Editors' Suggestion

Cosmology without window functions. II. Cubic estimators for the galaxy bispectrum

Oliver H. E. Philcox*

Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey 08540, USA and School of Natural Sciences, Institute for Advanced Study, 1 Einstein Drive, Princeton, New Jersey 08540, USA

(Received 16 July 2021; accepted 1 October 2021; published 14 December 2021)

When analyzing the galaxy bispectrum measured from spectroscopic surveys, it is imperative to account for the effects of nonuniform survey geometry. Conventionally, this is done by convolving the theory model with the window function; however, the computational expense of this prohibits full exploration of the bispectrum likelihood. In this work, we provide a new class of estimators for the *unwindowed* bispectrum, a quantity that can be straightforwardly compared to theory. This builds upon the work of Philcox [Phys. Rev. D **103**, 103504 (2021)] for the power spectrum and comprises two parts (both obtained from an Edgeworth expansion): a cubic estimator applied to the data and a Fisher matrix, which deconvolves the bispectrum components. In the limit of weak non-Gaussianity, the estimator is minimum variance; furthermore, we give an alternate form based on Feldman-Kaiser-Peacock weights that is close to optimal and easy to compute. As a demonstration, we measure the binned bispectrum monopole of a suite of simulations using both conventional estimators and our unwindowed equivalents. Computation times are comparable, except that the unwindowed approach requires a Fisher matrix, computable in an additional $\mathcal{O}(100)$ CPU hours. Our estimator may be straightforwardly extended to measure redshift-space distortions and the components of the bispectrum in arbitrary separable bases. The techniques of this work will allow the bispectrum to straightforwardly be included in the cosmological analysis of current and upcoming survey data.

DOI: 10.1103/PhysRevD.104.123529

I. INTRODUCTION

In the standard inflationary paradigm, the early Universe is well described by statistics that are Gaussian and close to scale invariant (see, e.g., [1,2]), an assumption that is in exquisite agreement with cosmic microwave background (CMB) observations (see, e.g., [3]). In this limit, all cosmological information is encapsulated within the power spectrum of the observed field or, equivalently, the two-point correlation function (2PCF), which can be easily measured, modeled, and fit. As the Universe evolves, nonlinear structure growth shifts information from the primordial power spectrum into higher-order moments (see, e.g., [4]), the simplest of which is the bispectrum, or three-point correlation function (3PCF). At the epoch of recombination, the power spectrum still encodes almost all relevant information; thus, we may safely neglect higher-order statistics when measuring ΛCDM parameters from the CMB. By the redshifts corresponding to current and future galaxy surveys, this is not the case; furthermore, higher-order statistics can carry signatures of *primordial* non-Gaussianity (PNG), which is a key probe of inflationary physics.

The next decade will see the advent of vast large-scale structure surveys such as DESI [5] and Euclid [6]. Unlike the CMB, current observations are far from being cosmic variance limited; thus, the upcoming data releases will lead to significant enhancements in our constraining power on cosmological parameters. To extract maximal information from this treasure trove of information, we must combine two-point statistics with their higher-order variants (including the bispectrum) or apply some transformation of the density field (see, e.g., [7]). Such a synergistic approach has been oft proposed in the literature and is expected to give notable improvements in ΛCDM parameter constraints [8,9], as well as nonstandard parameters such as those underlying PNG [10,11], the neutrino sector [12,13], and modified gravity [14].

While there is a long history of estimating PNG parameters from the CMB using higher-order statistics (see, e.g., [15–22]), comparatively few works have attempted to make use of the higher-point functions of the late Universe. There are notable exceptions, however, in particular, the historical bispectrum analyses of Refs. [23–25], as well as some more recent works [26–30], which make use of Sloan Digital Sky Survey (SDSS) data, albeit with several caveats. In configuration space, the situation is similar. Early 3PCF estimates appear in the mid-1970s (see, e.g., [31–35]), and

*ohep2@cantab.ac.uk

there has been a recent resurgence of interest, proving the 3PCF to be a useful statistic in its own right [36–39]. Such approaches are now being extended to even higher-point functions, such as the 4PCF [40,41], though we caution that configuration-space statistics are generically more difficult to model.

Performing robust inference using the galaxy bispectrum does not come without challenges. First, the statistic is nontrivial to measure. Costs associated with its computation are significantly larger than those of the power spectrum, though a number of efficient estimators have recently been developed (see, e.g., [42–46]). Theoretical modeling is similarly difficult, despite being the subject of many works across a number of decades (see, e.g., [47–54]). In particular, while the matter bispectrum can be well modeled perturbatively including its first-order corrections (one loop) [54], we still lack a one-loop theory model for the *galaxy* bispectrum, limiting any analysis to comparatively large scales. Many of the current tree-level models, while capable of modeling the bispectrum alone, are not self-consistent and will introduce systematic errors if they are combined with the power spectrum in future high-precision analyses. An important goal of future work must be to develop and test bispectrum theory models, including via blind challenges analogous to Ref. [55] for the power spectrum.

An additional difficulty in bispectrum analyses concerns dimensionality. With the power spectrum, the number of bins in the data vector is usually far smaller than the number of available simulations, facilitating straightforward mock-based Gaussian analyses. In the bispectrum, this is rarely the case, which has led to some analyses opting to reduce the dimensionality by increasing the bin width, potentially losing cosmological information [27]. A number of low-dimensional analogs of the bispectrum have been proposed to counter this, including the skew spectrum [56–58], line-correlation function [59,60], integrated bispectrum [61,62], and modal projections [63–66], though some lead to significant loss of information [66]. An alternative route is to compress the sample and model bispectra, significantly decreasing the dimensionality while preserving the main information content [52,66–72]. Previous work has shown this to be a powerful option, capable of reducing the size of the bispectrum data vector to ∼30 numbers for a DESI-like survey, without appreciable loss of constraining power [71].

Finally, the situation is complicated by the effects of nonuniform survey geometry. In practice, one does not have access to the galaxy overdensity field; rather, we have only the unnormalized density of galaxies and random particles. This leads to the following transformation in real and Fourier space:

$$\begin{aligned}\delta(\mathbf{x}) &\to \delta_W(\mathbf{x}) \equiv W(\mathbf{x})\delta(\mathbf{x}),\\ \delta(\mathbf{k}) &\to \delta_W(\mathbf{k}) \equiv \int \frac{d\mathbf{p}}{(2\pi)^3} W(\mathbf{k}-\mathbf{p})\delta(\mathbf{p}),\end{aligned} \tag{1}$$

where δ and δ_W are the true and windowed density fields, respectively, and $W(\mathbf{x})$ is the survey mask. In Fourier space (where theory models are most naturally formulated), the action of the window is that of a convolution. This applies similarly to the bispectrum itself:

$$\begin{aligned}B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) &\to B_W(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\\ &\equiv \int \frac{d\mathbf{p}_1}{(2\pi)^3}\frac{d\mathbf{p}_2}{(2\pi)^3}\frac{d\mathbf{p}_3}{(2\pi)^3} W(\mathbf{k}_1-\mathbf{p}_1)\\ &\quad\times W(\mathbf{k}_2-\mathbf{p}_2)W(\mathbf{k}_3-\mathbf{p}_3)B(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}_3),\end{aligned} \tag{2}$$

where B is the true bispectrum (predicted by theory) and B_W its windowed equivalent, which is the quantity computed by most estimators. To compare data and theory, we have two choices: (a) Convolve the theory model with the window via Eq. (2) or (b) estimate the unwindowed bispectrum B directly. Traditionally, approach (a) is chosen, yet this is nontrivial. Unlike for the power spectrum, it is highly expensive to perform the window convolution in full (though see [73] for an approach involving double Hankel transforms), especially considering that this must be done at each step in the eventual Markov chain Monte Carlo (MCMC) analysis. Simplifying assumptions are usually adopted; for example, Refs. [26,27,30] applied the window only to the two power spectra appearing in the tree-level bispectrum model. This assumption is uncontrolled and unwarranted; indeed, Ref. [30] discarded any bispectrum modes with $k_{\rm min} \leq 0.04h$ Mpc^{-1} for this reason. While the current size of survey error bars permits such approximations on all but the largest scales, this will soon change.

In this work, we construct estimators for the *unwindowed* bispectrum. Measuring such a quantity allows data and theory to be directly compared without the need for window convolution, significantly simplifying the eventual MCMC analysis. Our approach is analogous to Ref. [74], which constructed unwindowed estimators for the galaxy power spectrum (see also [75] and the historical approaches of Refs. [76–85]). By optimizing the (non-Gaussian) likelihood for the pixelized galaxy field itself, we obtain a bispectrum estimator that is unbiased and (under certain assumptions) minimum variance, forward modeling the effects of survey geometry. This is similar to the CMB bispectrum estimators of Ref. [86]. Below, we derive the estimators in full and consider their practical implementation, for both the binned bispectrum and an arbitrary (separable) basis decomposition. Contrary to that claimed in Ref. [74], the estimators do not have dependence on external simulations; rather, they can be formulated in a manner requiring only the data and knowledge of the survey geometry. This approach will enable a future measurement of the bispectrum of SDSS galaxies (following Ref. [87] for the power spectrum), including a robust treatment of the window function and a perturbative theory

model encapsulating all effects relevant on quasilinear scales. Although we focus on the rotation-averaged bispectrum monopole in this work, the higher-order anisotropic moments are known to be a useful source of cosmological information [73,88–90]. Our approach may be similarly applied to this scenario and, furthermore, can be used to extract the information from the field directly, without the need for binning or multipole decompositions, if a suitable basis can be found.

The remainder of this paper is structured as follows. In Sec. II, we give an overview of our power spectrum estimator (building upon Ref. [74]), laying the necessary groundwork for Sec. III, wherein our bispectrum estimator is derived. In Sec. IV, we discuss specialization to spectroscopic surveys, before considering practical implementation of the algorithm in Sec. V. Bispectra of realistic mock surveys are presented in Sec. VI, before we conclude in Sec. VII. Appendixes A and B provide details of our modeling of pixelation effects and limiting forms of the bispectrum estimators, respectively. A PYTHON implementation of our algorithm can be found on GitHub [91].

II. OVERVIEW OF PREVIOUS WORK

We begin with a brief summary of the optimal power spectrum estimators discussed in Ref. [74] (itself building on Refs. [77,79,80,83–85]), serving as an introduction to our bispectrum estimators. The treatment below is an updated version of the previous work and removes the need for a suite of simulations.

A. Quadratic estimators

Consider a vector $\mathbf{d}$ of observations, for example, the unnormalized galaxy overdensity measured by some survey in a set of pixels. Usually, this can be modeled as a sum of two components: a theory model $\mathbf{m}$ (i.e., the underlying galaxy density field) and a stochastic noise contribution $\mathbf{n}$ (i.e., Poisson noise), with covariances $\mathsf{S} \equiv \langle \mathbf{m}\mathbf{m}^T \rangle$ and $\mathsf{N} \equiv \langle \mathbf{n}\mathbf{n}^T \rangle$, respectively, assuming $\langle \mathbf{m} \rangle = \langle \mathbf{n} \rangle = 0$.[1] Since both $\mathbf{m}$ and $\mathbf{n}$ are random fields, the dependence on the underlying physical parameters $\mathbf{p}$ enters only through the moments of the dataset; thus, it is useful to first write down the likelihood of the data, $L[\mathbf{d}]$. Under Gaussian assumptions, the negative log-likelihood is given by

$$\ell_G[\mathbf{d}](\mathbf{p}) = -\log L_G[\mathbf{d}](\mathbf{p}) = \frac{1}{2}\mathbf{d}^T\mathsf{C}^{-1}(\mathbf{p})\mathbf{d} + \frac{1}{2}\mathrm{Tr}\,\log\mathsf{C}(\mathbf{p}) + \mathrm{const}, \quad (3)$$

where $\mathsf{C}(\mathbf{p})$ is the pixel covariance matrix, such that $\mathsf{C}(\mathbf{p}^{\rm true}) = \langle \mathbf{d}\mathbf{d}^T \rangle \equiv \mathsf{C}_D$, where $\mathbf{p}^{\rm true}$ is the parameter set that generates $\mathbf{d}$. In this section, $\mathbf{p}$ is the set of *power spectrum amplitudes*, according to some binning. The assumption of Gaussianity holds if the number of modes is large and the underlying density field is linear. Violation of these constraints is the subject of Sec. III.

To obtain an estimate for the parameter vector $\mathbf{p}$, we need simply extremize the likelihood (3). In general, this is nontrivial but can be performed straightforwardly if one first expands $\ell_G(\mathbf{p})$ around some fiducial power spectrum $\mathbf{p}^{\rm fid}$:

$$\ell_G[\mathbf{d}](\mathbf{p}^{\rm fid}+\delta\mathbf{p}) \approx \ell_G(\mathbf{p}^{\rm fid}) + \delta\mathbf{p}^T\nabla_{\mathbf{p}}\ell_G + \frac{1}{2}\delta\mathbf{p}^T(\nabla_{\mathbf{p}}\nabla_{\mathbf{p}'}\ell_G)\delta\mathbf{p}', \quad (4)$$

writing $\delta\mathbf{p} = \mathbf{p} - \mathbf{p}^{\rm fid}$ and assuming all gradients to be evaluated at $\mathbf{p}^{\rm fid}$. If one assumes $\mathbf{p}$ to be close to $\mathbf{p}^{\rm fid}$ [such that $\mathsf{C}(\mathbf{p}^{\rm fid}) \approx \mathsf{C}_D$], this leads to the following Newton-Raphson estimate:

$$\hat{\mathbf{p}} \approx \mathbf{p}^{\rm fid} - [\nabla_{\mathbf{p}}\nabla_{\mathbf{p}'}]^{-1}\nabla_{\mathbf{p}'}\ell_G. \quad (5)$$

Inserting the gradients, we obtain the maximum-likelihood (ML) power spectrum estimator

$$\hat{p}_\alpha^{\rm ML} = p_\alpha^{\rm fid} + \frac{1}{2}\sum_\beta F^{-1,{\rm ML}}_{\alpha\beta}\mathrm{Tr}[\mathsf{C}^{-1}\mathsf{C}_{,\beta}\mathsf{C}^{-1}(\mathbf{d}\mathbf{d}^T - \mathsf{C})] = \frac{1}{2}\sum_\beta F^{-1,{\rm ML}}_{\alpha\beta}\mathrm{Tr}[\mathsf{C}^{-1}\mathsf{C}_{,\beta}\mathsf{C}^{-1}(\mathbf{d}\mathbf{d}^T - \mathsf{N})], \quad (6)$$

where $\mathbf{p} \equiv \{p_\alpha\}$, $\mathsf{C}_{,\alpha} \equiv \partial\mathsf{C}/\partial p_\alpha$, and we define the (realization-averaged) Fisher matrix

$$F^{\rm ML}_{\alpha\beta} = \frac{1}{2}\mathrm{Tr}[\mathsf{C}^{-1}\mathsf{C}_{,\alpha}\mathsf{C}^{-1}\mathsf{C}_{,\beta}]. \quad (7)$$

In the above, we assume all quantities to be evaluated at the fiducial cosmology, i.e., $\mathsf{C} \equiv \mathsf{C}(\mathbf{p}^{\rm fid})$. The estimator is quadratic in the data and involves applying a filter $\mathsf{C}^{-1}\mathsf{C}_{,\alpha}\mathsf{C}^{-1}$ to two copies of $\mathbf{d}$ and then removing additive and multiplicative bias terms. As shown in Ref. [74], it is unbiased [assuming $\mathsf{C}(\mathbf{p}^{\rm fid})$ can be robustly computed] and minimum variance if (a) the likelihood is Gaussian and (b) $\mathsf{C}(\mathbf{p}^{\rm fid}) = \mathsf{C}(\mathbf{p}^{\rm true})$.[2] Extensions to include non-Gaussian noise were also considered in the former work, resulting in a cubic correction term.

While Eq. II A is the optimal estimator (subject to the above caveats), it is just a special case of a more general quadratic estimator, given by

[1] In this work, angle brackets represent an average over realizations of the dataset and the underlying random fields.

[2] Note that the degree of nonoptimality is quadratic in $\mathsf{C}(\mathbf{p}^{\rm fid}) - \mathsf{C}_D$ and, thus, expected to be small.

$$\hat{p}_\alpha = p_\alpha^{\rm fid} + \frac{1}{2}\sum_\beta F^{-1}_{\alpha\beta}{\rm Tr}[\mathsf{H}^{-1}\mathsf{C}_{,\beta}\mathsf{H}^{-1}(\mathbf{d}\mathbf{d}^T - \mathsf{C})]$$
$$= \frac{1}{2}\sum_\beta F^{-1}_{\alpha\beta}{\rm Tr}[\mathsf{H}^{-1}\mathsf{C}_{,\beta}\mathsf{H}^{-1}(\mathbf{d}\mathbf{d}^T - \mathsf{N})],$$
$$F_{\alpha\beta} = \frac{1}{2}{\rm Tr}[\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathsf{C}_{,\beta}], \tag{8}$$

where H is a positive-definite weighting matrix and $\mathsf{C}_{,\alpha}$ is again the derivative of the full covariance with respect to the parameters of interest. The ML estimator simply requires the data to be inverse-covariance weighted, i.e., $\mathsf{H}^{-1} = \mathsf{C}^{-1}$. Given that the full C^{-1} matrix is often expensive to compute, Eq. (8) can lead to significant expedition (though a slight loss of optimality) if some straightforwardly invertible weighting matrix H can be found that is close to C_D.

Two forms of the power spectrum estimators are given in Eqs. II A and (8). Previous work [74] focused on the first choice, which explicitly requires a fiducial model, $\mathbf{p}^{\rm fid}$, for the power spectrum parameters. Both this and the pixel covariance C can be defined using a suite of simulations; thus, the equations take the form of difference estimators, which are unbiased by construction if the data and simulations share the same parameters. However, the reliance on a suite of simulations is a practical restriction on the estimator's use. For this reason, we adopt the second form of the estimators in this work; these do not require mocks and significantly reduce the dependence on some fiducial model [with the general quadratic estimator obviating it entirely, assuming $\mathsf{C}(\mathbf{p})$ to be linear in $\mathbf{p}$].

B. Power spectrum estimation

The principal utility of the above estimators is to measure the galaxy power spectrum multipoles binned in wave number (hereafter the "band powers"). This is convenient since, neglecting non-Gaussianity of the noise, the band powers enter only in the two-point covariance of the signal. To implement the estimators, we must first specify the covariance matrix and its dependence on $\mathbf{p}$. Following Ref. [74], we assume our dataset to be the measured overdensity of galaxy survey, i.e., $d(\mathbf{r}) = \hat{n}_g(\mathbf{r}) - \hat{n}_r(\mathbf{r})$, where $\hat{n}_g(\mathbf{r})$ and $\hat{n}_r(\mathbf{r})$ are discretely sampled galaxy and random density fields, respectively, at position $\mathbf{r}$. Ignoring weights and pixelation for simplicity, these have the pairwise expectations

$$\langle \hat{n}_g(\mathbf{r})\hat{n}_g(\mathbf{r}')\rangle = n(\mathbf{r})n(\mathbf{r}')[1 + \xi(\mathbf{r},\mathbf{r}')] + n(\mathbf{r})\delta_{\rm D}(\mathbf{r}-\mathbf{r}'),$$
$$\langle \hat{n}_r(\mathbf{r})\hat{n}_r(\mathbf{r}')\rangle = n(\mathbf{r})n(\mathbf{r}') + n(\mathbf{r})\delta_{\rm D}(\mathbf{r}-\mathbf{r}'),$$
$$\langle \hat{n}_g(\mathbf{r})\hat{n}_r(\mathbf{r}')\rangle = n(\mathbf{r})n(\mathbf{r}'), \tag{9}$$

where $n(\mathbf{r})$ is the background number density of the survey and $\xi(\mathbf{r},\mathbf{r}') = \langle\delta(\mathbf{r})\delta(\mathbf{r}')\rangle$ is the two-point correlation function of the galaxy overdensity field. The Dirac functions $\delta_{\rm D}$ arise due to the discrete nature of the density fields and source the Poisson noise term. From this, we obtain the signal and noise covariances between two points, $\mathbf{r}$ and $\mathbf{r}'$:

$$\mathsf{C}(\mathbf{r},\mathbf{r}') \equiv \mathsf{S}(\mathbf{r},\mathbf{r}') + \mathsf{N}(\mathbf{r},\mathbf{r}'),$$
$$\mathsf{S}(\mathbf{r},\mathbf{r}') \approx n(\mathbf{r})n(\mathbf{r}')\int_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}P(\mathbf{k}),$$
$$\mathsf{N}(\mathbf{r},\mathbf{r}') \approx n(\mathbf{r})\delta_{\rm D}(\mathbf{r}-\mathbf{r}'), \tag{10}$$

where $P(\mathbf{k})$ is the galaxy power spectrum and $\int_{\mathbf{k}} \equiv \int \frac{d\mathbf{k}}{(2\pi)^3}$. If we wish to measure the monopole power spectrum in some bin α, the relevant covariance derivative becomes

$$\mathsf{C}_{,\alpha}(\mathbf{r},\mathbf{r}') \approx n(\mathbf{r})n(\mathbf{r}')\int_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\Theta^\alpha(k), \tag{11}$$

where we have written $P(\mathbf{k}) \approx \sum_\alpha p_\alpha \Theta^\alpha(k)$, introducing the binning function $\Theta^\alpha(k)$, which is unity in bin α and zero otherwise. In reality, the situation is complicated by the effects of pixelation, particle weights, and redshift-space distortions. These can be straightforwardly included, and the full forms for C are presented in Appendix A 1, including a more advanced treatment of pixelation than Ref. [74].

Given the above covariances, one may compute estimates for the band powers $\{p_\alpha\}$ using either the ML estimator II A or the general form (8). This requires application of both $\mathsf{C}_{,\alpha}$ and C^{-1} or H^{-1} to the data $\mathbf{d}$. Since the pixelized density fields are usually of high dimension, it is impractical to store any of the covariance matrices in full and infeasible to invert them; we can avoid this by considering only the matrices' action on a pixelized field (which can be straightforwardly computed using fast Fourier transforms) and using conjugate-gradient descent methods to invert C [80]. On small scales, a useful approximation to C^{-1} is provided by

$$\mathsf{C}^{-1}(\mathbf{r},\mathbf{r}') \approx \mathsf{H}^{-1}_{\rm FKP} \equiv \frac{\delta_{\rm D}(\mathbf{r}-\mathbf{r}')}{n(\mathbf{r})[1 + n(\mathbf{r})P_{\rm FKP}]}, \tag{12}$$

with $P_{\rm FKP} \sim 10^4 h^{-3}$ Mpc^{-3}. This is analogous to the well-known Feldman-Kaiser-Peacock (FKP) weighting scheme of Ref. [92] and found to be a useful approximation in Ref. [74], since it does not require numerical matrix inversion. An analogous form for this including pixelation effects and particle weights is given in Eq. (A10).

