Parity-Odd Power Spectra: Concise Statistics for Cosmological Parity Violation

Drew Jamieson, Angelo Caravano, Jiamin Hou, Zachary Slepian, Eiichiro Komatsu

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **533,** 2582–2598 (2024) https://doi.org/10.1093/mnras/stae1924
Advance Access publication 2024 August 8

Parity-odd power spectra: concise statistics for cosmological parity violation

Drew Jamieson [1★] Angelo Caravano,[2★] Jiamin Hou,[3,4★] Zachary Slepian [3] and Eiichiro Komatsu[1,5,6]

[1]*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Straße 1, D-85748 Garching, Germany*
[2]*Institut d'Astrophysique de Paris, UMR 7095 du CNRS et de Sorbonne Université, 98 bis Bd Arago, F-75014 Paris, France*
[3]*Department of Astronomy, University of Florida, Gainesville, FL 32611, USA*
[4]*Max-Planck-Institut für extraterrestrische Physik, Gießenbachstraße 1, D-85748 Garching, Germany*
[5]*Ludwig-Maximilians-Universität München, Schellingstr. 4, D-80799 München, Germany*
[6]*Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU, WPI), University of Tokyo, Chiba 277-8582, Japan*

Accepted 2024 August 5. Received 2024 August 1; in original form 2024 July 15

ABSTRACT

We introduce the parity-odd power (POP) spectra, a novel set of observables for probing parity violation in cosmological N-point statistics. POP spectra are derived from composite fields obtained by applying non-linear transformations, involving also gradients, curls, and filtering functions, to a scalar field. This compresses the parity-odd trispectrum into a power spectrum. These new statistics offer several advantages: they are computationally fast to construct, estimating their covariance is less demanding compared to estimating that of the full parity-odd trispectrum, and they are simple to model theoretically. We measure the POP spectra on simulations of a scalar field with a specific parity-odd trispectrum shape. We compare these measurements to semi-analytic theoretical calculations and find agreement. We also explore extensions and generalizations of these parity-odd observables.

Key words: methods: analytical – methods: data analysis – methods: statistical – inflation – large-scale structure of Universe.

1 INTRODUCTION

Modern physics has advanced considerably by proposing theoretical symmetries and testing whether or not those symmetries are respected or violated in nature. A crucial breakthrough for the standard model of particle physics occurred with the surprising discovery of weak nuclear parity violation in 1957 (Lee & Yang 1956; Garwin, Lederman & Weinrich 1957; Wu et al. 1957). Parity is the negation of spatial coordinates: $\boldsymbol{x} \rightarrow -\boldsymbol{x}$, and its violation implies that the laws of physics distinguish between right-handed and left-handed chiralities. While parity is known to be maximally violated on small scales by weak nuclear interactions, the status of parity on cosmological scales remains an open question.

Recent investigations using the method proposed in Cahn, Slepian & Hou (2023) and Cahn & Slepian (2023) have uncovered intriguing evidence of cosmological parity violation in the large-scale structure (LSS) of the Universe. Notably, analyses of the BOSS galaxy four-point correlation function detected parity violation with significance of up to 7σ (Philcox 2022; Hou, Slepian & Cahn 2023b). Confirmation of parity violation in the galaxy clustering would have profound implications. If the signal is primordial in origin, it is a form of primordial non-Gaussianity and would greatly inform our early-Universe inflationary models (Bartolo et al. 2004). Current cosmic microwave background (CMB) observations are consistent with parity conservation (Philcox 2023; Philcox & Shiraishi 2024). However, since the CMB is measured on a 2D surface and is sensitive to different scales than the LSS, it is still unclear what this implies for the BOSS signal. The central challenge in the current BOSS analysis lies in robustly estimating the covariance (Cahn et al. 2023), which can be influenced by both instrumental systematics and observational effects (Hou et al. 2023b) as well as the type of mocks used to calibrate or compute the covariance (Hou et al. 2022; Philcox & Ereza 2024).

Next-generation 3D spectroscopic surveys such as DESI (Aghamousa et al. 2016; Adame et al. 2024), Euclid (Amendola et al. 2018), Subaru PFS (Takada et al. 2014), Roman (Wang et al. 2022), and SPHEREx (Doré et al. 2014), as well as photometric efforts such as *LSST* (Ivezić et al. 2019), will provide significantly more data than is currently available. To constrain large-scale parity violation robustly with this wealth of data, we need analysis tools that facilitate accurate covariance estimation and characterization of observational systematics.

For scalar fields like the primordial curvature perturbation or the matter density contrast, N-point statistics are sensitive to parity at orders four and above (Shiraishi 2016). Measurements of these correlators have an enormous number of degrees of freedom [e.g. the BOSS analysis of Hou et al. (2023b) had 18 000]. Covariance estimation for such large data vectors requires a large number of simulated mock data sets, often

★ E-mails: jamieson@mpa-garching.mpg.de (DJ); caravano@iap.fr (AC); jiamin.hou@ufl.edu (JH)

