002).

[63] E. Krause and T. Eifler, Mon. Not. R. Astron. Soc. **470,** 2100 (2017).

[64] J. Harnois-Déraps, S. Vafaei, and L. Van Waerbeke, Mon. Not. R. Astron. Soc. **426,** 1262 (2012).

[65] M. Sato and T. Nishimichi, Phys. Rev. D **87,** 123538 (2013).

[66] A. Cooray and W. Hu, Astrophys. J. **534,** 533 (2000).

[67] F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro, Phys. Rep. **367**, 1 (2002).

[68] R. A. Janik and M. A. Nowak, J. Phys. A **36**, 3629 (2003).

[69] C. S. Herz, Ann. Math. **61**, 474 (1955).

[70] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions* (Chapman and Hall, CRC Press, Boca Raton, FL, 2000).

[71] K. B. Petersen and M. S. Pedersen, The Matrix Cookbook (2008), version 20081110, http://www2.imm.dtu.dk/pubdb/p.php?3274.