Finally, the quadratic estimators require the traces

$${\rm Tr}[\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathsf{N}], \qquad {\rm Tr}[\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathsf{C}_{,\beta}] \tag{13}$$

[using the form of Eq. (8) which does not involve a fiducial spectrum]. Given that the matrices are too large to be directly computed, these may seem difficult to obtain.

However, as in Refs. [80,86], they may be computed via Monte Carlo methods, first writing

$$\mathrm{Tr}[\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathsf{C}] = \langle \mathbf{a}^T\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathsf{C}\mathsf{A}^{-1}\mathbf{a}\rangle,$$
$$\mathrm{Tr}[\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathsf{C}_{,\beta}] = \langle \mathbf{a}^T\mathsf{H}^{-1}\mathsf{C}_{,\alpha}\mathsf{H}^{-1}\mathsf{C}_{,\beta}\mathsf{A}^{-1}\mathbf{a}\rangle, \quad (14)$$

where $\{\mathbf{a}\}$ are a set of simulated maps that satisfy $\langle \mathbf{a}\mathbf{a}^T\rangle = \mathsf{A}$, for some covariance matrix A. While one could set $\mathsf{A} = \mathsf{C}$ and use mock catalogs for this purpose (as proposed in Ref. [74]), this requires perfect knowledge of C in the mock cosmology, which may be difficult to obtain. In this work, we will use uniformly distributed particles for this purpose, which have a simple, and invertible, form for A given in Eqs. (A11) and (A12). Computing the traces in this way is efficient and incurs a Monte Carlo error scaling as $\sqrt{1+1/N_{\rm mc}}$ when using $N_{\rm mc}$ maps.

C. Properties

Before continuing, we note a number of important properties of the optimal power spectrum estimator.

(i) *Window function*.—Unlike the standard power spectrum estimator of Ref. [92] (and later Refs. [93,94]), the quadratic estimators measure the unwindowed power spectrum; i.e., the output is not convolved with the survey window function. This occurs since the effects of survey geometry are forward modeled through the covariance matrix and allow us to compare measured and theoretical power spectra directly. When using the FKP weights of Eq. (12), our approach is equivalent to that of Ref. [75].

(ii) *Optimality*.—Assuming Gaussianity, the quadratic estimator with $\mathsf{H} = \mathsf{C}$ achieves the tightest possible constraints on the band powers, in the sense that it saturates the Cramér-Rao limit. This will be particularly notable for large-scale analyses, such as those constraining primordial non-Gaussianity (see, e.g., [95]).[3]

(iii) *Compression*.—As shown in Ref. [74], quadratic estimators can be used to directly measure the coefficients of the power spectrum under some linear compression scheme. This significantly reduces dimensionality and obviates the need for k-space binning.

(iv) *Gridding and shot noise*.—The estimators may be formulated as the difference between a quantity measured in simulations and data [as in the first line of Eqs. (6) and (8)]. This removes the leading-order effects of unmeasured modes, discretization, and non-Poissonian shot noise, facilitating the use of coarser pixelation grids. Alternatively, the estimators can be constructed without this [as in the second line of Eqs. (6) and (8), which will be generally assumed here], which removes the need for a fiducial cosmology, and, thus, a suite of simulations.

(v) *Integral constraint*.—When analyzing the output from conventional (windowed) power spectrum estimators, we must account for the integral constraint, i.e., the fact that the overall survey density is not known. In the quadratic estimator framework, such effects are shunted into the fundamental k mode and, thus, may be ignored if this mode is not analyzed.

We refer the reader to Ref. [74] for a more in-depth discussion of these effects, as well as the estimator's application to the measurement of power spectrum multipoles from a suite of simulations.

III. THE CUBIC BISPECTRUM ESTIMATOR

A maximum-likelihood estimator for the bispectrum may be derived in an analogous manner to the above. Before doing so, we outline a number of reasons why this is of use.

(i) While the window function can be straightforwardly included in the theory model for the power spectrum via 1D Hankel transforms, this is considerably more difficult for the bispectrum, since it requires a six-dimensional convolution integral. While some procedures do exist (see, e.g., [73]), they are computationally expensive, prohibiting full parameter exploration.

(ii) The maximum-likelihood estimator derived below is optimal (in the Cramér-Rao sense) in the limit of weak non-Gaussianity. This allows us to extract maximal information from the bispectrum, effectively giving an increase in survey volume.

(iii) We do not require a fiducial model for the bispectrum to use the cubic estimator, just a fiducial power spectrum model (cf. Sec. II), if C^{-1} weights are assumed. This is useful, since obtaining an accurate fiducial bispectrum model, and simulations which reproduce it, is difficult. Furthermore, our estimator is stand alone; i.e., it does not require a suite of realistic simulations.

(iv) Bispectra are generally high dimensional, often involving measurements from $\gtrsim 10^3$ triangles. This requires a large number of mocks to compute a sample covariance, which has limited some previous analyses [27]. Using analogous techniques to those developed for the power spectrum in Ref. [74], we

[3]There is an important caveat to this statement, as noted in Ref. [75]. If one uses the same maximum wave number $k_{\rm max}$ for both unwindowed and windowed power spectrum estimates, the signal to noise of the windowed estimates will be generically slightly larger. This occurs since the window function mixes in modes of larger k, which have smaller variance. If one restricts to the same *preconvolved* k modes in both cases, the quadratic estimator will be at least as constraining as the windowed FKP approach.

may avoid this by directly estimating a set of basis coefficients rather than the full bispectrum, in a similar vein to Ref. [66].

(v) To compare theoretical and observed bispectra, the theory model should properly be integrated over the finite k bins. This is expensive in practice and can be avoided by directly estimating basis coefficients, which are defined from the unbinned bispectra.

Below, we consider the derivation of the general bispectrum estimator, before specializing to the case of spectroscopic surveys in Sec. IV. Part of the below parallels a similar derivation for the CMB bispectrum in Ref. [19], and we will adopt several of the associated tricks used in Ref. [86].

A. Non-Gaussian likelihood

To derive constraints on bispectrum parameters $\mathbf{b}$ (which could be the binned bispectrum estimates or some other summary statistic), we first require a likelihood for the data which contains them. Since such parameters are necessarily absent in the covariance, we require the non-Gaussian likelihood, which, in the limit of mild non-Gaussianity, is given by the Edgeworth expansion:

$$L[\mathbf{d}](\mathbf{b}) = L_G[\mathbf{d}]\left[1 + \frac{1}{3!}\mathsf{B}^{ijk}\mathcal{H}_{ijk} + \frac{1}{4!}\mathsf{T}^{ijkl}\mathcal{H}_{ijkl} + \frac{1}{6!}(\mathsf{B}^{ijk}\mathsf{B}^{lmn} + 9\,\text{perms})\mathcal{H}_{ijklmn} + \cdots\right] \tag{15}$$

(see, e.g., [19,96]). Here, we have written the Gaussian likelihood of Eq. (3) as L_G, which is independent of $\mathbf{b}$, and denoted the connected three- and four-point expectations of the density field by

$$\mathsf{B}^{ijk} = \langle d^i d^j d^k\rangle, \qquad \mathsf{T}^{ijkl} = \langle d^i d^j d^k d^l\rangle - [\langle d^i d^j\rangle\langle d^k d^l\rangle + 2\,\text{perms}]. \tag{16}$$

We use Latin indices to denote pixels in the dataset, such that $d^i \equiv d(\mathbf{r}_i)$, and assume Einstein summation over repeated indices. The Hermite tensors $\mathcal{H}$ used in Eq. (15) may be defined in terms of $h_i \equiv [\mathsf{C}^{-1}\mathbf{d}]_i$ by

$$\begin{aligned}
\mathcal{H}_{ijk} &= h_i h_j h_k - [h_i \mathsf{C}^{-1}_{jk} + 2\,\text{perms}],\\
\mathcal{H}_{ijkl} &= h_i h_j h_k h_l - [h_i h_j \mathsf{C}^{-1}_{kl} + 5\,\text{perms}] + [\mathsf{C}^{-1}_{ij}\mathsf{C}^{-1}_{kl} + 2\,\text{perms}],\\
\mathcal{H}_{ijklmn} &= h_i h_j h_k h_l h_m h_n - [h_i h_j h_k h_l \mathsf{C}^{-1}_{mn} + 14\,\text{perms}]\\
&\quad + [h_i h_j \mathsf{C}^{-1}_{kl}\mathsf{C}^{-1}_{mn} + 44\,\text{perms}] - [\mathsf{C}^{-1}_{ij}\mathsf{C}^{-1}_{kl}\mathsf{C}^{-1}_{mn} + 14\,\text{perms}]
\end{aligned} \tag{17}$$

[96], where $\mathsf{C}^{-1}_{ij} \equiv \mathsf{C}^{-1}(\mathbf{r}_i, \mathbf{r}_j)$. Each tensor is symmetric under any permutation of its arguments, and all involve the data weighted by the inverse covariance, just as in Sec. II. The likelihood (15) is an expansion in non-Gaussianity and, thus, valid if its effects are small. In practice, this can be ensured by restricting the analysis to relatively large scales.

From the above, we may construct a negative log-likelihood for the data as

$$\ell[\mathbf{d}](\mathbf{b}) = \ell_G[\mathbf{d}](\mathbf{b}) - \frac{1}{3!}\mathsf{B}^{ijk}\mathcal{H}_{ijk} - \frac{1}{4!}\mathsf{T}^{ijkl}\mathcal{H}_{ijkl} + \frac{1}{72}\mathsf{B}^{ijk}\mathsf{B}^{lmn}[\mathcal{H}_{ijk}\mathcal{H}_{lmn} - \mathcal{H}_{ijklmn}] + \mathcal{O}(\mathsf{B}^3), \tag{18}$$

absorbing the ten permutations of $\mathsf{B}^{ijk}\mathsf{B}^{lmn}$ into the totally symmetric tensor $\mathcal{H}_{ijklmn}$.

B. Cubic estimators

To obtain an estimator for the bispectrum parameters $\mathbf{b}$, we need to simply extremize (18). Here, we assume $\mathbf{b}$ to be encapsulated solely in the three-point expectation B^{ijk}; i.e., we ignore contributions from the higher-order correlators (arising due to Poisson corrections). As for the power spectrum, we first expand the negative log-likelihood in a Taylor series about some fiducial parameter set, here chosen as $\mathbf{b}^{\rm fid} = \mathbf{0}$:

$$\begin{aligned}
\ell[\mathbf{d}](\mathbf{b}) &= \ell[\mathbf{d}](\mathbf{0}) + \mathbf{b}^T\nabla_{\mathbf{b}}\ell[\mathbf{d}](\mathbf{0})\\
&\quad + \frac{1}{2}\mathbf{b}^T(\nabla_{\mathbf{b}}\nabla_{\mathbf{b}'}\ell[\mathbf{d}](\mathbf{0}))p' + \cdots.
\end{aligned} \tag{19}$$

This is simply an expansion in non-Gaussianity; i.e., the nth-order term contains a product of n three-point correlators. While one could instead expand around $\mathbf{b}^{\rm fid} \neq \mathbf{0}$ (and, thus, obtain a closer-to-optimal estimator if $\mathbf{b}^{\rm fid}$ is well chosen), this will introduce dependence on a fiducial bispectrum model and is, thus, ignored. Maximization of Eq. (19) leads to the estimator

$$\widehat{\mathbf{b}} = \langle -[\nabla_{\mathbf{b}}\nabla_{\mathbf{b}'}\ell[\mathbf{d}](\mathbf{0})]\rangle^{-1}\nabla_{\mathbf{b}'}\ell[\mathbf{d}](\mathbf{0}) + \cdots, \tag{20}$$

where we have additionally taken the expectation of the inverse term, as for the power spectrum case. In the limit of vanishing three- and higher-point correlators in the data (i.e., equivalence of true and fiducial parameters), this is the ML solution, as we show below. The relevant derivatives are easily obtained from Eq. (18):

$$\partial_\alpha \ell[\mathbf{d}](\mathbf{0}) = -\frac{1}{6}\mathsf{B}^{ijk}_{,\alpha}\mathcal{H}_{ijk} + \cdots,$$
$$\partial_\alpha\partial_\beta \ell[\mathbf{d}](\mathbf{0}) = -\frac{1}{36}\mathsf{B}^{ijk}_{,\alpha}\mathsf{B}^{lmn}_{,\beta}[\mathcal{H}_{ijklmn} - \mathcal{H}_{ijk}\mathcal{H}_{lmn}] + \cdots, \tag{21}$$

noting that the Gaussian part ℓ_G and the four-point correlator T^{ijkl} are independent of the bispectrum coefficients. We have additionally used that the Hermite tensors are totally symmetric and asserted that the three-point correlator B^{ijk} is linear in $\mathbf{b}$ (which is true for both the binned bispectrum and the basis components in some linear compression scheme). The ML bispectrum estimator becomes

$$\hat{b}^{\rm ML}_\alpha = \sum_\beta F^{-1,{\rm ML}}_{\alpha\beta}\hat{q}^{\rm ML}_\beta, \tag{22}$$

subject to the definitions

$$\hat{q}^{\rm ML}_\alpha \equiv \frac{1}{6}\mathsf{B}^{ijk}_{,\alpha}\mathcal{H}_{ijk},$$
$$F^{\rm ML}_{\alpha\beta} \equiv \frac{1}{36}\mathsf{B}^{ijk}_{,\alpha}\mathsf{B}^{lmn}_{,\beta}[\langle\mathcal{H}_{ijk}\mathcal{H}_{lmn}\rangle - \langle\mathcal{H}_{ijklmn}\rangle]. \tag{23}$$

This requires the expectation of a Hermite 6-tensor and a pair of 3-tensors [defined in Eq. (17)]; following a tedious, but elementary, calculation, these give

$$\begin{aligned}\langle\mathcal{H}_{ijk}\mathcal{H}_{lmn}\rangle &= (\langle h_i h_j h_k\rangle\langle h_l h_m h_n\rangle + 9\,\text{perms})\\ &\quad + (\langle h_i h_l\rangle\langle h_j h_m\rangle\langle h_k h_n\rangle + 5\,\text{perms})\\ &= (\mathsf{C}^{-1}_{ii'}\mathsf{C}^{-1}_{jj'}\mathsf{C}^{-1}_{kk'}\mathsf{B}^{i'j'k'}\mathsf{C}^{-1}_{ll'}\mathsf{C}^{-1}_{mm'}\mathsf{C}^{-1}_{nn'}\mathsf{B}^{l'm'n'}\\ &\quad + 9\,\text{perms}) + (\mathsf{C}^{-1}_{il}\mathsf{C}^{-1}_{jm}\mathsf{C}^{-1}_{kn} + 5\,\text{perms}),\\ \langle\mathcal{H}_{ijklmn}\rangle &= \langle h_i h_j h_k\rangle\langle h_l h_m h_n\rangle + 9\,\text{perms}\\ &= \mathsf{C}^{-1}_{ii'}\mathsf{C}^{-1}_{jj'}\mathsf{C}^{-1}_{kk'}\mathsf{B}^{i'j'k'}\mathsf{C}^{-1}_{ll'}\mathsf{C}^{-1}_{mm'}\mathsf{C}^{-1}_{nn'}\mathsf{B}^{l'm'n'}\\ &\quad + 9\,\text{perms},\end{aligned} \tag{24}$$

dropping correlators above second order in B and noting that the two-point piece of $\langle\mathcal{H}_{ijklmn}\rangle$ vanishes by orthonormality (Hermite tensors are Gaussian orthonormal with the zeroth-order tensor being unity). This leads to the simplified forms

$$\hat{q}^{\rm ML}_\alpha = \frac{1}{6}\mathsf{B}^{ijk}_{,\alpha}[\mathsf{C}^{-1}\mathbf{d}]_i([\mathsf{C}^{-1}\mathbf{d}]_j[\mathsf{C}^{-1}\mathbf{d}]_k - 3\mathsf{C}^{-1}_{jk}), \tag{25}$$

$$\begin{aligned}F^{\rm ML}_{\alpha\beta} &= \frac{1}{36}\mathsf{B}^{ijk}_{,\alpha}\mathsf{B}^{lmn}_{,\beta}[\mathsf{C}^{-1}_{il}\mathsf{C}^{-1}_{jm}\mathsf{C}^{-1}_{kn} + 5\,\text{perms}]\\ &= \frac{1}{6}\mathsf{B}^{ijk}_{,\alpha}\mathsf{B}^{lmn}_{,\beta}\mathsf{C}^{-1}_{il}\mathsf{C}^{-1}_{jm}\mathsf{C}^{-1}_{kn},\end{aligned} \tag{26}$$

assuming B^{ijk} to be symmetric under $\{i,j,k\}$ permutations and working to lowest nontrivial order in B. This matches that found in, for example, Refs. [63,86].

Much like the power spectrum estimator of Eq. II A involved a filter $\mathsf{C}_{,\alpha}$ applied to two copies of the inverse-covariance-weighted data, $\mathsf{C}^{-1}\mathbf{d}$, the bispectrum estimator of Eq. (22), involves a filter $\mathsf{B}_{,\alpha}$ applied to three copies of $\mathsf{C}^{-1}\mathbf{d}$. In both cases, the Fisher matrix is a trace over two copies of the filter, weighted by some number of inverse covariance matrices. However, the bispectrum Fisher matrix does *not* depend on any bispectrum amplitudes, unlike for the optimal power spectrum estimator. This occurs since we have assumed the non-Gaussianity to be weak, fixing the fiducial bispectrum to zero. This choice does not lead to a bias in the estimator, though it will give a slight loss of optimality if the true bispectrum is large. We further note the subtraction of a term involving $[\mathsf{C}^{-1}\mathbf{d}]_i\mathsf{C}^{-1}_{jk}$ in Eq. (25); this averages to zero and is not included in the simple estimators of Refs. [25–27], though Ref. [63] notes it to be important. In the limit of uniform survey density, this contributes only to the $\mathbf{k} = \mathbf{0}$ mode (Appendix B).

As for the power spectrum case, the form of Eq. (22) motivates a more general cubic bispectrum estimator:

$$\begin{aligned}\hat{b}_\alpha &= \sum_\beta F^{-1}_{\alpha\beta}\hat{q}_\beta,\\ \hat{q}_\alpha &= \frac{1}{6}\mathsf{B}^{ijk}_{,\alpha}[\mathsf{H}^{-1}\mathbf{d}]_i([\mathsf{H}^{-1}\mathbf{d}]_j[\mathsf{H}^{-1}\mathbf{d}]_k - 3\mathsf{H}^{-1}_{jk}),\\ F_{\alpha\beta} &= \frac{1}{6}\mathsf{B}^{ijk}_{,\alpha}\mathsf{B}^{lmn}_{,\beta}\mathsf{H}^{-1}_{il}\mathsf{H}^{-1}_{jm}\mathsf{H}^{-1}_{kn},\end{aligned} \tag{27}$$

for invertible weight matrix H. This approaches the ML solution in the limit $\mathsf{H} \to \mathsf{C}_D$, where C is the covariance of the data.

C. Estimator properties

1. Bias

Taking the expectation of $\hat{q}_\alpha$, the bias of the general cubic estimator (27) is easily considered:

$$\begin{aligned}\mathbb{E}[\hat{q}_\alpha] &= \frac{1}{6}\mathsf{B}^{ijk}_{,\alpha}\mathsf{H}^{-1}_{ii'}\mathsf{H}^{-1}_{jj'}\mathsf{H}^{-1}_{kk'}(\langle d_{i'}d_{j'}d_{k'}\rangle - 3\mathsf{H}^{-1}_{j'k'}\langle d_{i'}\rangle)\\ &= \frac{1}{6}\mathsf{H}^{-1}_{ii'}\mathsf{H}^{-1}_{jj'}\mathsf{H}^{-1}_{kk'}\mathsf{B}^{ijk}_{,\alpha}\sum_\beta b^{\rm true}_\beta\mathsf{B}^{i'j'k'}_{,\beta} = \sum_\beta b^{\rm true}_\beta F_{\alpha\beta}\\ &\Rightarrow \mathbb{E}[\hat{b}_\alpha] = b^{\rm true}_\alpha,\end{aligned} \tag{28}$$

where we have written $[\mathsf{H}^{-1}\mathbf{d}]_i \equiv \mathsf{H}^{-1}_{ii'}d_{i'}$ and assumed that the bispectrum parameters $\mathbf{b}$ fully determine B, i.e., that $\mathsf{B} = \sum_\alpha b^{\rm true}_\alpha\mathsf{B}_{,\alpha}$. This holds true also for the ML estimator of Eq. (22), setting $\mathsf{H} = \mathsf{C}$. Note that we ignore any biases from higher-order terms in the likelihood, i.e., those second order in B. These contribute if $\mathbf{b} \neq \mathbf{0}$ and could be removed at leading order by including an additional "bias" term in the estimator.

When considering binned bispectrum estimates, relation (28) strictly holds only if we measure all possible wave number bins; thus, a more appropriate expansion is $\mathsf{B} = \sum_\alpha b_\alpha^{\rm true}\mathsf{B}_{,\alpha} + \Delta\mathsf{B}$, where $\Delta\mathsf{B}$ contains the bispectrum contributions outside the region of interest. This leads to a bias

$$\Delta b_\alpha \equiv \mathbb{E}[\hat{b}_\alpha] - b_\alpha^{\rm true} = \sum_\beta F^{-1}_{\alpha\beta}\mathsf{B}^{ijk}_{,\beta}\mathsf{H}^{-1}_{ii'}\mathsf{H}^{-1}_{jj'}\mathsf{H}^{-1}_{kk'}\Delta\mathsf{B}_{i'j'k'}. \tag{29}$$

For optimal weights and a uniform survey of infinite volume, $\mathsf{C}^{-1}_{ii'} \propto \delta^{\rm K}_{ii'}$; thus, the bias arises from the term $\mathsf{B}^{ijk}_{,\beta}\Delta\mathsf{B}_{ijk}$. This is vanishing for all β, since $\Delta\mathsf{B}$ contains contributions only from modes not in $\{b_\beta\}$. In more realistic scenarios, this mode-coupling bias is expected to be small, assuming the survey window to be relatively compact. Reference [74] suggested that one should estimate a slightly larger range of wave number bins than used in an eventual analysis to ameliorate such a bias; we apply this principle in Sec. V.

Special care is required if some bispectrum decomposition scheme is adopted in which there is an additional linear term $\bar{\mathsf{B}}$ not included in the parameter set $\mathbf{b}$, i.e., if $\mathsf{B} = \sum_\alpha b_\alpha\mathsf{B}_{,\alpha} + \bar{\mathsf{B}}$. An example of this is the singular value decomposition scheme proposed in Ref. [71], which expands the statistic around some fiducial mean. In this case, we must modify the general estimator to ensure it remains unbiased:

$$\bar{b}_\alpha \to \sum_\beta F^{-1}_{\alpha\beta}[\hat{q}_\beta - \bar{q}_\beta], \qquad \bar{q}_\alpha = \frac{1}{6}\mathsf{H}^{-1}_{ii'}\mathsf{H}^{-1}_{jj'}\mathsf{H}^{-1}_{kk'}\mathsf{B}^{ijk}_{,\alpha}\bar{\mathsf{B}}^{i'j'k'}. \tag{30}$$

An alternative approach, which is somewhat easier to estimate, is to expand the log-likelihood around the mean $\bar{\mathsf{B}}$ instead of zero; in this instance, we estimate the *difference* between b_α and $\bar{b}_\alpha$ and may drop the $\bar{q}_\alpha$ term. Assuming $\bar{\mathsf{B}}$ to be a reasonable bispectrum estimate, this will not lead to a significant loss of optimality.

2. Covariance

Assuming the optimal weighting $\mathsf{H} = \mathsf{C}$, the covariance of the bispectrum estimator (22) can be derived via Wick's theorem:

$$\mathrm{cov}(\hat{b}^{\rm ML}_\alpha, \hat{b}^{\rm ML}_\beta) = \sum_{\gamma\delta} F^{-1}_{\alpha\gamma}F^{-1}_{\beta\delta}\mathrm{cov}(\hat{q}^{\rm ML}_\gamma, \hat{q}^{\rm ML}_\delta), \tag{31}$$

with

$$\begin{aligned}\mathrm{cov}(\hat{q}^{\rm ML}_\alpha, \hat{q}^{\rm ML}_\beta) &\equiv \mathbb{E}[\hat{q}^{\rm ML}_\alpha\hat{q}^{\rm ML}_\beta] - \mathbb{E}[\hat{q}^{\rm ML}_\alpha]\mathbb{E}[\hat{q}^{\rm ML}_\beta] = \frac{1}{36}\mathsf{B}^{ijk}_{,\alpha}\mathsf{B}^{lmn}_{,\beta}[\langle\mathcal{H}_{ijk}\mathcal{H}_{lmn}\rangle - \langle\mathcal{H}_{ijk}\rangle\langle\mathcal{H}_{lmn}\rangle] \\ &= \frac{1}{6}\mathsf{B}^{ijk}_{,\alpha}\mathsf{B}^{lmn}_{,\beta}\mathsf{C}^{-1}_{il}\mathsf{C}^{-1}_{jm}\mathsf{C}^{-1}_{kn} + \frac{5}{18}\mathsf{C}^{-1}_{ii'}\mathsf{C}^{-1}_{jj'}\mathsf{C}^{-1}_{kk'}\mathsf{C}^{-1}_{ll'}\mathsf{C}^{-1}_{mm'}\mathsf{C}^{-1}_{nn'}\mathsf{B}^{i'j'k'}\mathsf{B}^{l'm'n'}\mathsf{B}^{ijk}_{,\alpha}\mathsf{B}^{lmn}_{,\beta}.\end{aligned} \tag{32}$$

This uses the expectation of two Hermite 3-tensors given in Eq. (24) and is valid at second order in B. In the limit of a small bispectrum, this is equal to the inverse Fisher matrix, such that

$$\mathrm{cov}(\hat{b}^{\rm ML}_\alpha, \hat{b}^{\rm ML}_\beta) = F^{-1,{\rm ML}}_{\alpha\beta} + \mathcal{O}(\mathsf{B}^2). \tag{33}$$

The calculation for general weighting H is analogous but lengthy and is, thus, omitted from this publication.

3. Optimality

According to the Cramér-Rao theorem, for the estimator $\hat{b}^{\rm ML}_\alpha$ to be optimal, it must satisfy

$$\mathrm{cov}(\hat{b}^{\rm ML}_\alpha, \hat{b}^{\rm ML}_\beta)|_{\rm CR\,bound} = \left\langle\frac{\partial^2\ell[\mathbf{d}](\mathbf{b})}{\partial b_\alpha\partial b_\beta}\right\rangle^{-1} \equiv \mathcal{I}^{-1}_{\alpha\beta}, \tag{34}$$

where the right-hand side is the inverse Fisher information for negative log-likelihood $\ell[\mathbf{d}](\mathbf{b})$ depending on parameters $\mathbf{b}$ and data $\mathbf{d}$ (see, e.g., [97]). Since our Fisher matrix $F^{\rm ML}_{\alpha\beta}$ is defined as the (realization-averaged) second derivative of the log-likelihood, $\mathcal{I}_{\alpha\beta} = F^{\rm ML}_{\alpha\beta}$, and Eq. (33) demonstrates that the ML estimator satisfies its Cramér-Rao bound in the limit $\mathsf{B} \to 0$. We thus obtain optimality in the limit of weak non-Gaussianity, assuming inverse covariance weighting and neglecting non-Gaussian contributions from stochastic effects. A corollary of this is that any estimator with $\mathsf{H} \neq \mathsf{C}$ must give larger error bars, since the covariance of $\hat{b}_\alpha$ will necessarily be different from Eq. (33), and no unbiased estimator can exceed its Cramér-Rao bound.

IV. SPECIALIZATION TO SPECTROSCOPIC SURVEYS

We now consider the form of the bispectrum estimators [Eqs. (22) and (27)] when applied to spectroscopic surveys. We will specialize to the binned isotropic (i.e., rotationally averaged) bispectrum in Sec. IV B before commenting on alternative bases in Sec. IV C.

A. General formalism

As in Sec. II, we assume the data vector to be a (pixelized) field of data-minus-randoms, such that $d(\mathbf{r}) \equiv \hat{n}_g(\mathbf{r}) - \hat{n}_r(\mathbf{r})$. For clarity, we ignore the effects of pixelation and particle weights in this section; the full estimators including such phenomena are presented in Appendix A 2.

Assuming Poisson statistics for the two discrete density fields, the two- and three-point correlators are given by

$$\begin{aligned}\mathsf{C}^{ij} &\equiv \langle d^i d^j\rangle = n(\mathbf{r}_i)n(\mathbf{r}_j)\xi(\mathbf{r}_i,\mathbf{r}_j) + n(\mathbf{r}_i)\delta_{\rm D}(\mathbf{r}_i-\mathbf{r}_j),\\ \mathsf{B}^{ijk} &\equiv \langle d^i d^j d^k\rangle = n(\mathbf{r}_i)n(\mathbf{r}_j)n(\mathbf{r}_k)\zeta(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k) + [n(\mathbf{r}_i)n(\mathbf{r}_j)\delta_{\rm D}(\mathbf{r}_i-\mathbf{r}_k)\xi(\mathbf{r}_i-\mathbf{r}_j) + 2\,\text{perms}]\\ &\quad + n(\mathbf{r}_i)\delta_{\rm D}(\mathbf{r}_i-\mathbf{r}_j)\delta_{\rm D}(\mathbf{r}_j-\mathbf{r}_k),\end{aligned} \tag{35}$$

applying the results of Eq. (9). This uses the background number density field $n(\mathbf{r})$, the 2PCF $\xi(\mathbf{r}_1,\mathbf{r}_2) = \langle\delta(\mathbf{r}_1)\delta(\mathbf{r}_2)\rangle$, and the 3PCF $\zeta(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3) = \langle\delta(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3)\rangle$, retaining dependence on all position vectors for generality. The terms involving Dirac deltas arise from Poisson contractions of the density field; these take a slightly different form after incorporating pixelation effects (Appendix A). Writing the 2PCF and 3PCF in terms of their Fourier-space counterparts $P(\mathbf{k})$ and $B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)$, respectively, we obtain

$$\begin{aligned}\mathsf{C}^{ij} &= n(\mathbf{r}_i)n(\mathbf{r}_j)\int_{\mathbf{k}} P(\mathbf{k})e^{i\mathbf{k}\cdot(\mathbf{r}_i-\mathbf{r}_j)} + n(\mathbf{r}_i)\delta_{\rm D}(\mathbf{r}_i-\mathbf{r}_j),\\ \mathsf{B}^{ijk} &= n(\mathbf{r}_i)n(\mathbf{r}_j)n(\mathbf{r}_k)\int_{\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3} B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)e^{i\mathbf{k}_1\cdot\mathbf{r}_i+i\mathbf{k}_2\cdot\mathbf{r}_j+i\mathbf{k}_3\cdot\mathbf{r}_k}(2\pi)^3\delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\\ &\quad + \left[n(\mathbf{r}_i)n(\mathbf{r}_j)\delta_{\rm D}(\mathbf{r}_i-\mathbf{r}_k)\int_{\mathbf{k}} P(\mathbf{k})e^{i\mathbf{k}\cdot(\mathbf{r}_i-\mathbf{r}_j)} + 2\,\text{perms}\right] + n(\mathbf{r}_i)\delta_{\rm D}(\mathbf{r}_i-\mathbf{r}_j)\delta_{\rm D}(\mathbf{r}_j-\mathbf{r}_k),\end{aligned} \tag{36}$$

where the Dirac delta function $(2\pi)^3\delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)$ arises from translation invariance. Additionally, we have made the flat-sky approximation, assuming the power spectrum and bispectrum to be independent of the position vectors $\mathbf{r}$. This is done only for clarity; we can retain dependence of the statistics on the *local* line of sight (e.g., in the Yamamoto approximation [93]), as done for the power spectrum multipoles in Ref. [74]. Full treatment of the three-pixel correlator allowing for this and other effects can be found in Eq. (A13).

In the estimators of Sec. III, B^{ijk} does not enter directly; instead, we require its derivatives with respect to a set of bispectrum parameters. Here, we assume the gravitational bispectrum $B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)$ to be a linear sum of templates, such that

$$B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) = \sum_\alpha b_\alpha \omega^\alpha(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3), \tag{37}$$

where $b \equiv \{b_\alpha\}$ are the parameters of interest and $\omega^\alpha(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)$ are some templates, which are of relevance only if $\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3 = \mathbf{0}$.[4] This decomposition is fully applicable to the simple case of binned bispectrum estimates as well as more nuanced schemes; the former is discussed in Sec. IV B.

Using Eq. (37), the derivatives of Eq. (36) are straightforward:

$$\mathsf{B}^{ijk}_{,\alpha} = n(\mathbf{r}_i)n(\mathbf{r}_j)n(\mathbf{r}_k)\int_{\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3}\omega^\alpha(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)e^{ik_1\cdot\mathbf{r}_i+i\mathbf{k}_2\cdot\mathbf{r}_j+i\mathbf{k}_3\cdot\mathbf{r}_k}(2\pi)^3\delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3), \tag{38}$$

leading to the following $\hat{q}_\alpha$ estimator [from Eq. (27)]:

$$\begin{aligned}\hat{q}_\alpha &= \frac{1}{6}\mathsf{B}^{ijk}_{,\alpha}[\mathsf{H}^{-1}d]_i([\mathsf{H}^{-1}d]_j[\mathsf{H}^{-1}d]_k - 3\mathsf{H}^{-1}_{jk})\\ &= \frac{1}{6}\int_{\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3}\omega^\alpha(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\mathcal{F}[n\mathsf{H}^{-1}d](\mathbf{k}_1)\mathcal{F}[n\mathsf{H}^{-1}d](\mathbf{k}_2)\mathcal{F}[n\mathsf{H}^{-1}\mathbf{d}](\mathbf{k}_3)(2\pi)^3\delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\\ &\quad -\frac{1}{2}\int_{\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3}\omega^\alpha(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\mathcal{F}[n\mathsf{H}^{-1}\mathbf{d}](\mathbf{k}_1)\langle\mathcal{F}[n\mathsf{H}^{-1}\mathbf{a}](\mathbf{k}_2)\mathcal{F}[n\mathsf{A}^{-1}\mathbf{a}](k_3)\rangle(2\pi)^3\delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3).\end{aligned} \tag{39}$$

[4]If necessary, we can introduce an offset such that $B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) = \sum b_\alpha\omega^\alpha(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) + \bar{B}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)$. This is of relevance when considering singular value decompositions and modifies the bispectrum estimator slightly, as discussed in Sec. III C.

In the above, we have performed the spatial integrals (which are simply Fourier transforms, denoted by $\mathcal{F}$) and, in the third line, written $\mathsf{H}^{-1}_{jk} \equiv \mathsf{H}^{-1}_{jj'}\mathsf{A}^{j'k'}\mathsf{A}^{-1}_{k'k} \equiv \langle \mathsf{H}^{-1}_{jj'} a^{j'} \mathsf{A}^{-1}_{kk'} a^{k'} \rangle$, where $\{\mathbf{a}\}$ are a set of simulated maps with known covariance A, i.e., $\langle \mathbf{a}\mathbf{a}^T \rangle = \mathsf{A}$. This is similar to Ref. [86] and is the same trick used to evaluate the ML power spectrum bias term in Sec. II. Unlike in previous work, we do not enforce $\mathsf{A} = \mathsf{H}$; instead, we use uniformly distributed particles for this purpose (as in Sec. II), removing the dependence on N-body simulations. The corresponding form for A^{-1} (including pixelation effects) is given in Eq. (A12).

In practice, the above expression is difficult to compute due to the momentum-conserving delta function. A useful assumption is that the filter functions ω^α are separable, such that

$$\omega^\alpha(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \prod_{i=1}^{3} \omega^{\alpha,i}(\mathbf{k}_i). \tag{40}$$

This is the case for the binned bispectrum estimates considered below. Writing the Dirac function as the integral of a complex exponential, this yields a straightforward form for the $\hat{q}_\alpha$ estimator:

$$\hat{q}_\alpha = \frac{1}{6}\int d\mathbf{r} g^{\alpha,1}[\mathbf{d}](\mathbf{r}) g^{\alpha,2}[\mathbf{d}](\mathbf{r}) g^{\alpha,3}[\mathbf{d}](\mathbf{r}) - \frac{1}{6}\int d\mathbf{r}[g^{\alpha,1}[\mathbf{d}](\mathbf{r})\langle g^{\alpha,2}[\mathbf{a}](\mathbf{r})\tilde{g}^{\alpha,3}[\mathbf{a}](\mathbf{r})\rangle + 2\,\text{perms}], \tag{41}$$

where we have defined the functions

$$g^{\alpha,i}[\mathbf{y}](\mathbf{r}) = \int_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{r}}\omega^{\alpha,i}(\mathbf{k}) \int d\mathbf{r}' e^{i\mathbf{k}\cdot\mathbf{r}'} n(\mathbf{r}')[\mathsf{H}^{-1}\mathbf{y}](\mathbf{r}'),$$
$$\tilde{g}^{\alpha,i}[\mathbf{y}](\mathbf{r}) = \int_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{r}}\omega^{\alpha,i}(\mathbf{k}) \int d\mathbf{r}' e^{i\mathbf{k}\cdot\mathbf{r}'} n(\mathbf{r}')[\mathsf{A}^{-1}\mathbf{y}](\mathbf{r}'). \tag{42}$$

These may be efficiently computed using fast Fourier transforms (FFTs).

In this manner, the $\hat{q}_\alpha$ estimator may be evaluated as a set of FFTs and, for the $\mathbf{r}$ integrals, real-space summations. Note that we have allowed the templates to depend explicitly on the direction $\mathbf{k}_i$; while no such dependence is required when measuring the rotationally averaged bispectrum, it will arise in generalizations to higher-order moments.

To form the full bispectrum estimator, we additionally require an explicit form for the Fisher matrix (26). When using ML weighting schemes, one option is to compute $F^{\rm ML}_{\alpha\beta}$ as the (inverse) sample covariance of a set of $N_{\rm mc}$ Monte Carlo realizations of $\hat{q}^{\rm ML}_\alpha$, using Eq. (32). However, this requires a suite of realistic simulations, works only for ML weights, and is slow to converge, since the Fisher matrix must be invertible. As in Refs. [74,86], we instead compute the full matrix via Monte Carlo averages. This proceeds similarly to that of $\hat{q}_\alpha$; we first rewrite

$$\frac{1}{2}\mathsf{H}^{-1}_{il}(\mathsf{H}^{-1}_{jm}\mathsf{H}^{-1}_{kn} + \mathsf{H}^{-1}_{jn}\mathsf{H}^{-1}_{km}) \equiv \frac{1}{2}\mathsf{H}^{-1}_{il}\mathsf{H}^{-1}_{jj'}\mathsf{H}^{-1}_{kk'}\mathsf{A}^{-1}_{mm'}\mathsf{A}^{-1}_{nn'}(\mathsf{A}^{j'm'}\mathsf{A}^{k'n'} + \mathsf{A}^{j'n'}\mathsf{A}^{k'm'}) \tag{43}$$

(assuming interchange symmetry under $\{j,m\} \leftrightarrow \{k,n\}$) and then replace the two-point cumulants with Monte Carlo averages, using

$$\mathsf{A}^{j'm'}\mathsf{A}^{k'n'} + \mathsf{A}^{j'n'}\mathsf{A}^{k'm'} = \langle a^{j'}a^{k'}a^{m'}a^{n'}\rangle - \langle a^{j'}a^{k'}\rangle\langle a^{m'}a^{n'}\rangle, \tag{44}$$

for Monte Carlo simulations $\mathbf{a}$ satisfying $\langle \mathbf{a}\mathbf{a}^T\rangle = \mathsf{A}$, assuming the contributions from higher-point correlators to be small.[5] This gives

$$F_{\alpha\beta} = \frac{1}{12}\langle(\mathsf{B}^{ijk}_{,\alpha}\mathsf{H}^{-1}_{jj'}\mathsf{H}^{-1}_{kk'}a^{j'}a^{k'})\mathsf{H}^{-1}_{il}(\mathsf{B}^{lmn}_{,\beta}\mathsf{A}^{-1}_{mm'}\mathsf{A}^{-1}_{nn'}a^{m'}a^{n'})\rangle - \frac{1}{12}\langle\mathsf{B}^{ijk}_{,\alpha}\mathsf{H}^{-1}_{jj'}\mathsf{H}^{-1}_{kk'}a^{j'}a^{k'}\rangle\mathsf{H}^{-1}_{il}\langle\mathsf{B}^{lmn}_{,\beta}\mathsf{A}^{-1}_{mm'}\mathsf{A}^{-1}_{nn'}a^{m'}a^{n'}\rangle. \tag{45}$$

While Eq. (45) seems complicated, it can be expressed as a vector product. To see this, first define

$$\begin{aligned}\phi^i_\alpha[\mathbf{a}] &= \mathsf{B}^{ijk}_{,\alpha}\mathsf{H}^{-1}_{jj'}\mathsf{H}^{-1}_{kk'}a^{j'}a^{k'}\\ &= \int_{\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3} e^{i\mathbf{k}_1\cdot\mathbf{r}_i}\omega^\alpha(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)(2\pi)^3\delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)n(\mathbf{r}_i)[n\mathsf{H}^{-1}a](\mathbf{k}_2)[n\mathsf{H}^{-1}\mathbf{a}](\mathbf{k}_3)\\ &= \frac{1}{3}\int_{\mathbf{k}_1} d\mathbf{r} e^{i\mathbf{k}_1\cdot(\mathbf{r}_i-\mathbf{r})}(\omega^{\alpha,1}(\mathbf{k}_1)n(\mathbf{r}_i)g^{\alpha,2}[\mathbf{a}](\mathbf{r})g^{\alpha,3}[\mathbf{a}](\mathbf{r}) + 2\,\text{perms})\\ &= \frac{1}{3}n(\mathbf{r}_i)(\mathcal{F}^{-1}[\omega^{\alpha,1}(\mathbf{k})\mathcal{F}[g^{\alpha,2}[\mathbf{a}]g^{\alpha,3}[\mathbf{a}]](\mathbf{k})](\mathbf{r}_i) + 2\,\text{perms}),\end{aligned} \tag{46}$$

[5]When using a set of N uniformly distributed particles to define $\mathbf{a}$, the two- and four-point correlators are $\mathsf{A}_{ij} = \delta^{\rm K}_{ij}a(\mathbf{r}_i)$ and $\mathsf{A}_{ijkl} = \delta^{\rm K}_{ij}\delta^{\rm K}_{jk}\delta^{\rm K}_{kl}a(\mathbf{r}_i)$, respectively, indexing galaxies by i, j, k, l and ignoring pixelation effects for simplicity. Taking the trace, $\sum_i \mathsf{A}_{iiii} = N$, $\sum_i \mathsf{A}_{ii}\mathsf{A}_{ii} = N^2$, thus, the four-point correlator is negligible, for $N \gg 1$.

and analogously $\tilde{\phi}^i_\alpha[\mathbf{a}]$ with $\mathsf{H}^{-1} \to \mathsf{A}^{-1}$, and, thus, $g \to \tilde{g}$. Each of these can be evaluated using FFTs, assuming separability of the bispectrum templates. The Fisher matrix is, thus,

$$F_{\alpha\beta} = \frac{1}{12}(\langle \phi^i_\alpha \mathsf{H}^{-1}_{il} \tilde{\phi}^l_\beta \rangle - \langle \phi^i_\alpha \rangle \mathsf{H}^{-1}_{il} \langle \tilde{\phi}^l_\beta \rangle), \tag{47}$$

which can be evaluated using a matrix inversion and a real-space sum. In practice, the second term is small except on the largest scales, as discussed in Appendix B. When considering the ML estimators, we must perform one matrix inversion per element of $\mathbf{b}$ (as well as one to define $\mathsf{C}^{-1}\mathbf{a}$); thus, the computation time scales as $(N_{\rm bins}+1)$, indicating the benefits of an efficient data compression scheme. As before, the error on $F_{\alpha\beta}$ scales as $\sqrt{1+1/N_{\rm mc}}$ for $N_{\rm mc}$ simulations, implying that $N_{\rm mc} \sim 50$ is requires to compute the statistic to percent-level accuracy. This applies also to the second term in $\hat{q}_\alpha$.

B. Application to the binned bispectrum

We now specialize to the measurement of the bispectrum monopole in a set of k bins. This is the quantity obtained from most standard estimators (see, e.g., [25–27]). Here, we will assume the parameter b_α to be the measured bispectrum amplitude in an ordered triplet of k bins, indexed by $\alpha \equiv \{a,b,c\}$ with $a \le b \le c$. In this case, the bispectrum can be decomposed separably as

$$B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) \approx \sum_\alpha \frac{1}{\Delta_\alpha} b_\alpha [\Theta^a(k_1)\Theta^b(k_2)\Theta^c(k_3) + 5\,\text{perms}], \tag{48}$$

where the binning function Θ^α is unity in bin a and zero otherwise. We additionally require the symmetry factor Δ_α, defined as

$$\Delta_\alpha = \begin{cases} 6 & \text{if } a=b=c, \\ 2 & \text{if } a=b \text{ or } a=c \text{ or } b=c, \\ 1 & \text{else}, \end{cases} \tag{49}$$

to account for the permutations of Eq. (48). This may be familiar from the Gaussian covariance of the bispectrum (see, e.g., [47]). The momentum conservation condition $\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3=0$ places strong constraints on the triangle bins, in particular, $|k_a-k_b| < k_c < k_a+k_b$, where k_i is the center of bin i, if we ignore the effects of finite bin widths. Note that our decomposition ignores any directional information about the triangle; i.e., we consider only the rotationally averaged bispectrum. In the absence of redshift-space distortions, this is expected to capture all possible bispectrum information; in the alternate case, it will encode only the dominant component. Generalizations to the anisotropic moments are considered in Sec. IV C.

The cumulant derivative $\mathsf{B}^{ijk}_{,\alpha}$ [Eq. (38)] becomes

$$\mathsf{B}^{ijk}_{,\alpha} = \frac{n(\mathbf{r}_i)n(\mathbf{r}_j)n(\mathbf{r}_k)}{\Delta_\alpha} \int_{\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3} [\Theta^a(k_1)\Theta^b(k_2)\Theta^c(k_3) + 5\,\text{perms}] e^{i\mathbf{k}_1\cdot\mathbf{r}_i + i\mathbf{k}_2\cdot\mathbf{r}_j + i\mathbf{k}_3\cdot\mathbf{r}_k} (2\pi)^3 \delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3), \tag{50}$$

leading to the following form for $\hat{q}_\alpha$ [from Eq. (41)]:

$$\hat{q}_\alpha = \frac{1}{\Delta_\alpha} \int d\mathbf{r} \{ g^a[\mathbf{d}](\mathbf{r}) g^b[\mathbf{d}](\mathbf{r}) g^c[\mathbf{d}](\mathbf{r}) - (g^a[\mathbf{d}](\mathbf{r}) \langle g^b[\mathbf{a}](\mathbf{r}) \tilde{g}^c[\mathbf{a}](\mathbf{r}) \rangle + 2\,\text{perms}) \}, \tag{51}$$

incorporating the permutation symmetries. This uses the definitions

$$\begin{aligned} g^a[\mathbf{y}](\mathbf{r}) &= \int_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{r}} \Theta^a(k) \int d\mathbf{r}' e^{i\mathbf{k}\cdot\mathbf{r}'} n(\mathbf{r}') [\mathsf{H}^{-1}\mathbf{y}](\mathbf{r}') \equiv \mathcal{F}^{-1}[\Theta^a(k) \mathcal{F}[n \mathsf{H}^{-1}\mathbf{y}](\mathbf{k})](\mathbf{r}), \\ \tilde{g}^a[\mathbf{y}](\mathbf{r}) &= \int_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{r}} \Theta^a(k) \int d\mathbf{r}' e^{i\mathbf{k}\cdot\mathbf{r}'} n(\mathbf{r}') [\mathsf{A}^{-1}\mathbf{y}](\mathbf{r}') \equiv \mathcal{F}^{-1}[\Theta^a(k) \mathcal{F}[n \mathsf{A}^{-1}\mathbf{y}](\mathbf{k})](\mathbf{r}), \end{aligned} \tag{52}$$

analogous to Eq. (42). The Fisher matrix is given by

$$F_{\alpha\beta} = \frac{1}{12}(\langle \phi^i_\alpha \mathsf{H}^{-1}_{il} \tilde{\phi}^l_\beta \rangle - \langle \phi^i_\alpha \rangle \mathsf{H}^{-1}_{il} \langle \tilde{\phi}^l_\beta \rangle), \tag{53}$$

as in Eq. (47), where the ϕ coefficients of Eq. (46) simplify to

$$\begin{aligned} \phi_\alpha[\mathbf{a}](\mathbf{r}) &= \frac{2n(\mathbf{r})}{\Delta_\alpha} \mathcal{F}^{-1}[\Theta^a(k) \mathcal{F}[g^b[\mathbf{a}] g^c[\mathbf{a}]](\mathbf{k})](\mathbf{r}) + 2\,\text{perms}, \\ \tilde{\phi}_\alpha[\mathbf{a}](\mathbf{r}) &= \frac{2n(\mathbf{r})}{\Delta_\alpha} \mathcal{F}^{-1}[\Theta^a(k) \mathcal{F}[\tilde{g}^b[\mathbf{a}] \tilde{g}^c[\mathbf{a}]](\mathbf{k})](\mathbf{r}) + 2\,\text{perms}. \end{aligned} \tag{54}$$

Analogous definitions including the effects of pixelation and particle weights can be found in Eqs. (A16) and (A17).

The approach described above has several differences from standard approach: (i) We subtract a (zero-mean) bias term in Eq. (51), (ii) the data are weighted by $n\mathsf{H}^{-1}$ in Eq. (52), (iii) we include a symmetry factor Δ_α, and (iv) we normalize by a geometry-dependent factor $F_{\alpha\beta}$ rather than the bin volumes. The first three are included for the sake of optimality and arise naturally from the ML solution, while (iv) allows measurement of the unwindowed statistic. In Appendix B, we demonstrate that the above procedure reduces to conventional bispectrum estimators (see, e.g., [42]) in the limit of uniform, and large, $\bar{n}$ (B11), additionally noting that we have not subtracted a Poissonian shot-noise term.

C. Alternative bases

In the above discussion, we have assumed that the bispectrum can be fully parametrized by the lengths of three triangle sides, $\{k_1, k_2, k_3\}$. In reality, redshift-space distortions (RSDs) make the bispectrum anisotropic, giving additional dependence on the relative orientation of $\{\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3\}$ triplet and the (local) line of sight $\widehat{\mathbf{n}}$. Thanks to azimuthal symmetry about $\widehat{\mathbf{n}}$, this is specified by only two additional variables. A number of parametrizations of the bispectrum anisotropy are possible (see, e.g., [43,73,89]); here, we will consider that of Ref. [43]:

$$B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3; \widehat{\mathbf{n}}) \approx \sum_\alpha \frac{1}{\Delta_\alpha} b_\alpha [\Theta^a(k_1)\Theta^b(k_2)\Theta^c(k_3)\mathcal{L}_\ell(\widehat{\mathbf{k}}_3 \cdot \widehat{\mathbf{n}}) + 5\,\text{perms}], \tag{55}$$

where $\mathcal{L}_\ell$ is a Legendre polynomial, $\widehat{\mathbf{k}}_3 \cdot \widehat{\mathbf{n}}$ is the angle between $\mathbf{k}_3$ and the line of sight $\widehat{\mathbf{n}}$, and we have neglected dependence on the second angular coordinate for simplicity. In this case, the binning is now specified by four numbers: $\alpha = \{a, b, c, \ell\}$, with $\ell = 0$ reproducing the bispectrum monopole of Sec. IV C.

Computation of the bispectrum multipoles b_α appearing in Eq. (55) is possible via similar methods to before. In particular, the cumulant derivative becomes

$$\begin{aligned}\mathsf{B}^{ijk}_{,\alpha} = \frac{n(\mathbf{r}_i)n(\mathbf{r}_j)n(\mathbf{r}_k)}{\Delta_\alpha} \int_{\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3} &[\Theta^a(k_1)\Theta^b(k_2)\Theta^c(k_3)\mathcal{L}_\ell(\widehat{\mathbf{k}}_3 \cdot \widehat{\mathbf{r}}_k) + 5\,\text{perms}] e^{i\mathbf{k}_1\cdot\mathbf{r}_i + i\mathbf{k}_2\cdot\mathbf{r}_j + i\mathbf{k}_3\cdot\mathbf{r}_k} \\ &\times (2\pi)^3 \delta_{\rm D}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3),\end{aligned} \tag{56}$$

setting the local line of sight to $\widehat{\mathbf{n}} = \widehat{\mathbf{r}}_k$ (as in the Yamamoto approximation of Ref. [93]). Only the $\mathbf{k}_3$ part is more difficult to compute: This leads to the form

$$\hat{q}_\alpha = \frac{1}{\Delta_\alpha} \int d\mathbf{r} \{ g^a[\mathbf{d}](\mathbf{r}) g^b[\mathbf{d}](\mathbf{r}) g^{c,\ell}[\mathbf{d}](\mathbf{r}) - (g^a[\mathbf{d}](\mathbf{r}) \langle \tilde{g}^b[\mathbf{a}](\mathbf{r}) g^{c,\ell}[\mathbf{a}](\mathbf{r}) \rangle + 2\,\text{perms}) \}, \tag{57}$$

with

$$\begin{aligned} g^{a,\ell}[\mathbf{y}](\mathbf{r}) &= \int_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{r}} \Theta^a(k) \int d\mathbf{r}' n(\mathbf{r}') [\mathsf{H}^{-1}\mathbf{y}](\mathbf{r}') \mathcal{L}_\ell(\widehat{\mathbf{k}} \cdot \widehat{\mathbf{r}}') \\ &= \frac{4\pi}{2\ell+1} \sum_m \mathcal{F}^{-1}[\Theta^a(k) Y^*_{\ell m}(\widehat{\mathbf{k}}) \mathcal{F}[n\mathsf{H}^{-1}\mathbf{y} Y_{\ell m}](\mathbf{k})], \end{aligned} \tag{58}$$

expanding the Legendre polynomial in terms of spherical harmonics. The Fisher matrix can be computed similarly. This estimator gives a practical manner with which to estimate the bispectrum multipoles, and we note that its computation time is not significantly greater than that of the isotropic estimator, except due to the larger number of bins.

The estimators of Sec. IV A may also be applied to the scenario in which the bispectrum is represented as a sum of templates [e.g., Eq. (37)] rather than a large number of binned estimates (cf. Ref. [74]). While a suitably chosen decomposition scheme can significantly reduce the number of bispectrum elements (and, thus, the computation time) [71], this is nontrivial, since we require the templates to follow the separable form of Eq. (40). A simple approach would be to measure only the galaxy bias parameters from the bispectrum, assuming tree-level perturbation theory to be valid. In this case, the bispectrum *is* separable into a sum of ∼6 distinct components [58,66]. However, this decomposition becomes more difficult when higher-loop effects, redshift-space distortions, and the Alcock-Paczynski effect are included or when cosmological parameters must also be constrained. It seems likely that approximate methods such

modal decompositions [63] may assist with this; we leave a comprehensive discussion of this and other subtleties to future work.

V. PRACTICAL IMPLEMENTATION

While Sec. IV gives the explicit formulas required to apply the bispectrum estimators to a galaxy survey, we are left with a number of practical questions, including how to estimate the underlying field $n(\mathbf{r})$ and how to efficiently implement the $\hat{q}_\alpha$ and $F_{\alpha\beta}$ estimators. We discuss this below, as well as details pertaining to our simulations and computation.

A. Computation strategy

To compute the first term in $\hat{q}_\alpha$ (51), we require the following intermediate quantities.

(i) $\mathsf{H}^{-1}\mathbf{d}$.—Assuming FKP weights, this can be computed via Eq. (A10) and requires four FFTs, or none, if pixelation effects are ignored (12). For ML weights with $\mathsf{H} = \mathsf{C}$, the application of the inverse is performed using conjugate gradient descent (CGD), as in Ref. [74]. This generally converges in a few tens of steps and requires repeated computation of $\mathsf{C}\mathbf{x}$ for some map $\mathbf{x}$. This is achieved using chained FFTs, as discussed in Appendix A 1.

(ii) $g^a[\mathbf{d}](\mathbf{r})$.—Using the pixelized definition given in Eq. (A16), we require $N_k + 3$ FFTs to compute all relevant g^a maps given $\mathsf{H}^{-1}\mathbf{d}$ (assuming N_k bins per dimension). This reduces to $N_k + 1$ FFTs if pixelation effects are not included. Assuming each map contains $N_{\rm pix}$ pixels and is stored as an array of floats, the set of all g^a maps requires $4N_kN_{\rm pix}$ bytes of memory, which may be large for high-resolution maps or fine binning. As an example, using 512 grid points per dimension with 30 k bins leads to a total memory requirement of ≈2 GB.

(iii) *Contribution to* $\hat{q}_\alpha$.—As in Eq. (51), this can be computed by performing (inexpensive) real-space sums over products of three g^a functions, subject to the triangle conditions on $\{a, b, c\}$. In total, this gives $N_{\rm bins} \sim N_k^3$ scalar coefficients. Depending on the machine requirements, one may need to store the g^a functions to disk and then load each sequentially to compute $\hat{q}_\alpha$.

The second term in $\hat{q}_\alpha$ requires the average of pairs of $g^a[\mathbf{a}]$ and $\tilde{g}^a[\mathbf{a}]$ functions. We use uniformly distributed particles to define $\{\mathbf{a}\}$, created by Poisson sampling a spatially invariant background number density (here set to $\bar{n} = 10^{-3}h^3$ Mpc^{-3}) and then assigning the sampled particles to a grid, with the normalization chosen such that $\langle\mathbf{a}\rangle = \mathbf{0}$. Given $\mathbf{a}$, we estimate both $g^a[\mathbf{a}]$ and $\tilde{g}^a[\mathbf{a}]$ filters via Eq. (52), using the exact form of A^{-1} given in Eq. (A12), whose implementation requires two FFTs. The pairwise product of these filters are saved to disk and can then be used in combination with the data $g^a[\mathbf{d}]$ maps to compute $\hat{q}_\alpha$ via Eq. (51). In practice, we use $N_{\rm mc} = 50$ realizations to define the average $\langle g^b[\mathbf{a}]\tilde{g}^c[\mathbf{a}]\rangle$; since the error scales as $\sqrt{1 + 1/N_{\rm mc}}$, this is sufficient to give errors well below the statistical threshold.

The Fisher matrix is obtained in a similar fashion using the same set of $\mathbf{a}$ maps. Given the previously computed $g^a[\mathbf{a}]$ and $\tilde{g}^a[\mathbf{a}]$ maps, we further compute the following.

(i) $\phi_\alpha[\mathbf{a}]$.—By Eq. (54), this is simply computed using two Fourier transforms per $\{a, b, c\}$ permutation or four when the pixelation effects are included [Eq. (A17)]. Since the Fisher matrix requires $\langle\phi_\alpha[\mathbf{a}]\rangle$, a simple approach would be to save each map to disk. We caution that this is *not* practically achievable in many contexts, since it requires $4N_{\rm bins}N_{\rm pix}N_{\rm mc}$ of (temporary) storage if all bins are computed at once. In practice, one needs only to save the partial sum of all computed maps, greatly reducing the storage requirements.

(ii) $\tilde{\phi}_\alpha[\mathbf{a}]$.—This is computed analogously to $\phi_\alpha[\mathbf{a}]$ but using the $\tilde{g}^a[\mathbf{a}]$ filters.

(iii) $\mathsf{H}^{-1}\phi_\alpha[\mathbf{a}]$.—The action of H^{-1} on the ϕ_α maps can be computed using FFTs and, if ML weights are applied, CGD inversion. Since $N_{\rm bins}(N_{\rm mc} + 1)$ such maps must be computed, this is the rate-limiting step of the ML algorithm, as each requires $\mathcal{O}(100)$ FFTs.

(iv) $F_{\alpha\beta}$.—The Fisher matrix is constructed using Eq. (53), requiring $\langle\phi_\alpha\mathsf{H}^{-1}\tilde{\phi}_\beta\rangle$ and $\langle\phi_\alpha\rangle\mathsf{H}^{-1}\langle\tilde{\phi}_\beta\rangle$. The MC averages can be accumulated from each a simulation and stored as a matrix of size $N_{\rm bins} \times N_{\rm bins}$.

It is clear that computation of the bispectrum using cubic estimators is a relatively intensive procedure. The principal work lies in computing g^a and $\mathsf{H}^{-1}\phi_\alpha$ maps, requiring $\mathcal{O}(N_k)$ and $\mathcal{O}(N_{\rm bins})$ FFTs, respectively. The latter dominates in practice, particularly when using ML weights, since we must apply an inverse matrix H^{-1} to each of $N_{\rm mc}$ random simulations in $N_{\rm bins}$ bins. However, computation of the Fisher matrix (and the $\langle g^b[\mathbf{a}]\tilde{g}^c[\mathbf{a}]\rangle$ bias term) is *independent* of both data and any N-body simulations, with its estimation requiring only the background number density $n(\mathbf{r})$ and the fiducial band powers (if using ML weights). Usually, one computes summary statistics on both the data and a large number of mock catalogs; each simulation requires only computation of $\hat{q}_\alpha$ and is, thus, fast.

B. Mock catalogs

Foreshadowing the eventual application of the above estimators to the BOSS DR12 galaxy sample [98], we first consider their use on a set of mock galaxy catalogs taken from the MultiDark-Patchy (hereafter "Patchy") suite [99,100].[6] These share the BOSS geometry and selection functions

[6]These are publicly available [101].

and are calibrated to have a similar power spectrum and bispectrum to the observational data. As in Ref. [74], we use only the north Galactic cap with the "z1" redshift cut $0.2 < z < 0.5$ (see, e.g., [87,102]). This has mean redshift $z = 0.38$ and a total volume $V = 1.46h^{-3}$ Gpc^3, containing $\sim 5 \times 10^5$ galaxies. We additionally use a Patchy random catalog which has the same selection functions as the data but 50× larger particle density. The observed redshifts and angular positions of both galaxies and randoms are converted to Cartesian coordinates using the fiducial matter density $\Omega_m = 0.31$ [102], before they are painted to a uniform grid using a triangular-shaped cloud mass assignment scheme, implemented using NBODYKIT [103]. The discretization grid uses pixels of twice the width as those in the official BOSS release; this is more than sufficient for our purposes (cf. Sec. VI) and gives a Nyquist frequency $k_{\rm Nyq} = 0.3h$ Mpc^{-1}. The data vector $\mathbf{d}$ is then computed as the difference between galaxies and (rescaled) random particles, including the incompleteness weights w_c of Ref. [102]. Note that we do not add FKP weights at this point.

To define the background density field $n(\mathbf{r})$, we make use of the publicly available survey mask files as well as the redshift distribution $\bar{n}(z)$ from the Patchy random catalogs (themselves calibrated from the BOSS data). In practice, we first histogram the randoms in redshift bins (weighted by w_c) before transforming these to volumetric bins via the Jacobian $dV/dz = 4\pi r^2(z)dr/dz$, where $r(z)$ is the comoving distance to redshift z. We then evaluate $\bar{n}(z(\mathbf{r}))$ at the center of each pixel, before multiplying by the angular MANGLE mask $\phi(\hat{\mathbf{r}})$ and normalizing such that $\int d\mathbf{r} n^3(\mathbf{r})$ matches the value obtained from the random catalog (including completeness weights). This differs from the approach of Ref. [74], which assumed $n(\mathbf{r})$ to be equal to the pixelized field of randoms, and is preferred since (a) it does not include pixelation effects and (b) it gives a smooth field; i.e., there are no regions within the survey area where $n(\mathbf{r}) = 0$, which can cause instabilities upon inversion.

To apply the ML estimators, we additionally require a fiducial power spectrum for the signal covariance, as in Eq. (A7). Since the bispectrum estimator is unbiased for all choices of invertible H, the precise power spectrum is not critical, though the procedure is suboptimal if it is far from the true power spectrum.[7] Here, we follow Ref. [74] and model the unwindowed power spectra using a fit to 1000 Patchy simulations using CLASS-PT [104]. No fiducial spectrum is required when using FKP weights, i.e., $\mathsf{H} = \mathsf{H}_{\rm FKP}$.

[7]This is unlike the approach of Ref. [74], whereupon the fiducial spectrum was a crucial component of the power spectrum estimator. This was a consequence of using the first form of Eqs. II A and (8) which contains explicit dependence on a fiducial set of band powers. In this work, we focus on the second form, which does not require such considerations.

C. Unwindowed bispectrum computation

Given the above, we may apply the bispectrum estimators of Secs. III and IV to the Patchy simulations. For this demonstration, we adopt a k-space binning with width $\Delta k = 0.01h$ Mpc^{-1} from $k_{\rm min} = 0h$ Mpc^{-1} to $k_{\rm max} = 0.16h$ Mpc^{-1}, giving $N_k = 16$. Allowing for the effects of finite bin widths, this gives $N_{\rm bins} = 508$ triplets obeying the triangle conditions, i.e., containing wave vectors $\{\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3\}$ satisfying $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = \mathbf{0}$.[8] After computation, we discard any bins with $k < 0.01h$ Mpc^{-1} or $k > 0.15h$ Mpc^{-1}, since these are not fully window corrected (cf. [74]); this leaves $N_{\rm bins} = 399$ elements. A finer binning would lead to a much larger dimensionality; this does not pose problems in the analysis if a compression scheme is applied before the likelihood is computed (see, e.g., [52,71]), though it will increase the computation time.

Here, we compute the bispectrum of 999 Patchy simulations, using $N_{\rm mc} = 50$ uniform random simulations to define the linear $\hat{q}_\alpha$ term and the Fisher matrix. For our fiducial run, we assume an FKP weighting matrix (12) and do not include the forward modeling of pixelation effects described in Appendix A.[9] To test the various hyperparameters, we consider a number of alternative runs, each with 100 simulations: (i) using a coarser pixel grid with $k_{\rm Nyq} = 0.2h$ Mpc^{-1}, (ii) using $N_{\rm mc} = 100$ MC simulations, (iii) including the full pixelation effects of Appendix A, (iv) omitting the second, variance reducing, term in Eq. (51), and (v) using ML weights. The impact of these assumptions will be assessed in Sec. VI B.

All computations are performed in PYTHON, making extensive use of the pyfftw package to perform FFTs.[10] When using FKP (ML) weights, we require 100 (140) CPU hours to compute the data bispectrum contributions from all 999 mocks (matching the computation time of any standard bispectrum estimator), plus an additional 150 (850) CPU hours in total for the uniform random contributions. While non-negligible, this run time is not unreasonable, especially considering that the bulk of the time is spent computing the Fisher matrix which needs to be done only once. This is further reduced if the bin width is increased or when using a coarser pixel grid.

[8]When comparing data and theory, it is common to evaluate the theory model in the center of each k bin, i.e., at $\{\bar{k}_1, \bar{k}_2, \bar{k}_3\}$. We caution that some of our allowed bins violate the triangle conditions on $\bar{k}_i$; i.e., they do not satisfy $|\bar{k}_1 - \bar{k}_2| \leq \bar{k}_3 \leq \bar{k}_1 + \bar{k}_2$ and should not be included in the fit. If the theory model is integrated over a finer grid of wave numbers, these bins are important to include, however.

[9]We do, however, include a factor $\tilde{\psi}^{-1}(\mathbf{k}_1)\tilde{\psi}^{-1}(\mathbf{k}_2)\tilde{\psi}^{-1}(\mathbf{k}_3)$ in the $\hat{q}_\alpha$ estimator (where $\tilde{\psi}$ is the Fourier-space pixelation window) to ensure that our bispectra are not biased. This is further discussed in Appendix A 2.

[10]Our code is publicly available [91].

D. Windowed bispectrum estimators

To assess the impact of the survey window function on the bispectrum, we additionally compute the statistic using conventional estimators. In this case, the output bispectrum is binned, FKP weighted (before gridding), and convolved with the survey window function. Following Ref. [27], the windowed bispectrum of data d can be written

$$\hat{b}_\alpha^{\rm win} = \frac{1}{I_3}\frac{1}{V_\alpha}\int_{\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3}(2\pi)^3\delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)d(\mathbf{k}_1) \times d(\mathbf{k}_2)d(\mathbf{k}_3)\Theta^a(k_1)\Theta^b(k_2)\Theta^c(k_3), \tag{59}$$

where the Dirac delta function enforces momentum conservation and the bin volume is given by

$$V_\alpha = \int_{\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3}(2\pi)^3\delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\Theta^a(k_1)\Theta^b(k_2)\Theta^c(k_3). \tag{60}$$

Equation (59) also includes the normalization factor $I_3 = \int d\mathbf{r} n^3(\mathbf{r}) w_{\rm FKP}^3(\mathbf{r})$, where $w_{\rm FKP}(\mathbf{r}) = [1+n(\mathbf{r})P_{\rm FKP}]^{-1}$ is the FKP weight. This can be computed from the random catalog (requiring the inclusion of completeness weights) as in Ref. [27]. Rewriting the Dirac functions in Eqs. (59) and (60) as complex exponentials as before gives a more practical form for the estimator:

$$\hat{b}_\alpha^{\rm win} = \frac{1}{I_3}\frac{\int d\mathbf{r}\mathcal{D}_a(\mathbf{r})\mathcal{D}_b(\mathbf{r})\mathcal{D}_c(\mathbf{r})}{\int d\mathbf{r}\mathcal{U}_a(\mathbf{r})\mathcal{U}_b(\mathbf{r})\mathcal{U}_c(\mathbf{r})}, \tag{61}$$

defining

$$\mathcal{D}_a(\mathbf{r}) = \int_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{r}} d(\mathbf{k})\Theta_a(\mathbf{k}), \quad \mathcal{U}_a(\mathbf{r}) = \int_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{r}}\Theta_a(\mathbf{k}), \tag{62}$$

the set of which can be computed using $(2N_k+1)$ FFTs. Here, we apply the estimator to 999 Patchy simulations, which requires ∼100 CPU hours in total.

VI. RESULTS

A. Bispectrum estimates

In Fig. 1, we display the main results of this work—a set of binned bispectrum measurements from 999 Patchy mocks using FKP weights (12), which are not convolved with the survey window function. The measurements exhibit a distinctive "sawtooth" pattern (analogous to those of Ref. [27]), which is a result of projecting a 3D dataset into one dimension. In full, this arises since $k_1k_2k_3B(k_1,k_2,k_3)$ is an increasing function of k_i. Considering the fractional error, it is clear that we have detected a nonzero bispectrum at high significance, with most bins having a signal to noise above unity (ignoring correlations). We note a clear distinction between the three types of triangles shown in Fig. 1; scalene triangles exhibit significantly smaller error bars than those of isosceles configurations, themselves smaller than the equilateral triangles error bars. This arises due to the number of ways in which one can obtain a triangle of a given configuration [i.e., Δ_α, Eq. (49)] and matches that found in previous works (e.g., Ref. [69]).

Also shown in Fig. 1 are the *windowed* bispectrum measurements from Patchy, computed using the estimators of Sec. V D. On short scales (high k, corresponding to the rightmost data points in Fig. 1), the mean bispectrum measurements from the unwindowed and windowed estimators appear highly consistent, implying that the window function's impact is minimal. Moving to larger scales (low k), the window function distortions become significant, particularly for squeezed triangles (containing two long sides and a short side). Given that the BOSS window function primarily contains power on large scales [74], this matches our expectations. Some of the most important bispectrum parameters (such as $f_{\rm NL}$) have signatures concentrated on large scales; thus, these results highlight the importance of a proper window function treatment.

To further assess the effects of window function convolution, it is useful to look at the Fisher matrix $F_{\alpha\beta}$ [Eq. (53)]. In the limit of an ideal survey geometry (Appendix B), the matrix is diagonal and simply encodes the bin volumes and normalization; outside this limit, it has a more complex structure, including off-diagonal contributions, which act to deconvolve the bispectrum (cf. Ref. [75]). In Fig. 2, we plot the Fisher matrix for the Patchy simulations, again assuming $\mathsf{H} = \mathsf{H}_{\rm FKP}$. Here, we find off-diagonal correlations up to ∼10% in neighboring bins (i.e., those in which a single element of $\alpha = \{a,b,c\}$ and $\beta = \{a',b',c'\}$ differs by one) but small corrections elsewhere. This indicates that the primary effect of the window function is to convolve neighboring bins (due to the compact nature of the window function) and motivates our choice of scale cuts (Sec. V C).

From the fractional errors shown in Fig. 1, it may seem that the windowed bispectrum estimator achieves significantly higher precision than the unwindowed estimator. In fact, this is not the case, since the individual windowed data points are far more correlated than their unwindowed equivalents, an effect sourced by the nonuniform geometry. As a demonstration, we plot the bispectrum correlation matrices (equal to the covariance matrices normalized by their diagonals) in Fig. 3. In the unwindowed case, the matrix is close to diagonal, though there remain some off-diagonal correlations in closely separated bins.[11]

[11]For ML weights, and in the limit of zero bispectrum, the covariance matrix is equal to the inverse of the Fisher matrix [cf. Eq. (33)].

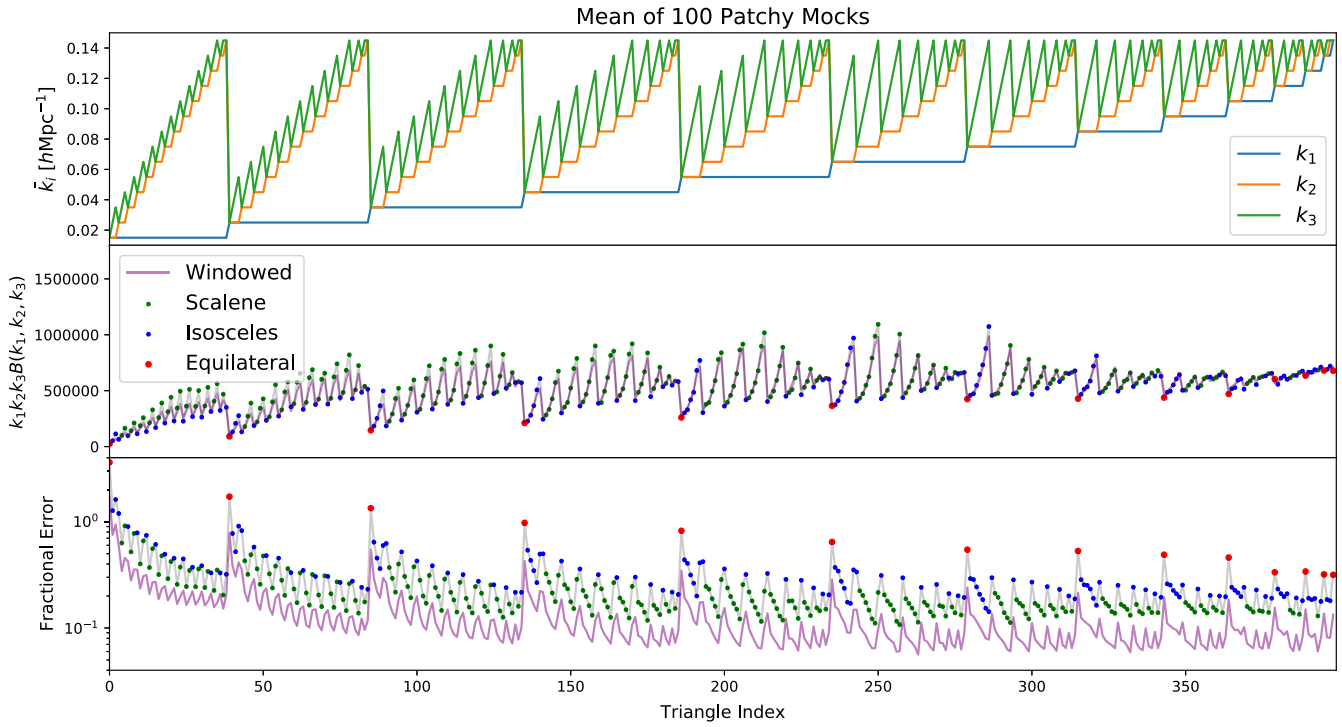

FIG. 1. Unwindowed bispectrum measurements from 999 Patchy simulations at $z = 0.57$, obtained using the cubic estimators of this work. In the first panel, we show the bin-averaged triangle sides $\{k_1, k_2, k_3\}$ corresponding to each one-dimension wavenumber bin, with the second panel giving the associated binned bispectrum estimates (normalized by $k_1k_2k_3$), averaged across all mocks. The third panel displays the fractional error in the bispectrum relevant for a single mock dataset. We show triangle configurations corresponding to scalene ($k_1 \neq k_2 \neq k_3$), isosceles ($k_1 = k_2 < k_3$ or $k_1 < k_2 = k_3$) and equilateral ($k_1 = k_2 = k_3$) triangles in green, blue and red respectively; the error-bars for isosceles (equilateral) triangles are inflated by a factor $\sim$2 ($\sim$6) as expected. To compute these bispectra, we assume an FKP weighting (12), a Nyquist frequency of $k_{\rm Nyq} = 0.3h$ Mpc^{-1}, $N_{\rm mc} = 50$ Monte Carlo simulations, and do not forward model pixelation effects (cf. Appendix A). We additionally show results from the *windowed* bispectrum estimator in purple; whilst these have smaller fractional errors, this is primarily due to increased bin-to-bin covariances, as demonstrated in Fig. 3. The figure shows results only for the bispectrum monopole (averaged over triangle rotations); the estimators of this work could be simply extended to include higher multipoles sourced by redshift-space distortions.

In contrast, we observe a much increased correlation structure in the windowed estimates, with correlation coefficients approaching unity at low k (where the effects of the window function are largest). This highlights the utility of the windowed estimator and matches the conclusions of Ref. [74] for the power spectrum.[12]

B. Dependence on hyperparameters

To test the dependence of the output spectra on the algorithm hyperparameters, we perform a number of additional bispectrum analyses, as noted in Sec. V C. First, Fig. 4 considers the effects of using a coarser Fourier-space grid, reducing the Nyquist frequency from $k_{\rm Nyq} = 0.3h$ Mpc^{-1} to

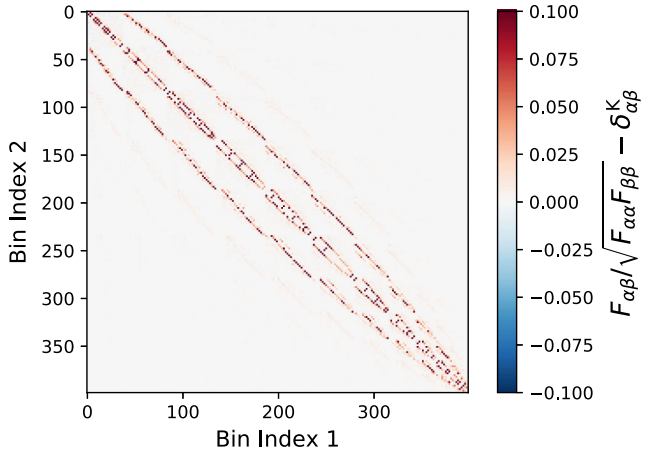

FIG. 2. Fisher matrix for the unwindowed bispectrum estimates shown in Fig. 1. As discussed in the main text, this removes effects of the survey window function, practically acting as a deconvolution matrix. Here, we plot the matrix using the same binning scheme as Fig. 1, normalizing by the diagonal elements for clarity. The largest correlations are between bins $k_1' = k_1 \pm \Delta k$, $k_2' = k_2$, $k_3' = k_3$ (or permutations); other correlations are found to be small.

[12]Given that the bispectra obtained from the estimators of Sec. IV are not convolved with the window function, it may be tempting to model their covariance perturbatively, ignoring window effects. This is not true, however, since the impact of the window on the covariance is different to that on the data.

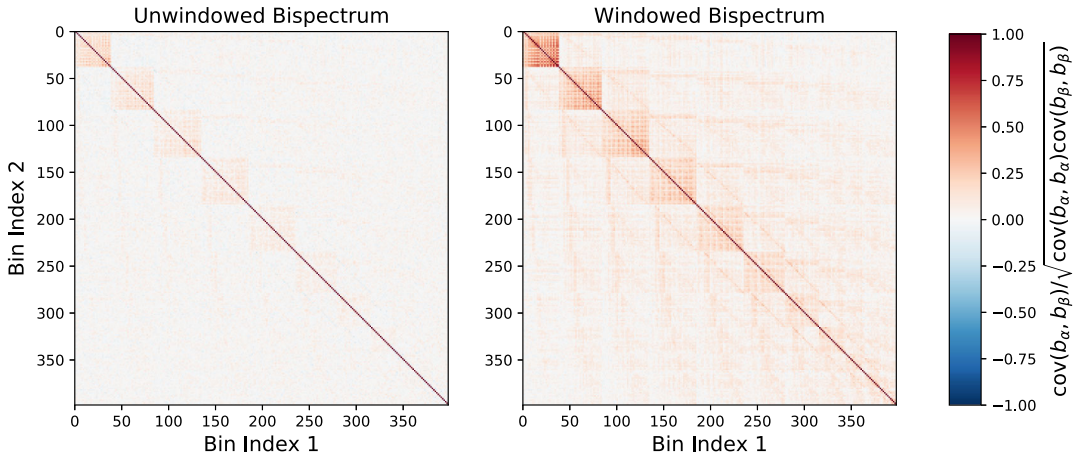

FIG. 3. Correlation matrices for the unwindowed and windowed bispectrum measurements shown in Fig. 1. These are computed from the sample covariance of 999 individual Patchy bispectra and follow the same bin ordering as Fig. 1. While the unwindowed spectra have a close-to-diagonal covariance, this is not true for the windowed measurements and highlights the smoothing effect of the window function.

$0.2h$ Mpc^{-1} (which can be compared to $k_{\rm max} = 0.15h$ Mpc^{-1}). Though the noise properties differ somewhat, we find consistent results between the two datasets across all triangle bins, except those in which all three triangle sides have large wave numbers. This matches our expectations and implies that the fiducial value of $k_{\rm Nyq} = 0.3h$ Mpc^{-1} is more than sufficient for our choice of $k_{\rm max}$.

Figure 5 assesses the impact of Monte Carlo noise. As noted in Sec. IV, we use a set of $N_{\rm mc} = 50$ uniform random realizations to compute the Fisher matrix (and the second term of $\hat{q}_\alpha$), in order to sidestep a computationally prohibitive matrix multiplication. When $N_{\rm mc}$ is increased to 100, the results are statistically consistent, yet we find an $\sim$2% reduction in the error bar. Given that the expected error is $\sqrt{1 + 1/N_{\rm mc}}$, this matches expectations. In future runs, it may be desired to use a larger number of simulations than the fiducial $N_{\rm mc} = 50$, though we note that this is still a

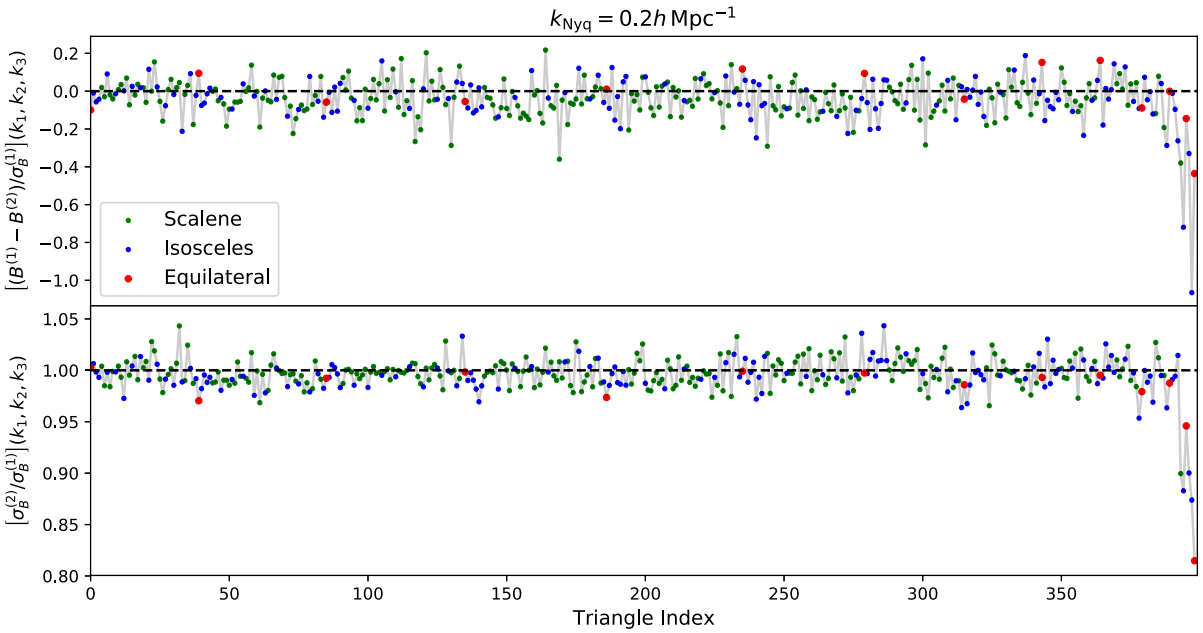

FIG. 4. Dependence of the unwindowed bispectra on the Fourier-space grid size. This compares two sets of bispectra: those computed with a Nyquist frequency $k_{\rm Nyq} = 0.3h$ Mpc^{-1} (shown in Fig. 1) and those with $k_{\rm Nyq} = 0.2h$ Mpc^{-1}. The top panel shows the difference between the two estimates as a fraction of the statistical error of the fiducial estimates, while, in the bottom panel, we give the ratio of statistical errors. The bins are ordered in the same fashion as Fig. 1, and both datasets include 100 Patchy simulations. Here, increasing the number of grid points is found to have an insignificant effect, except on the smallest scales.

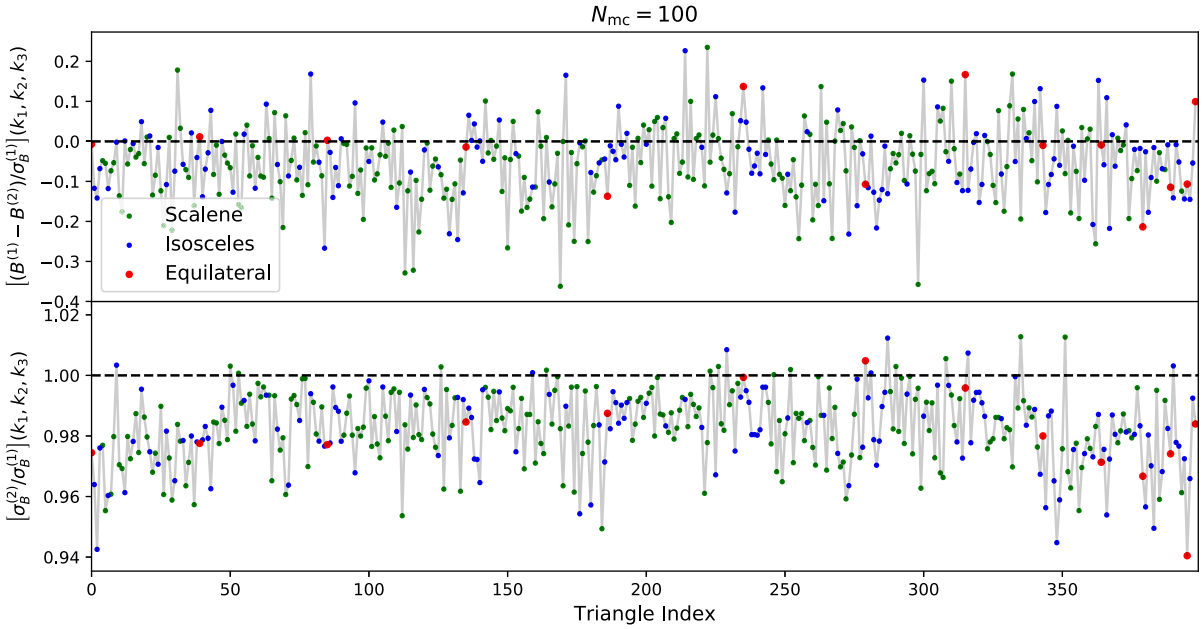

FIG. 5. As in Fig. 4, but assessing the impact of the number of Monte Carlo simulations used to define the Fisher matrix. Here we compare the case of $N_{\rm mc} = 100$ to the fiducial results with $N_{\rm mc} = 50$. The error is expected to scale as $\sqrt{1 + 1/N_{\rm mc}}$; this matches that found in the figure.

small (though multiplicative) contribution to the overall error budget.

In Appendix A, we present a full treatment of the two- and three-point cumulants of the *gridded* density field. This allows one to forward model the pixelation effects instead of simply dividing by the pixelation mask in the final step of the bispectrum estimator computation. Bispectra computed using this approach are compared to the fiducial set in Fig. 6. In this case, we do not find any marked improvement with this approach; rather, there appears to be a slight increase in the variances (though we caution that this may be sourced by a decrease in bin-to-bin correlations). While this approach may be of use for analyses with more complex window functions, it is not

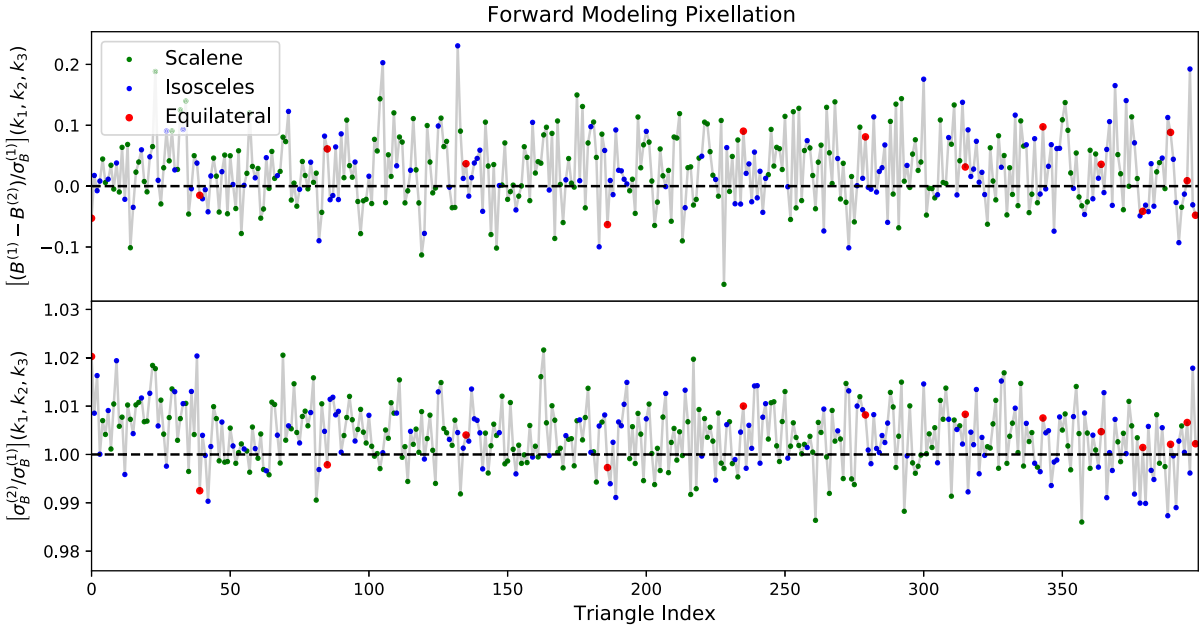

FIG. 6. As in Fig. 4, but testing the impact of forward modeling the pixelation effects, as described in Appendix A. This is found to have a very small effect in practice, since the survey size is much larger than the pixel width.

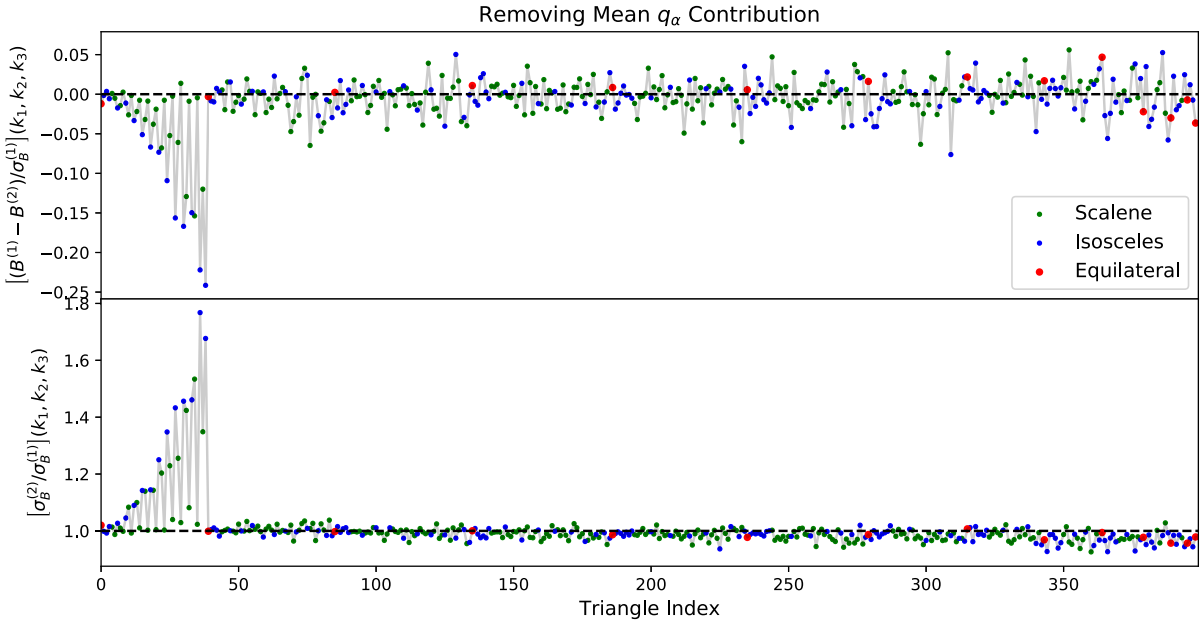

FIG. 7. As in Fig. 4, but assessing the importance of the second term in $\hat{q}_\alpha$ [Eq. (51)], which is ignored in conventional bispectrum estimators. Removal of this term is found to significantly increase the error on large scales; this is consistent with the expected behavior found in Appendix B and highlights the importance of this contribution in $f_{\rm NL}$-based studies.

found to be important here and is disfavored, since it requires more FFT operations.

An additional test concerns the second term appearing in the $\hat{q}_\alpha$ quantity, i.e., that involving $\mathsf{C}_{ij}^{-1}[\mathsf{C}^{-1}\mathbf{d}]_k$ (and permutations) in Eq. (25). On average, this is expected to be zero, and, furthermore, it contributes only to the $\mathbf{k} = \mathbf{0}$ mode for a uniform survey geometry (Appendix B). Figure 7 shows the effects of removing it from the

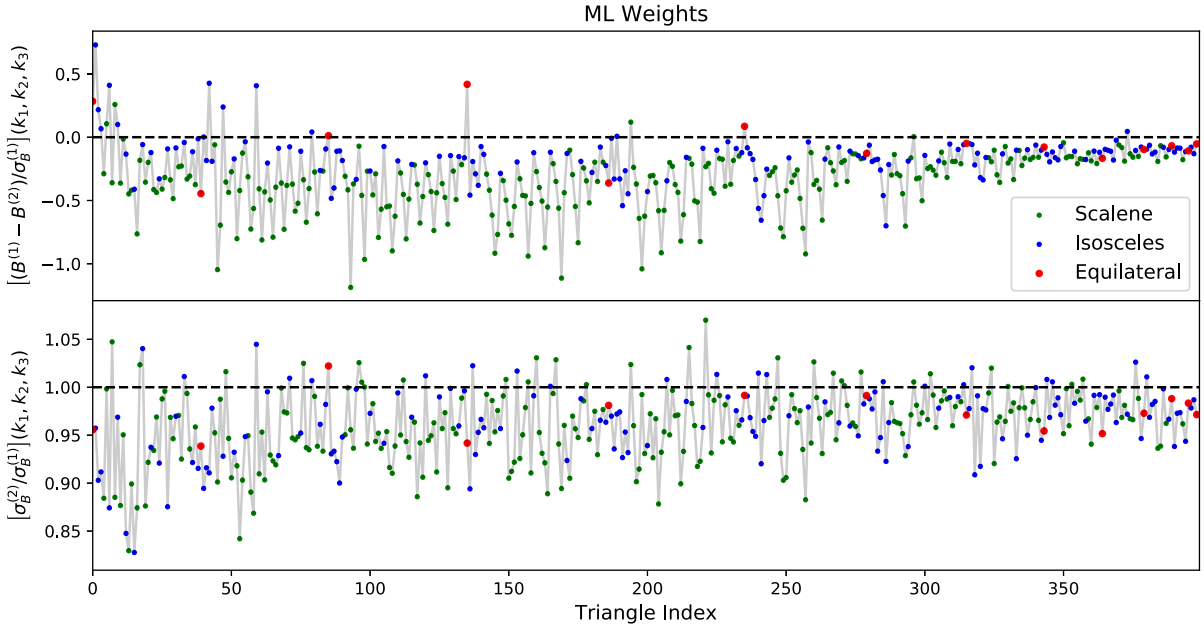

FIG. 8. As in Fig. 4, but adopting maximum-likelihood weights C^{-1} [Eq. (A4)] rather than the short-scale FKP approximation [Eq. (12)]. As in Ref. [74], this is not expected to significantly improve constraints from a BOSS-like survey, since the window function contains mostly large-scale power and the shot noise is large. Here, we find an ∼5% decrease in the error bars; however, this is associated with a nontrivial underprediction of the bispectrum. This may arise from the difficulty of inverting the survey mask (since it contains zeros, which do not quite match those of the dataset due to pixelation effects). In practice, we expect the difference to be absorbed in bispectrum bias parameters in any cosmological analysis; thus, this is unlikely to be a cause for concern.

bispectrum estimator; we find a large increase in the bispectrum variances in the lowest k bin (with one leg in the range $[0.01, 0.02]h$ Mpc^{-1}) but negligible impact to other bins.[13] We conclude that this term should certainly be included when the measuring large-scale bispectrum modes but is of little importance if these are not included in the analysis.

Finally, we assess the impact of using ML weights in the bispectrum estimator [rather than the FKP scheme of Eq. (12)]. A similar comparison was performed for the power spectrum in Ref. [74]; in that case, the ML weights were found to have negligible impact on both the output power spectra and the corresponding cosmological parameter constraints. This was attributed to the compact nature of the window function (i.e., with power concentrated on large scales), the relatively high shot noise, and the narrow k bins adopted. The analogous results for the bispectrum are shown in Fig. 8. Unlike for the power spectrum, we find a significant ($\sim$5%) reduction in the bispectrum variances, somewhat increasing on large scales. More troublingly, we find also a shift in the spectrum compared to that of FKP weights, which is not expected (since the bispectrum estimators should be unbiased for any choice of weighting matrix H). We hypothesize that this is caused by zeros in the survey mask and the effects of gridding; since the data are smoothed by the pixelation window but the underlying density $n(\mathbf{r})$ is not, there will be pixels in which $n(\mathbf{r})$ is zero but the data are nonzero. These can cause instabilities upon inversion, which will affect the ML and FKP weights differently, particularly impacting the ML weights due to their more complex character. We defer further consideration of this effect to future work; however, we note that the principle effect is that of a global renormalization, which may be degenerate with the galaxy bias parameters.

VII. SUMMARY

In the effort to harvest maximal information from current and future spectroscopic surveys, the bispectrum plays a key role. In combination with the galaxy power spectrum, it has been shown to significantly enhance constraining power on ΛCDM parameters [8,9], as well as strengthening the bounds on additional phenomena such as primordial non-Gaussianity, neutrino masses, and modified gravity [10–14]. However, several challenges must be overcome before its full potential can be realized, in particular, regarding its measurement, modeling, and dimensionality. In this work, we have introduced new estimators for the bispectrum, which, unlike previous approaches, output spectra which are not convolved with the survey window. These are analogous to the "quadratic estimators" used in power spectrum analyses of old (and Ref. [74]) and are constructed by maximizing a pixel likelihood based on the Edgeworth expansion. We provide general forms for the estimator using both optimal and FKP-like [92] weighting schemes, with the latter providing a simple-to-implement approximation to the ML solution. The general estimator consists of two parts: a cubic estimator applied to the data and a Fisher matrix computed from the survey mask, which acts as a deconvolution matrix (cf. Ref. [75]).

When considering the binned bispectrum monopole, the bispectrum estimators may be straightforwardly implemented using FFTs and, if optimal weights are adopted, conjugate gradient descent methods. Their computation requires only the data, a random catalog, and knowledge of the survey mask. The estimators were applied to a suite of Patchy simulations and shown to be efficient, with the window function corrections taking $\sim$150 (850) CPU hours to compute with FKP (ML) weights, irrespective of the number of mocks analyzed. Comparison of the windowed and unwindowed bispectra showed highly consistent results on small scales but significant differences at low k, due to the nonuniform survey geometry. Furthermore, the unwindowed bispectrum bins were found to be substantially less correlated than the windowed bins, simplifying their interpretation.

Measuring bispectra in this manner has a number of benefits over standard approaches. First and foremost, it avoids the need to window convolve the theory model in an MCMC analysis. This is a computationally expensive procedure (involving double Hankel transforms [73]); thus, its removal will significantly expedite parameter inference studies and obviate the need for survey-specific analysis pipelines. Second, the estimators approach the minimum-variance solution (assuming weak non-Gaussianity). While this is unlikely to be of importance for relatively uniform surveys such as BOSS [74] (with the ML weights, in fact, causing a slight bias due to pixelation effects and FKP weights being close to optimal), the unwindowed bispectrum estimator does feature an additional linear term which is shown to significantly reduce the error bars on the large-scale bispectrum. An additional use concerns the global integral constraint, arising from the unknown mean survey density. Since our measurements are, by definition, unwindowed, these effects should be restricted to the first k bin and, thus, may be ignored if this is not included in the analysis.

There are a number of ways in which the ideas of this work can be extended. In particular, one may wish to construct analogous estimators for the *anisotropic* bispectrum multipoles [43,73,89,90]. This follows the procedure noted in Sec. IV C and will allow all the information present in the three-point statistic to be captured. More abstract bases for the bispectrum would also be of use; essentially, one decomposes the bispectrum into a set of separable shapes, whose amplitudes can be directly estimated. For the simple case of constraining galaxy biases from a tree-level model, such a formalism exists [58,105], though we caution that finding such a decomposition in a

[13]There is also a slight bias, though we again caution that error bars are most strongly correlated on these scales.

more general case may be difficult. An additional application would be to compute windowless *skew spectra* using cubic estimators; such statistics have been shown to capture equivalent information to the galaxy bispectra [105,106] in a friendlier form. Finally, an important application of the windowless bispectra presented herein is to the BOSS galaxy sample. Combining the above measurements with a bispectrum theory model in a full-shape framework will tighten parameter constraints from power-spectrum-based analyses [87,107], further pinning down the Universe's composition and evolution.

ACKNOWLEDGMENTS

It is a pleasure to thank Giovanni Cabass, Emanuele Castorina, Mikhail Ivanov, Azadeh Moradinezhad Dizgah, Marko Simonović, David Spergel, and Matias Zaldarriaga for insightful conversations without which this work would not have been possible. I am additionally grateful for insightful comments from the referee. O. H. E. P. acknowledges funding from the Nancy Grace Roman Space Telescope program through NNG26PJ30C and NNN12AA01C and thanks the Simons Foundation for additional support. The author is pleased to acknowledge that the work reported in this paper was substantially performed using the Princeton Research Computing resources at Princeton University, which is a consortium of groups led by the Princeton Institute for Computational Science and Engineering (PICSciE) and the Office of Information Technology's Research Computing Division. Additional computations were performed on the Helios cluster at the Institute for Advanced Study.

APPENDIX A: PIXEL CORRELATION FUNCTIONS

Below, we present a full discussion of the pixel correlators used in this work, including the effects of discretization, particle weights, and redshift-space distortions. Our covariance modeling represents an improvement to that of Ref. [74], particularly with regards to the treatment of pixelation, which was previously ignored.[14]

First, we define the data and random density fields as weighted sums over Dirac delta functions:

$$\hat{n}'_g(\mathbf{r}) = \sum_a w_d^a \delta_{\rm D}(\mathbf{r}-\mathbf{r}_a), \qquad \hat{n}'_r(\mathbf{r}) = \alpha \sum_b w_r^b \delta_{\rm D}(\mathbf{r}-\mathbf{r}_b), \tag{A1}$$

where a and b represent data and random particle indices, respectively, and $\alpha \equiv \sum_a w_d^a / \sum_b w_r^b$. We add primes to indicate weighted fields. Under a Poisson average, $\langle \hat{n}'_g(\mathbf{r})\rangle = n(\mathbf{r})[1+\delta(\mathbf{r})]$ and $\langle \alpha \hat{n}'_r(\mathbf{r})\rangle = n(\mathbf{r})$, where $n(\mathbf{r})$ is the background number density (assuming that the weights correct for sampling errors), i.e., $n(\mathbf{r}) = \lim_{\alpha\to 0}\alpha \hat{n}_r(\mathbf{r})$.

After painting the density fields to a grid, the overdensity in pixel i is given by

$$\begin{aligned} d^i &= \int d\mathbf{r}[\hat{n}'_g(\mathbf{r}) - \alpha \hat{n}'_r(\mathbf{r})]\psi(\mathbf{r}-\mathbf{r}_i) \\ &= \sum_a w_d^a \psi(\mathbf{r}_i - \mathbf{r}_a) - \alpha \sum_b w_r^b \psi(\mathbf{r}_i - \mathbf{r}_b), \end{aligned} \tag{A2}$$

where $\psi(\mathbf{r})$ is some mass-assignment (compensation) function with compact support. A typical choice for this is the "triangle-shaped-cloud" interpolation scheme, whose functional form can be found in Ref. [108].

1. Two-point statistics

Given the pixelized density field (A2), we may construct the two-point covariance $\mathsf{C}^{ij} \equiv \langle d^i d^j\rangle$. To do so, we first define the pairwise expectations of the weighted fields, analogously to Eq. (9):

$$\begin{aligned} \langle \hat{n}'_g(\mathbf{r})\hat{n}'_g(\mathbf{r}')\rangle &= n(\mathbf{r})n(\mathbf{r}')[1+\xi(\mathbf{r}-\mathbf{r}')] \\ &\quad + \delta_D(\mathbf{r}-\mathbf{r}')\left\langle \sum_a (w_d^a)^2 \delta_D(\mathbf{r}-\mathbf{r}_a)\right\rangle, \\ \langle \alpha \hat{n}'_r(\mathbf{r})\hat{n}'_g(\mathbf{r}')\rangle &= n(\mathbf{r})n(\mathbf{r}'), \\ \langle \alpha^2 \hat{n}'_r(\mathbf{r})\hat{n}'_r(\mathbf{r}')\rangle &= n(\mathbf{r})n(\mathbf{r}') \\ &\quad + \alpha^2 \delta_D(\mathbf{r}-\mathbf{r}')\left\langle \sum_b (w_r^b)^2 \delta_D(\mathbf{r}-\mathbf{r}_b)\right\rangle, \end{aligned} \tag{A3}$$

giving

$$\begin{aligned} \langle d^i d^j\rangle &\equiv \mathsf{C}^{ij} = \mathsf{S}^{ij} + \mathsf{N}^{ij}, \\ \mathsf{S}^{ij} &= \int d\mathbf{r} d\mathbf{r}' n(\mathbf{r}) n(\mathbf{r}') \xi(\mathbf{r},\mathbf{r}') \psi(\mathbf{r}-\mathbf{r}_i)\psi(\mathbf{r}'-\mathbf{r}_j), \\ \mathsf{N}^{ij} &\equiv \int d\mathbf{r} n'(\mathbf{r}) \psi(\mathbf{r}-\mathbf{r}_i)\psi(\mathbf{r}-\mathbf{r}_j)[1+\alpha^2/\beta], \end{aligned} \tag{A4}$$

where $\xi(\mathbf{r},\mathbf{r}')$ is the 2PCF of the underlying field δ. Notably, the gridded number density field N is not diagonal in pixel space. Equation (A4) uses the twice-weighted field $n'(\mathbf{r})$, defined by

$$n'(\mathbf{r}) = \lim_{\beta\to 0}\beta \sum_b (w_r^b)^2 \delta_D(\mathbf{r}-\mathbf{r}_b) = \left\langle \sum_a (w_d^a)^2 \delta_D(\mathbf{r}-\mathbf{r}_a)\right\rangle, \tag{A5}$$

assuming data and randoms to have the same weight distribution with $\beta = \sum_a (w_d^a)^2 / \sum_b (w_r^b)^2$. For an alternative approximation, we can write

[14] This does not affect the conclusions of the earlier work, since their estimators were formulated as differences from a set of simulations and, thus, unbiased by construction.

$$\left\langle \sum_a (w_d^a)^2 \delta_D(\mathbf{r}-\mathbf{r}_a) + \alpha^2 \sum_b (w_r^b)^2 \delta_D(\mathbf{r}-\mathbf{r}_b) \right\rangle \approx \frac{\langle w_d^2\rangle + \alpha\langle w_r^2\rangle}{\langle w_d\rangle} n(\mathbf{r}) \equiv \mathcal{S} n(\mathbf{r}), \tag{A6}$$

in terms of the average squared weights. Note that the above relations depend on the *ungridded* density field $n(\mathbf{r})$. In Ref. [74], it was assumed that this could be approximated by the gridded random field; however, this necessarily introduces an extra factor of ψ and can cause inversion errors when the number of randoms in a cell is small. Motivated by this and further testing of the power spectrum estimators, we instead use a smooth model for $n(\mathbf{r})$ computed from the survey mask and redshift distribution, as discussed in Sec. V.

To obtain a tractable form of Eq. (A4), we rewrite the 2PCF in Fourier space, leading to

$$\mathsf{S}^{ij} = \int_{\mathbf{k}} d\mathbf{r} d\mathbf{r}' n(\mathbf{r}) n(\mathbf{r}') \sum_\ell P_\ell(k) \mathcal{L}_\ell(\widehat{\mathbf{k}}\cdot\widehat{\mathbf{r}}') e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \times \psi(\mathbf{r}-\mathbf{r}_i)\psi(\mathbf{r}'-\mathbf{r}_j), \tag{A7}$$

where we have expanded the power spectrum as a Legendre series about the first galaxy's line of sight using the Legendre polynomial $\mathcal{L}_\ell(\widehat{\mathbf{k}}\cdot\widehat{\mathbf{r}}')$, adopting the Yamamoto approximation [93].[15] Given that the background density $n(\mathbf{r})$ is also a pixelized (but not ψ-convolved) field, the $\mathbf{r}$ and $\mathbf{r}'$ summations can be evaluated as FFTs. Considering the action of S^{ij} and N^{ij} on a map y_j, we find

$$\begin{aligned}
[\mathsf{S}y]_i &= \frac{1}{V_{\rm cell}} \mathrm{FFT}^{-1}\left[\tilde{\psi}\mathrm{FFT}\left[n\mathrm{FFT}^{-1}\left[\sum_\ell P_\ell \frac{4\pi}{2\ell+1}\sum_{m=-\ell}^{\ell} Y_{\ell m}\mathrm{FFT}[Y^*_{\ell m} n \mathrm{FFT}^{-1}[\tilde{\psi}\mathrm{FFT}[y]]]\right]\right]\right]_i, \\
[\mathsf{N}y]_i &= \frac{\mathcal{S}}{V_{\rm cell}} \mathrm{FFT}^{-1}[\tilde{\psi}\mathrm{FFT}[n\mathrm{FFT}^{-1}[\tilde{\psi}\mathrm{FFT}[y]]]]_i,
\end{aligned} \tag{A8}$$

where we have additionally normalized by the pixel volume $V_{\rm cell}$, written ψ in terms of its Fourier transform $\tilde{\psi}$, and expanded $\mathcal{L}_\ell(\widehat{\mathbf{k}}\cdot\widehat{\mathbf{r}}') = (4\pi)/(2\ell+1)\sum_{m=-\ell}^{\ell} Y_{\ell m}(\widehat{\mathbf{k}}) Y^*_{\ell m}(\widehat{\mathbf{r}}')$ for spherical harmonics $Y_{\ell m}$. While somewhat imposing, Eq. (A8) is simple to implement and requires just the gridded power spectrum, background number density field, and spherical harmonics. We caution that $\tilde{\psi}$ is the *forward* gridding transform, which is the reciprocal of that usually applied in conventional power spectrum estimators.

A similar form may be derived for the FKP covariance (12), including the effects of pixelation. Explicitly, we obtain

$$[\mathsf{H}_{\rm FKP}\mathbf{y}]_i = \frac{1}{V_{\rm cell}} \mathrm{FFT}^{-1}[\tilde{\psi}\mathrm{FFT}[n(\mathcal{S}+nP_{\rm FKP})\mathrm{FFT}^{-1}[\tilde{\psi}\mathrm{FFT}[y]]]]_i, \tag{A9}$$

for FKP power $P_{\rm FKP} \sim 10^4 h^{-3}$ Mpc^3. This can be straightforwardly inverted:

$$[\mathsf{H}^{-1}_{\rm FKP}\mathbf{y}]_i = \mathrm{FFT}^{-1}\left[\tilde{\psi}^{-1}\mathrm{FFT}\left[\frac{V_{\rm cell}}{n(\mathcal{S}+P_{\rm FKP}n)}\mathrm{FFT}^{-1}[\tilde{\psi}^{-1}\mathrm{FFT}[y]]\right]\right]_i. \tag{A10}$$

Additionally of interest to this work is the covariance of uniformly distributed random particles with constant number density $\bar{n}$. In continuous form, this can be shown to equal

$$\mathsf{A}^{ij} = \bar{n}\int d\mathbf{r}\psi(\mathbf{r}-\mathbf{r}_i)\psi(\mathbf{r}-\mathbf{r}_j). \tag{A11}$$

The action of A^{-1} on a field $\mathbf{y}$ is straightforwardly expressed in discrete form as

$$[\mathsf{A}^{-1}y]_i = \frac{V_{\rm cell}}{\bar{n}}\mathrm{FFT}^{-1}[\tilde{\psi}^{-2}\mathrm{FFT}[y]]_i. \tag{A12}$$

This can be straightforwardly tested by generating uniform maps $\mathbf{a}$ and verifying that $\langle \mathbf{a}^T\mathsf{A}^{-1}\mathbf{a}\rangle = N_{\rm pix}$.

[15]More nuanced approaches are also available, such as pairwise, or multiple, lines of sight [109,110].

2. Three-point statistics

We now consider the pixel-space three-point correlator $\mathsf{B}^{ijk} \equiv \langle d^i d^j d^k \rangle$. Applying the same reasoning as before, this takes the full form [cf. Eq. (36)]

$$\begin{aligned}\mathsf{B}^{ijk} = &\int d\mathbf{r}_1 d\mathbf{r}_2 d\mathbf{r}_3 n(\mathbf{r}_1) n(\mathbf{r}_2) n(\mathbf{r}_3) \int_{\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3} B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3;\widehat{\mathbf{r}}_1,\widehat{\mathbf{r}}_2,\widehat{\mathbf{r}}_3) e^{i\mathbf{k}_1\cdot\mathbf{r}_1 + i\mathbf{k}_2\cdot\mathbf{r}_2 + i\mathbf{k}_3\cdot\mathbf{r}_3} \\ &\times (2\pi)^3 \delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3) \psi(\mathbf{r}_1-\mathbf{r}_i)\psi(\mathbf{r}_2-\mathbf{r}_j)\psi(\mathbf{r}_3-\mathbf{r}_k) \\ &+ \left[\int d\mathbf{r} d\mathbf{r}' n'(\mathbf{r}) n(\mathbf{r}') \int_{\mathbf{k}} P(\mathbf{k};\widehat{\mathbf{r}},\widehat{\mathbf{r}}') e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \psi(\mathbf{r}-\mathbf{r}_i)\psi(\mathbf{r}'-\mathbf{r}_j)\psi(\mathbf{r}-\mathbf{r}_k) + 2\,\text{perms}\right] \\ &+ [1+\alpha^3/\gamma] \int d\mathbf{r} n''(\mathbf{r}) \psi(\mathbf{r}-\mathbf{r}_i)\psi(\mathbf{r}-\mathbf{r}_j)\psi(\mathbf{r}-\mathbf{r}_k),\end{aligned} \tag{A13}$$

where we allow for dependence of the power spectrum and bispectrum on the various lines of sight, with, for example, $P(\mathbf{k};\widehat{\mathbf{r}},\widehat{\mathbf{r}}') = \sum_\ell P_\ell(k)\mathcal{L}_\ell(\widehat{\mathbf{k}}\cdot\widehat{\mathbf{r}}')$ in the Yamamoto formalism. Analogous to the two-point noise covariance N, we have introduced the doubly weighted field $n''(\mathbf{r}) = \langle \sum_a (w_d^a)^3 \delta_D(\mathbf{r}-\mathbf{r}_a)\rangle$, with $\gamma = \sum_a (w_d^a)^3 / \sum_b (w_r^b)^3$, to account for the reweighted shot-noise contributions. These do not affect the statistics of this work.

From Eq. (A13), we can obtain the parameter derivatives:

$$\begin{aligned}\mathsf{B}^{ijk}_{,\alpha} = &\int d\mathbf{r}_1 d\mathbf{r}_2 d\mathbf{r}_3 n(\mathbf{r}_1) n(\mathbf{r}_2) n(\mathbf{r}_3) \int_{\mathbf{k}_1\mathbf{k}_2\mathbf{k}_3} \omega_\alpha(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3;\widehat{\mathbf{r}}_1,\widehat{\mathbf{r}}_2,\widehat{\mathbf{r}}_3) e^{i\mathbf{k}_1\cdot\mathbf{r}_1 + i\mathbf{k}_2\cdot\mathbf{r}_2 + i\mathbf{k}_3\cdot\mathbf{r}_3} \\ &\times (2\pi)^3 \delta_{\rm D}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3) \psi(\mathbf{r}_1-\mathbf{r}_i)\psi(\mathbf{r}_2-\mathbf{r}_j)\psi(\mathbf{r}_3-\mathbf{r}_k),\end{aligned} \tag{A14}$$

where ω_α arises from the basis decomposition of Eq. (37). Specializing to rotationally averaged binned bispectrum estimates, as in Eq. (48), and performing the integral over $\mathbf{r}$, we find

$$\begin{aligned}\mathsf{B}^{ijk}_{,\alpha} = &\frac{1}{\Delta_\alpha} \int d\mathbf{r} \left[\int_{\mathbf{k}_1} d\mathbf{r}_1 e^{i\mathbf{k}_1\cdot(\mathbf{r}_1-\mathbf{r})} \Theta^a(k_1) n(\mathbf{r}_1)\psi(\mathbf{r}_1-\mathbf{r}_i)\right]\left[\int_{\mathbf{k}_2} d\mathbf{r}_2 e^{i\mathbf{k}_2\cdot(\mathbf{r}_2-\mathbf{r})}\Theta^b(k_2) n(\mathbf{r}_2)\psi(\mathbf{r}_2-\mathbf{r}_j)\right] \\ &\times \left[\int_{\mathbf{k}_3} d\mathbf{r}_3 e^{i\mathbf{k}_3\cdot(\mathbf{r}_3-\mathbf{r})}\Theta^c(k_3) n(\mathbf{r}_3)\psi(\mathbf{r}_3-\mathbf{r}_k)\right] + 5\,\text{perms}.\end{aligned} \tag{A15}$$

Following a little algebra, we obtain the binned $\hat{q}_\alpha$ estimate of Eq. (51), but with the redefinition

$$\begin{aligned} g^a[\mathbf{y}](\mathbf{r}) &= \int_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{r}}\Theta^a(k) \int d\mathbf{r}' e^{i\mathbf{k}\cdot\mathbf{r}'} n(\mathbf{r}') \int d\mathbf{x}\psi(\mathbf{r}'-\mathbf{x})[\mathsf{H}^{-1}\mathbf{y}](\mathbf{x}), \\ g^a[\mathbf{y}]_i &= \text{FFT}^{-1}[\Theta^a \text{FFT}[n\text{FFT}^{-1}[\tilde{\psi}\text{FFT}[\mathsf{H}^{-1}\mathbf{y}]]]]_i \end{aligned} \tag{A16}$$

[cf. Eq. (52)], where we give the (properly normalized) gridded form involving discrete Fourier transforms in the second line. $\tilde{g}^a[\mathbf{y}]_i$ follows similarly. Similarly, the Fisher matrix retains the form of Eq. (53) but with the redefined ϕ coefficients (and likewise the $\tilde{\phi}$ coefficients)

$$\begin{aligned}\phi_\alpha[\mathbf{y}](\mathbf{x}) &= \frac{2}{\Delta_\alpha}\int d\mathbf{r} g^b[\mathbf{y}](\mathbf{r}) g^c[\mathbf{y}](\mathbf{r}) \int_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{r}}\Theta^a(k) \int d\mathbf{r}' n(\mathbf{r}')\psi(\mathbf{r}'-\mathbf{x}) + 2\,\text{perms}, \\ \phi^i_\alpha[\mathbf{y}] &= \frac{2}{V_{\rm cell}\Delta_\alpha}\text{FFT}^{-1}[\tilde{\psi}\text{FFT}[n\text{FFT}^{-1}[\Theta^a\text{FFT}[g^b[\mathbf{y}]g^c[\mathbf{y}]]]]]_i + 2\,\text{perms}\end{aligned} \tag{A17}$$

[cf. Eq. (54)], which can again be implemented using FFTs. We note an additional overall normalization factor of $V_{\rm cell}$.

One further comment is of note. In Sec. VI, we consider the estimators both with and without the pixelation corrections given above. Even when these effects are not included, it is important to add a factor $\tilde{\psi}^{-2}(\mathbf{k})$ to $\hat{q}_\alpha$ for the power spectrum estimators or $\tilde{\psi}^{-1}(\mathbf{k}_1)\tilde{\psi}^{-1}(\mathbf{k}_2)\tilde{\psi}^{-1}(\mathbf{k}_3)$ for the bispectrum estimators. This is necessary to remove the leading-order pixelation effects in the statistic and is included in all standard estimators. The treatment above forward models this effect

rather than simply dividing it out in the final step and, thus, should somewhat reduce the dependence on the grid size and slightly decrease the estimator error.

APPENDIX B: LIMITING FORMS OF THE ESTIMATORS

It is instructive to consider the limiting forms of the ML bispectrum estimator on large and small scales. Assuming a uniform number density $n(\mathbf{r}) \approx \bar{n}$ and ignoring pixelation effects and RSD, the covariance matrix given in Eq. (10) becomes

$$\mathsf{C}(\mathbf{r},\mathbf{r}') \to \bar{n}^2 \int_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} P(\mathbf{k}) + \bar{n}\delta_{\rm D}(\mathbf{r}-\mathbf{r}') \approx \begin{cases} \bar{n}^2 \int_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} P(\mathbf{k}) & \bar{n}P \gg 1, \\ \bar{n}\delta_{\rm D}(\mathbf{r}-\mathbf{r}') & \bar{n}P \ll 1 \end{cases} \tag{B1}$$

for some characteristic power spectrum amplitude P. The two approximations are appropriate on large [$P(\mathbf{k})$-dominated] and small (shot-noise-dominated) scales, respectively. Both forms are invertible:

$$\mathsf{C}^{-1}(\mathbf{r},\mathbf{r}') \to \begin{cases} \frac{1}{\bar{n}^2} \int_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \frac{1}{P(\mathbf{k})} & \bar{n}P \gg 1, \\ \frac{1}{\bar{n}} \delta_{\rm D}(\mathbf{r}-\mathbf{r}') & \bar{n}P \ll 1, \end{cases} \tag{B2}$$

such that the full form of the estimators may be computed analytically.

1. Large-scale limit

Following some computation, the large-scale limiting forms of g^a, $\tilde{g}^a$, ϕ_α, and $\tilde{\phi}_\alpha$ are found to be

$$\begin{aligned} g^a[\mathbf{y}](\mathbf{r}) &\approx \frac{1}{\bar{n}} \int_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{r}} \Theta^a(k) \frac{y(\mathbf{k})}{P(\mathbf{k})}, \qquad \tilde{g}^a[\mathbf{y}](\mathbf{r}) \approx \frac{\bar{n}}{\bar{n}_A^2} \int_k e^{-i\mathbf{k}\cdot\mathbf{r}} \Theta^a(k) y(\mathbf{k}), \\ \phi_\alpha[\mathbf{a}](\mathbf{r}) &\approx \frac{2}{\bar{n}\Delta_\alpha} \int_{\mathbf{k}\mathbf{k}'\mathbf{k}''} (2\pi)^3 \delta_{\rm D}(\mathbf{k}+\mathbf{k}'+\mathbf{k}'') e^{i\mathbf{k}\cdot\mathbf{r}} \Theta^a(k)\Theta^b(k')\Theta^c(k'') \frac{a(\mathbf{k}')a(\mathbf{k}'')}{P(\mathbf{k}')P(\mathbf{k}'')} + 2\,\text{perms}, \\ \tilde{\phi}_\alpha[\mathbf{a}](\mathbf{r}) &= \frac{2\bar{n}^3}{\bar{n}_A^4\Delta_\alpha} \int_{\mathbf{k}\mathbf{k}'\mathbf{k}''} (2\pi)^3 \delta_{\rm D}(\mathbf{k}+\mathbf{k}'+\mathbf{k}'') e^{i\mathbf{k}\cdot\mathbf{r}} \Theta^a(k)\Theta^b(k')\Theta^c(k'') a(\mathbf{k}')a(\mathbf{k}'') + 2\,\text{perms} \end{aligned} \tag{B3}$$

[from Eqs. (52) and (54)], where we have assumed that the quantity of interest is the binned bispectrum monopole as in Eq. (48) and written the number density of uniform randoms as $\bar{n}_A$. Note that all C^{-1}-weighted density fields are multiplied $1/P(\mathbf{k})$ in this limit, as in Ref. [63]. The expectations of these are also of use:

$$\begin{aligned} \langle g^b[\mathbf{a}](\mathbf{r}) \tilde{g}^c[\mathbf{a}](\mathbf{r}) \rangle &\approx \delta^{\rm K}_{bc} \int_{\mathbf{k}} \Theta^b(k) \frac{1}{P(\mathbf{k})}, \\ \langle \phi_\alpha[\mathbf{a}](\mathbf{r}) \rangle &\approx \delta^{\rm K}_{a0}\delta^{\rm K}_{bc} \times \frac{2\bar{n}_A^2}{\bar{n}\Delta_\alpha} \int_{\mathbf{k}'} \frac{1}{P^2(\mathbf{k})} \Theta^b(k) + 2\,\text{perms}, \\ \langle \tilde{\phi}_\alpha[\mathbf{a}](\mathbf{r}) \rangle &\approx \delta^{\rm K}_{a0}\delta^{\rm K}_{bc} \times \frac{2\bar{n}^3}{\bar{n}_A^2\Delta_\alpha} \int_{\mathbf{k}} \Theta^b(k) + 2\,\text{perms}, \end{aligned} \tag{B4}$$

where we assume $\langle a(\mathbf{k})a(\mathbf{k}')\rangle = \bar{n}_A^2(2\pi)^3\delta_{\rm D}(\mathbf{k}+\mathbf{k}')$ for Fourier-space maps $a(\mathbf{k})$. Notably, $\langle g^b\tilde{g}^c\rangle$ is diagonal with respect to a, b and the $\langle\phi^\alpha\rangle$ and $\langle\tilde{\phi}^\alpha\rangle$ contribute only if one k bin includes $\mathbf{k}=\mathbf{0}$ (designated by $\delta^{\rm K}_{a0}$).

These lead to the following $\hat{q}^{\rm ML}_\alpha$ and $F^{\rm ML}_{\alpha\beta}$ forms [from Eqs. (51) and (53)]:

$$\begin{aligned} \hat{q}^{\rm ML}_\alpha &\approx \frac{1}{\bar{n}^3\Delta_\alpha} \int_{\mathbf{k}\mathbf{k}'\mathbf{k}''} (2\pi)^3 \delta_{\rm D}(\mathbf{k}+\mathbf{k}'+\mathbf{k}'') \Theta^a(k)\Theta^b(k')\Theta^c(k'') \frac{d(\mathbf{k})d(\mathbf{k}')d(\mathbf{k}'')}{P(\mathbf{k})P(\mathbf{k}')P(\mathbf{k}'')}, \\ F^{\rm ML}_{\alpha\beta} &\approx \frac{1}{12} \langle \phi^i_\alpha \mathsf{C}^{-1}_{il} \phi^l_\alpha \rangle \\ &\approx \frac{\delta^{\rm K}_{aa'}(\delta^{\rm K}_{bb'}\delta^{\rm K}_{cc'} + \delta^{\rm K}_{bc'}\delta^{\rm K}_{cb'})}{3\Delta_\alpha\Delta_\beta} \int_{\mathbf{k}\mathbf{k}'\mathbf{k}''} (2\pi)^3 \delta_{\rm D}(\mathbf{k}+\mathbf{k}'+\mathbf{k}'') \Theta^a(k)\Theta^b(k')\Theta^c(k'') \frac{1}{P(\mathbf{k})P(\mathbf{k}')P(\mathbf{k}'')} + 8\,\text{perms}, \end{aligned} \tag{B5}$$

dropping any zero-lag terms. Notably, neither the second term in $\hat{q}_\alpha^{\rm ML}$ nor the Fisher matrix terms involving $\langle\phi^i_\alpha\rangle$ contribute in this limit. In the final line of Eq. (B5), we label the bins by $\alpha = \{a, b, c\}, \beta = \{a', b', c'\}$ and note that the Fisher matrix is nonzero only when the two triplets of indices contain the same members. Fixing $a \le b \le c$, the permutation symmetries allow simplification:

$$F^{\rm ML}_{\alpha\beta} \approx \delta^{\rm K}_{\alpha\beta} \times \frac{1}{\Delta_\alpha}\int_{\mathbf{k}\mathbf{k}'\mathbf{k}''}(2\pi)^3\delta_{\rm D}(\mathbf{k}+\mathbf{k}'+\mathbf{k}'')\Theta^a(k)\Theta^b(k')\Theta^c(k'')\frac{1}{P(\mathbf{k})P(\mathbf{k}')P(\mathbf{k}'')}, \tag{B6}$$

giving a diagonal Fisher matrix. This estimator matches that of Ref. [63].

2. Small-scale limit

The small-scale (shot-noise-dominated) form is simpler still. Via a similar calculation, we find

$$\begin{aligned}
g^a[\mathbf{y}](\mathbf{r}) &\approx \int_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{r}}\Theta^a(k)y(\mathbf{k}), \qquad \tilde{g}^a[\mathbf{y}](\mathbf{r}) \approx \frac{\bar{n}}{\bar{n}_A^2}\int_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{r}}\Theta^a(k)y(\mathbf{k}),\\
\phi_\alpha[\mathbf{a}](\mathbf{r}) &\approx \frac{2\bar{n}}{\Delta_\alpha}\int_{\mathbf{k}\mathbf{k}'\mathbf{k}''}(2\pi)^3\delta_{\rm D}(\mathbf{k}+\mathbf{k}'+\mathbf{k}'')e^{i\mathbf{k}\cdot\mathbf{r}}\Theta^a(k)\Theta^b(k')\Theta^c(k'')a(\mathbf{k}')a(\mathbf{k}'') + 2\,\text{perms},\\
\tilde{\phi}_\alpha[\mathbf{a}](\mathbf{r}) &\approx \frac{2\bar{n}^3}{\bar{n}_A^4\Delta_\alpha}\int_{\mathbf{k}\mathbf{k}'\mathbf{k}''}(2\pi)^3\delta_{\rm D}(\mathbf{k}+\mathbf{k}'+\mathbf{k}'')e^{i\mathbf{k}\cdot\mathbf{r}}\Theta^a(k)\Theta^b(k')\Theta^c(k'')a(\mathbf{k}')a(\mathbf{k}'') + 2\,\text{perms},
\end{aligned} \tag{B7}$$

with expectations

$$\begin{aligned}
\langle g^b[\mathbf{a}](\mathbf{r})\tilde{g}^c[\mathbf{a}](\mathbf{r})\rangle &\approx \delta^{\rm K}_{bc}\bar{n}\int_{\mathbf{k}}\Theta^b(k),\\
\langle\phi_\alpha[\mathbf{a}](\mathbf{r})\rangle &\approx \delta^{\rm K}_{bc}\delta^{\rm K}_{a0}\times\frac{2\bar{n}\bar{n}_A^2}{\Delta_\alpha}\int_{\mathbf{k}}\Theta^b(k) + 2\,\text{perms},\\
\langle\tilde{\phi}_\alpha[\mathbf{a}](\mathbf{r})\rangle &\approx \delta^{\rm K}_{a0}\delta^{\rm K}_{bc}\times\frac{2\bar{n}^3}{\bar{n}_A^2\Delta_\alpha}\int_{\mathbf{k}}\Theta^b(k) + 2\,\text{perms}.
\end{aligned} \tag{B8}$$

The $\langle\phi^\alpha\rangle$ and $\langle\tilde{\phi}^\alpha\rangle$ terms contribute only to zero-lag configurations, as before. These lead to the short-scale limits of the estimators

$$\begin{aligned}
\hat{q}_\alpha^{\rm ML} &\approx \frac{1}{\Delta_\alpha}\int_{\mathbf{k}\mathbf{k}'\mathbf{k}''}(2\pi)^3\delta_{\rm D}(\mathbf{k}+\mathbf{k}'+\mathbf{k}'')\Theta^a(k)\Theta^b(k')\Theta^c(k'')d(\mathbf{k})d(\mathbf{k}')d(\mathbf{k}''),\\
F^{\rm ML}_{\alpha\beta} &\approx \frac{\bar{n}^3\delta^{\rm K}_{aa'}(\delta^{\rm K}_{bb'}\delta^{\rm K}_{cc'}+\delta^{\rm K}_{bc'}\delta^{\rm K}_{cb'})}{3\Delta_\alpha\Delta_\beta}\int_{\mathbf{k}\mathbf{k}'\mathbf{k}''}(2\pi)^3\delta_{\rm D}(\mathbf{k}+\mathbf{k}'+\mathbf{k}'')\Theta^a(k)\Theta^b(k')\Theta^c(k'') + 8\,\text{perms},
\end{aligned} \tag{B9}$$

again ignoring zero-lag contributions. As before, incorporating the permutation symmetries leads to the simplified Fisher matrix

$$F^{\rm ML}_{\alpha\beta} \approx \delta^{\rm K}_{\alpha\beta}\times\frac{\bar{n}^3}{\Delta_\alpha}\int_{\mathbf{k}\mathbf{k}'\mathbf{k}''}(2\pi)^3\delta_{\rm D}(\mathbf{k}+\mathbf{k}'+\mathbf{k}'')\Theta^a(k)\Theta^b(k')\Theta^c(k'') \equiv \delta^{\rm K}_{\alpha\beta}\frac{\bar{n}^3}{\Delta_\alpha}V_\alpha, \tag{B10}$$

where V_α is the combined bin volume. The limit of the full binned bispectrum estimator is, thus,

$$\hat{b}_\alpha \approx \frac{1}{V_\alpha}\int_{\mathbf{k}\mathbf{k}'\mathbf{k}''}(2\pi)^3\delta_{\rm D}(\mathbf{k}+\mathbf{k}'+\mathbf{k}'')\Theta^a(k)\Theta^b(k')\Theta^c(k'')\delta(\mathbf{k})\delta(\mathbf{k}')\delta(\mathbf{k}''), \tag{B11}$$

where we have written $d(\mathbf{k}) = \bar{n}\delta(\mathbf{k})$ for fractional overdensity $\delta(\mathbf{k})$. This is the simple bispectrum estimator used in N-body simulations (see, e.g., [42]), often implemented by Fourier transforming the density field and then counting triangles.

[1] A. D. Linde, Scalar field fluctuations in the expanding universe and the new inflationary universe scenario, Phys. Lett. **116B**, 335 (1982).

[2] A. Albrecht and P. J. Steinhardt, Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking, Phys. Rev. Lett. **48**, 1220 (1982).

[3] N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi *et al.* (Planck Collaboration), Planck 2018 results. VI. Cosmological parameters, Astron. Astrophys. **641**, A6 (2020).

[4] M. Schmittfull, Y. Feng, F. Beutler, B. Sherwin, and M. Y. Chu, Eulerian BAO reconstructions and N -point statistics, Phys. Rev. D **92**, 123522 (2015).

[5] A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, L. E. Allen *et al.* (DESI Collaboration), The DESI Experiment Part I: Science, targeting, and survey design, arXiv:1611.00036.

[6] R. Laureijs, J. Amiaux, S. Arduini, J. L. Augueres, J. Brinchmann, R. Cole *et al.*, Euclid definition study report, arXiv:1110.3193.

[7] D. J. Eisenstein, H.-J. Seo, E. Sirko, and D. N. Spergel, Improving cosmological distance measurements by reconstruction of the baryon acoustic peak, Astrophys. J. **664**, 675 (2007).

[8] P. Gagrani and L. Samushia, Information content of the angular multipoles of redshift-space galaxy bispectrum, Mon. Not. R. Astron. Soc. **467**, 928 (2017).

[9] N. Agarwal, V. Desjacques, D. Jeong, and F. Schmidt, Information content in the redshift-space galaxy power spectrum and bispectrum, J. Cosmol. Astropart. Phys. 03 (2021) 021.

[10] D. Karagiannis, A. Lazanu, M. Liguori, A. Raccanelli, N. Bartolo, and L. Verde, Constraining primordial non-Gaussianity with bispectrum and power spectrum from upcoming optical and radio surveys, Mon. Not. R. Astron. Soc. **478**, 1341 (2018).

[11] A. Moradinezhad Dizgah, M. Biagetti, E. Sefusatti, V. Desjacques, and J. Noreña, Primordial non-Gaussianity from biased tracers: likelihood analysis of real-space power spectrum and bispectrum, J. Cosmol. Astropart. Phys. 05 (2021) 015.

[12] A. Chudaykin and M. M. Ivanov, Measuring neutrino masses with large-scale structure: Euclid forecast with controlled theoretical error, J. Cosmol. Astropart. Phys. 11 (2019) 034.

[13] C. Hahn and F. Villaescusa-Navarro, Constraining M_ν with the bispectrum. Part II. The information content of the galaxy bispectrum monopole, J. Cosmol. Astropart. Phys. 04 (2021) 029.

[14] S. Alam, A. Aviles, R. Bean, Y.-C. Cai, M. Cautun, J. L. Cervantes-Cota *et al.*, Testing the theory of gravity with DESI: estimators, predictions and simulation requirements, arXiv:2011.05771.

[15] A. F. Heavens, Estimating non-Gaussianity in the microwave background, Mon. Not. R. Astron. Soc. **299**, 805 (1998).

[16] L. Verde, L. Wang, A. F. Heavens, and M. Kamionkowski, Large-scale structure, the cosmic microwave background and primordial non-Gaussianity, Mon. Not. R. Astron. Soc. **313**, 141 (2000).

[17] A. Gangui and J. Martin, Best unbiased estimators for the three-point correlators of the cosmic microwave background radiation, Phys. Rev. D **62**, 103004 (2000).

[18] M. G. Santos, A. Heavens, A. Balbi, J. Borrill, P. G. Ferreira, S. Hanany *et al.*, Multiple methods for estimating the bispectrum of the cosmic microwave background with application to the MAXIMA data, Mon. Not. R. Astron. Soc. **341**, 623 (2003).

[19] D. Babich, Optimal estimation of non-Gaussianity, Phys. Rev. D **72**, 043003 (2005).

[20] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga, Limits on non-Gaussianities from WMAP data, J. Cosmol. Astropart. Phys. 05 (2006) 004.

[21] J. R. Fergusson and E. P. S. Shellard, Shape of primordial non-Gaussianity and the CMB bispectrum, Phys. Rev. D **80**, 043510 (2009).

[22] P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown *et al.* (Planck Collaboration), Planck 2013 results. XXIV. Constraints on primordial non-Gaussianity, Astron. Astrophys. **571**, A24 (2014).

[23] J. N. Fry and M. Seldner, Transform analysis of the high-resolution Shane-Wirtanen Catalog—The power spectrum and the bispectrum, Astrophys. J. **259**, 474 (1982).

[24] R. Scoccimarro, H. A. Feldman, J. N. Fry, and J. A. Frieman, The bispectrum of IRAS Redshift catalogs, Astrophys. J. **546**, 652 (2001).

[25] E. Sefusatti and R. Scoccimarro, Galaxy bias and halo-occupation numbers from large-scale clustering, Phys. Rev. D **71**, 063001 (2005).

[26] H. Gil-Marín, J. Noreña, L. Verde, W. J. Percival, C. Wagner, M. Manera, and D. P. Schneider, The power spectrum and bispectrum of SDSS DR11 BOSS galaxies—I. Bias and gravity, Mon. Not. R. Astron. Soc. **451**, 539 (2015).

[27] H. Gil-Marín, W. J. Percival, L. Verde, J. R. Brownstein, C.-H. Chuang, F.-S. Kitaura, S. A. Rodríguez-Torres, and M. D. Olmstead, The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: RSD measurement from the power spectrum and bispectrum of the DR12 BOSS galaxies, Mon. Not. R. Astron. Soc. **465**, 1757 (2017).

[28] Z. Slepian, D. J. Eisenstein, F. Beutler, C.-H. Chuang, A. J. Cuesta, J. Ge *et al.*, The large-scale three-point correlation function of the SDSS BOSS DR12 CMASS galaxies, Mon. Not. R. Astron. Soc. **468**, 1070 (2017).

[29] D. W. Pearson and L. Samushia, A detection of the baryon acoustic oscillation features in the SDSS BOSS DR12 galaxy bispectrum, Mon. Not. R. Astron. Soc. **478**, 4500 (2018).

[30] G. d'Amico, J. Gleyzes, N. Kokron, K. Markovic, L. Senatore, P. Zhang, F. Beutler, and H. Gil-Marín, The cosmological analysis of the SDSS/BOSS data from the effective field theory of large-scale structure, J. Cosmol. Astropart. Phys. 05 (2020) 005.

[31] P. J. E. Peebles and E. J. Groth, Statistical analysis of catalogs of extragalactic objects. V. Three-point correlation function for the galaxy distribution in the Zwicky catalog, Astrophys. J. **196**, 1 (1975).

[32] P. J. E. Peebles, The galaxy and mass N-point correlation functions: A blast from the past, Astron. Soc. Pac. Conf. Ser. **252**, 201 (2001), arXiv:astro-ph/0103040.
[33] Y. P. Jing and G. Börner, The three-point correlation function of galaxies determined from the Las Campanas redshift survey, Astrophys. J. **503**, 37 (1998).
[34] I. Kayo, Y. Suto, R. C. Nichol, J. Pan, I. Szapudi, A. J. Connolly *et al.*, Three-point correlation functions of SDSS galaxies in redshift space: Morphology, color, and luminosity dependence, Publ. Astron. Soc. Jpn. **56**, 415 (2004).
[35] R. C. Nichol, R. K. Sheth, Y. Suto, A. J. Gray, I. Kayo, R. H. Wechsler *et al.*, The effect of large-scale structure on the SDSS galaxy three-point correlation function, Mon. Not. R. Astron. Soc. **368**, 1507 (2006).
[36] F. Marín, The large-scale three-point correlation function of Sloan digital sky survey luminous red galaxies, Astrophys. J. **737**, 97 (2011).
[37] H. Guo, Z. Zheng, Y. P. Jing, I. Zehavi, C. Li, D. H. Weinberg *et al.*, Modelling the redshift-space three-point correlation function in SDSS-III., Mon. Not. R. Astron. Soc. **449**, L95 (2015).
[38] Z. Slepian, D. J. Eisenstein, J. R. Brownstein, C.-H. Chuang, H. Gil-Marín, S. Ho *et al.*, Detection of baryon acoustic oscillation features in the large-scale three-point correlation function of SDSS BOSS DR12 CMASS galaxies, Mon. Not. R. Astron. Soc. **469**, 1738 (2017).
[39] Z. Slepian, D. J. Eisenstein, J. A. Blazek, J. R. Brownstein, C.-H. Chuang, H. Gil-Marín *et al.*, Constraining the baryon-dark matter relative velocity with the large-scale three-point correlation function of the SDSS BOSS DR12 CMASS galaxies, Mon. Not. R. Astron. Soc. **474**, 2109 (2018).
[40] O. H. E. Philcox, Z. Slepian, J. Hou, C. Warner, R. N. Cahn, and D. J. Eisenstein, ENCORE: Estimating galaxy N-point correlation functions in $\mathcal{O}(N_g^2)$ time, arXiv:2105.08722.
[41] O. H. E. Philcox, J. Hou, and Z. Slepian, A First Measurement of the BOSS Non-Gaussian Four-Point Function, arXiv:2108.01670 [Phys. Rev. D (to be published)].
[42] C. A. Watkinson, S. Majumdar, J. R. Pritchard, and R. Mondal, A fast estimator for the bispectrum and beyond—A practical method for measuring non-Gaussianity in 21-cm maps, Mon. Not. R. Astron. Soc. **472**, 2436 (2017).
[43] R. Scoccimarro, Fast estimators for redshift-space clustering, Phys. Rev. D **92**, 083532 (2015).
[44] D. Regan, An inventory of bispectrum estimators for redshift space distortions, J. Cosmol. Astropart. Phys. 12 (2017) 020.
[45] O. H. E. Philcox and D. J. Eisenstein, Computing the small-scale galaxy power spectrum and bispectrum in configuration space, Mon. Not. R. Astron. Soc. **492**, 1214 (2020).
[46] O. H. E. Philcox, A faster Fourier transform? Computing small-scale power spectra and bispectra for cosmological simulations in $\mathcal{O}(N^2)$ time, Mon. Not. R. Astron. Soc. **501**, 4004 (2021).
[47] R. Scoccimarro, S. Colombi, J. N. Fry, J. A. Frieman, E. Hivon, and A. Melott, Nonlinear evolution of the bispectrum of cosmological perturbations, Astrophys. J. **496**, 586 (1998).
[48] S. Matarrese, L. Verde, and A. F. Heavens, Large-scale bias in the universe: Bispectrum method, Mon. Not. R. Astron. Soc. **290**, 651 (1997).
[49] L. Verde, A. F. Heavens, S. Matarrese, and L. Moscardini, Large-scale bias in the universe—II. Redshift-space bispectrum, Mon. Not. R. Astron. Soc. **300**, 747 (1998).
[50] R. Scoccimarro, H. M. P. Couchman, and J. A. Frieman, The bispectrum as a signature of gravitational instability in redshift space, Astrophys. J. **517**, 531 (1999).
[51] R. Scoccimarro and H. M. P. Couchman, A fitting formula for the non-linear evolution of the bispectrum, Mon. Not. R. Astron. Soc. **325**, 1312 (2001).
[52] R. Scoccimarro, The bispectrum: From theory to observations, Astrophys. J. **544**, 597 (2000).
[53] T. Baldauf, L. Mercolli, M. Mirbabayi, and E. Pajer, The bispectrum in the effective field theory of large scale structure, J. Cosmol. Astropart. Phys. 05 (2015) 007.
[54] R. E. Angulo, S. Foreman, M. Schmittfull, and L. Senatore, The one-loop matter bispectrum in the effective field theory of large scale structures, J. Cosmol. Astropart. Phys. 10 (2015) 039.
[55] T. Nishimichi, G. D'Amico, M. M. Ivanov, L. Senatore, M. Simonović, M. Takada, M. Zaldarriaga, and P. Zhang, Blinded challenge for precision cosmology with large-scale structure: Results from effective field theory for the redshift-space galaxy power spectrum, Phys. Rev. D **102**, 123541 (2020).
[56] A. Cooray, Squared temperature-temperature power spectrum as a probe of the CMB bispectrum, Phys. Rev. D **64**, 043516 (2001).
[57] D. Munshi, A. L. Melott, and P. Coles, Generalised cumulant correlators and hierarchical clustering, arXiv:astro-ph/9812271.
[58] M. Schmittfull, T. Baldauf, and U. Seljak, Near optimal bispectrum estimators for large-scale structure, Phys. Rev. D **91**, 043530 (2015).
[59] D. Obreschkow, C. Power, M. Bruderer, and C. Bonvin, A robust measure of cosmic structure beyond the power spectrum: Cosmic filaments and the temperature of dark matter, Astrophys. J. **762**, 115 (2013).
[60] A. Eggemeier, T. Battefeld, R. E. Smith, and J. Niemeyer, The anisotropic line correlation function as a probe of anisotropies in galaxy surveys, Mon. Not. R. Astron. Soc. **453**, 797 (2015).
[61] C.-T. Chiang, C. Wagner, F. Schmidt, and E. Komatsu, Position-dependent power spectrum of the large-scale structure: a novel method to measure the squeezed-limit bispectrum, J. Cosmol. Astropart. Phys. 05 (2014) 048.
[62] C.-T. Chiang, C. Wagner, A. G. Sánchez, F. Schmidt, and E. Komatsu, Position-dependent correlation function from the SDSS-III baryon oscillation spectroscopic survey data release 10 CMASS sample, J. Cosmol. Astropart. Phys. 09 (2015) 028.
[63] J. R. Fergusson, D. M. Regan, and E. P. S. Shellard, Rapid separable analysis of higher order correlators in large-scale structure, Phys. Rev. D **86**, 063511 (2012).
[64] D. M. Regan, M. M. Schmittfull, E. P. S. Shellard, and J. R. Fergusson, Universal non-Gaussian initial conditions for N-body simulations, Phys. Rev. D **86**, 123524 (2012).

[65] M. M. Schmittfull, D. M. Regan, and E. P. S. Shellard, Fast estimation of gravitational and primordial bispectra in large scale structures, Phys. Rev. D **88**, 063512 (2013).
[66] J. Byun, A. Oddo, C. Porciani, and E. Sefusatti, Towards cosmological constraints from the compressed modal bispectrum: a robust comparison of real-space bispectrum estimators, J. Cosmol. Astropart. Phys. 03 (2021) 105.
[67] A. F. Heavens, R. Jimenez, and O. Lahav, Massive lossless data compression and multiple parameter estimation from galaxy spectra, Mon. Not. R. Astron. Soc. **317**, 965 (2000).
[68] D. Gualdi, M. Manera, B. Joachimi, and O. Lahav, Maximal compression of the redshift-space galaxy power spectrum and bispectrum, Mon. Not. R. Astron. Soc. **476**, 4045 (2018).
[69] D. Gualdi, H. Gil-Marín, R. L. Schuhmann, M. Manera, B. Joachimi, and O. Lahav, Enhancing BOSS bispectrum cosmological constraints with maximal compression, Mon. Not. R. Astron. Soc. **484**, 3713 (2019).
[70] D. Gualdi, H. Gil-Marín, M. Manera, B. Joachimi, and O. Lahav, Geometrical compression: a new method to enhance the BOSS galaxy bispectrum monopole constraints, Mon. Not. R. Astron. Soc. **484**, L29 (2019).
[71] O. H. E. Philcox, M. M. Ivanov, M. Zaldarriaga, M. Simonović, and M. Schmittfull, Fewer mocks and less noise: Reducing the dimensionality of cosmological observables with subspace projections, Phys. Rev. D **103**, 043508 (2021).
[72] J. Alsing and B. Wandelt, Generalized massive optimal data compression, Mon. Not. R. Astron. Soc. **476**, L60 (2018).
[73] N. S. Sugiyama, S. Saito, F. Beutler, and H.-J. Seo, A complete FFT-based decomposition formalism for the redshift-space bispectrum, Mon. Not. R. Astron. Soc. **484**, 364 (2019).
[74] O. H. E. Philcox, Cosmology without window functions: Quadratic estimators for the galaxy power spectrum, Phys. Rev. D **103**, 103504 (2021).
[75] F. Beutler and P. McDonald, Unified galaxy power spectrum measurements from 6dFGS, BOSS, and eBOSS, arXiv:2106.06324.
[76] M. Tegmark, How to measure CMB power spectra without losing information, Phys. Rev. D **55**, 5895 (1997).
[77] M. Tegmark, A. J. S. Hamilton, M. A. Strauss, M. S. Vogeley, and A. S. Szalay, Measuring the galaxy power spectrum with future redshift surveys, Astrophys. J. **499**, 555 (1998).
[78] M. Tegmark, A. N. Taylor, and A. F. Heavens, Karhunen-Loève Eigenvalue problems in cosmology: How should we tackle large data sets?, Astrophys. J. **480**, 22 (1997).
[79] J. R. Bond, A. H. Jaffe, and L. Knox, Estimating the power spectrum of the cosmic microwave background, Phys. Rev. D **57**, 2117 (1998).
[80] S. P. Oh, D. N. Spergel, and G. Hinshaw, An efficient technique to determine the power spectrum from cosmic microwave background sky maps, Astrophys. J. **510**, 551 (1999).
[81] M. Tegmark, A. J. S. Hamilton, and Y. Xu, The power spectrum of galaxies in the 2dF 100k redshift survey, Mon. Not. R. Astron. Soc. **335**, 887 (2002).
[82] M. Tegmark, S. Dodelson, D. J. Eisenstein, V. Narayanan, R. Scoccimarro, R. Scranton *et al.*, The angular power spectrum of galaxies from early sloan digital sky survey data, Astrophys. J. **571**, 191 (2002).
[83] A. J. S. Hamilton, Power spectrum estimation I. Basics, Lect. Notes Phys. **665**, 415 (2008).
[84] A. J. S. Hamilton, Power spectrum estimation II. Linear maximum likelihood, Lect. Notes Phys. **665**, 433 (2008).
[85] M. Tegmark, M. R. Blanton, M. A. Strauss, F. Hoyle, D. Schlegel, R. Scoccimarro *et al.*, The three-dimensional power spectrum of galaxies from the sloan digital sky survey, Astrophys. J. **606**, 702 (2004).
[86] K. M. Smith and M. Zaldarriaga, Algorithms for bispectra: Forecasting, optimal analysis and simulation, Mon. Not. R. Astron. Soc. **417**, 2 (2011).
[87] M. M. Ivanov, M. Simonović, and M. Zaldarriaga, Cosmological parameters from the BOSS galaxy power spectrum, J. Cosmol. Astropart. Phys. 05 (2020) 042.
[88] N. S. Sugiyama, S. Saito, F. Beutler, and H.-J. Seo, Perturbation theory approach to predict the covariance matrices of the galaxy power spectrum and bispectrum in redshift space, Mon. Not. R. Astron. Soc. **497**, 1684 (2020).
[89] D. Gualdi and L. Verde, Galaxy redshift-space bispectrum: the importance of being anisotropic, J. Cosmol. Astropart. Phys. 06 (2020) 041.
[90] D. Gualdi, H. Gil-Marin, and L. Verde, Joint analysis of anisotropic power spectrum, bispectrum and trispectrum: Application to N-body simulations, J. Cosmol. Astropart. Phys. 07 (2021) 008.
[91] http://github.com/oliverphilcox/BOSS-Without-Windows.
[92] H. A. Feldman, N. Kaiser, and J. A. Peacock, Power-spectrum analysis of three-dimensional redshift surveys, Astrophys. J. **426**, 23 (1994).
[93] K. Yamamoto, M. Nakamichi, A. Kamino, B. A. Bassett, and H. Nishioka, A measurement of the quadrupole power spectrum in the clustering of the 2dF QSO survey, Publ. Astron. Soc. Jpn. **58**, 93 (2006).
[94] N. Hand, Y. Li, Z. Slepian, and U. Seljak, An optimal FFT-based anisotropic power spectrum estimator, J. Cosmol. Astropart. Phys. 07 (2017) 002.
[95] E. Castorina, N. Hand, U. Seljak, F. Beutler, C.-H. Chuang, C. Zhao *et al.*, Redshift-weighted constraints on primordial non-Gaussianity from the clustering of the eBOSS DR14 quasars in Fourier space, J. Cosmol. Astropart. Phys. 09 (2019) 010.
[96] E. Sellentin, A. H. Jaffe, and A. F. Heavens, On the use of the Edgeworth expansion in cosmology I: How to foresee and evade its pitfalls, arXiv:1709.03452.
[97] A. Heavens, Statistical techniques in cosmology, arXiv:0906.0664.
[98] S. Alam, M. Ata, S. Bailey, F. Beutler, D. Bizyaev, J. A. Blazek *et al.*, The clustering of galaxies in the completed SDSS-III baryon oscillation spectroscopic survey: Cosmological analysis of the DR12 galaxy sample, Mon. Not. R. Astron. Soc. **470**, 2617 (2017).
[99] S. A. Rodríguez-Torres, C.-H. Chuang, F. Prada, H. Guo, A. Klypin, P. Behroozi *et al.*, The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: Modelling the clustering and halo occupation distribution

of BOSS CMASS galaxies in the final data release, Mon. Not. R. Astron. Soc. **460**, 1173 (2016).
[100] F.-S. Kitaura, S. Rodríguez-Torres, C.-H. Chuang, C. Zhao, F. Prada, H. Gil-Marín *et al.*, The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: Mock galaxy catalogues for the BOSS final data release, Mon. Not. R. Astron. Soc. **456**, 4156 (2016).
[101] http://data.sdss.org/sas/dr12/boss/lss.
[102] F. Beutler, H.-J. Seo, S. Saito, C.-H. Chuang, A. J. Cuesta, D. J. Eisenstein *et al.*, The clustering of galaxies in the completed SDSS-III baryon oscillation spectroscopic survey: Anisotropic galaxy clustering in Fourier space, Mon. Not. R. Astron. Soc. **466**, 2242 (2017).
[103] N. Hand, Y. Feng, F. Beutler, Y. Li, C. Modi, U. Seljak, and Z. Slepian, nbodykit: An open-source, massively parallel toolkit for large-scale structure, Astron. J. **156**, 160 (2018).
[104] A. Chudaykin, M. M. Ivanov, O. H. E. Philcox, and M. Simonović, Nonlinear perturbation theory extension of the Boltzmann code CLASS, Phys. Rev. D **102**, 063533 (2020).
[105] M. Schmittfull and A. M. Dizgah, Galaxy skew-spectra in redshift-space, J. Cosmol. Astropart. Phys. 03 (2021) 020.
[106] A. Moradinezhad Dizgah, H. Lee, M. Schmittfull, and C. Dvorkin, Capturing non-Gaussianity of the large-scale structure with weighted skew-spectra, J. Cosmol. Astropart. Phys. 04 (2020) 011.
[107] O. H. E. Philcox, B. D. Sherwin, G. S. Farren, and E. J. Baxter, Determining the Hubble constant without the sound horizon: Measurements from galaxy surveys, Phys. Rev. D **103**, 023538 (2021).
[108] Y. P. Jing, Correcting for the alias effect when measuring the power spectrum using a Fast Fourier Transform, Astrophys. J. **620**, 559 (2005).
[109] E. Castorina and M. White, Beyond the plane-parallel approximation for redshift surveys, Mon. Not. R. Astron. Soc. **476**, 4403 (2018).
[110] O. H. E. Philcox and Z. Slepian, Beyond the Yamamoto approximation: Anisotropic power spectra and correlation functions with pairwise lines of sight, Phys. Rev. D **103**, 123509 (2021).