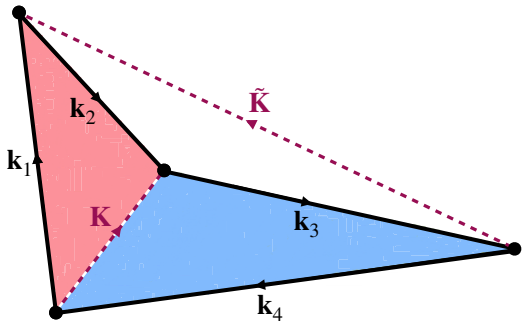

Figure 1. Here we display a Fourier-space tetrahedron representing the trispectrum configuration $T(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)$. The diagonals (in dashed magenta) are defined as $\boldsymbol{K} = \boldsymbol{k}_1 + \boldsymbol{k}_2$ and $\tilde{\boldsymbol{K}} = \boldsymbol{k}_1 + \boldsymbol{k}_4$. The tetrahedron is formed by joining two triangles, $\{\boldsymbol{k}_1, \boldsymbol{k}_2, -\boldsymbol{K}\}$ on the left and $\{\boldsymbol{k}_3, \boldsymbol{k}_4, \boldsymbol{K}\}$ on the right, along their shared edge.

combined with approximate analytic covariance templates [e.g. Hou et al. (2022); also Slepian & Eisenstein (2015) and Xu et al. (2012)]. Even with statistically accurate covariance estimation, systematics in observational data can bias both the detected signal and its significance.

A complementary approach to high-order N-point statistics is lower-order statistics, such as power spectra, that are computed on composite fields. These composite fields are created by applying non-linear transformations to the original scalar field. Composite-field statistics average or compress the high-dimensional correlators down to low-dimensional correlators. As a result, the covariances of these composite-field power spectra are less demanding to estimate, due to having fewer degrees of freedom, while still carrying higher-order information. The trade-off involved in this approach is the loss of information through compression. Examples of weighted, compressed analyses for bispectra and parity-even trispectra have been done before and are known as *skew-spectra* and *kurt-spectra* (Schmittfull, Baldauf & Seljak 2015; Moradinezhad Dizgah et al. 2020; Schmittfull & Moradinezhad Dizgah 2021; Munshi et al. 2022; Hou et al. 2023a), as well as works on the unweighted integrated three-point correlation function (Slepian et al. 2017), trispectrum (Gualdi, Gil-Marín & Verde 2021; Gualdi & Verde 2022), and four-point function (Sabiu et al. 2019); the clustering fossil estimators (Jeong & Kamionkowski 2012); and a Fourier-transform-based algorithm for the full four-point function is Sunseri et al. (2023). An estimator of the CMB lensing power spectrum can also be constructed in this manner (Hu 2001).

In the current work, we introduce new methods for detecting cosmological parity violation by constructing power-spectrum-like statistics sensitive to parity. We call our new observables *parity-odd power spectra*, or *POP spectra*. The POP spectra are compressions of the 6D four-point statistics down to 1D power spectra. They are computationally efficient to construct and their lower dimensionality alleviates the burden of full four-point covariance estimation.

Importantly, POP spectra are sensitive to soft limits of parity-violating trispectrum shapes, which can help to place strong constraints on primordial parity violation. Moreover, being formulated in Fourier space, they are easy to interpret and can facilitate theoretical understanding of the signal, which is crucial to distinguish it from potential observational systematics. It is also possible to construct real-space and spherical-harmonic-space versions of these statistics. As we will demonstrate below, constructing parity-odd statistics is non-trivial, requiring more sophisticated techniques than those required for parity-even compressed estimators.

This work is structured as follows. In Section 2, we discuss the primordial parity-odd trispectrum. In Section 3, we present two POP spectra constructions based on composite vector and scalar fields. In Section 4, we validate the estimator on simulated data, including comparisons with semi-analytical calculations. In Section 5, we explore extensions and generalizations of our parity-odd observables. Finally, we conclude and discuss prospects for measuring these parity-odd statistics on observational data in Section 6.

2 PARITY-VIOLATING TRISPECTRUM

Under statistical homogeneity and isotropy, the four-point correlation function is the lowest-order parity-sensitive statistic for a single scalar field. In Fourier space, this corresponds to the trispectrum, $T(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)$, defined through the four-point correlator,

$$\left\langle \Phi(\boldsymbol{k}_1)\Phi(\boldsymbol{k}_2)\Phi(\boldsymbol{k}_3)\Phi(\boldsymbol{k}_4) \right\rangle \equiv (2\pi)^3 \delta_{\rm D}^{(3)}(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3 + \boldsymbol{k}_4)\, T(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)\,. \tag{1}$$

The Dirac delta function imposes translational invariance by requiring that the sum of the four wave vectors vanishes. Thus, they form a closed loop in Fourier space, and define a tetrahedron, as displayed in Fig. 1. More precisely, the four wave vectors define an equivalence class of tetrahedra, consisting of the 24 tetrahedra produced by permuting them. These are not all geometrically distinct, since a cyclic permutation of wave vectors within a given tetrahedron yields a congruent tetrahedron. The three independent wave vectors, which we take to be $\boldsymbol{k}_1$, $\boldsymbol{k}_2$, and $\boldsymbol{k}_3$, uniquely determine these tetrahedra. Statistical isotropy, however, reduces the dimensionality of the trispectrum from nine to six continuous degrees of freedom. We will choose the wave vector magnitudes k_1, k_2, k_3, and k_4 as four of these continuous parameters, and refer to them as *sides* of the tetrahedron. Then we take $K = |\boldsymbol{k}_1 + \boldsymbol{k}_2| = |\boldsymbol{k}_3 + \boldsymbol{k}_4|$ and $\tilde{K} = |\boldsymbol{k}_1 + \boldsymbol{k}_4| = |\boldsymbol{k}_2 + \boldsymbol{k}_3|$ as the other two parameters and refer to these as *diagonals* to distinguish them from the wave vector magnitudes.

Among the set of 24 tetrahedra, there is an additional discrete degree of freedom that splits them into two groups of 12. These are distinguished by the sign of the vector triple product $\boldsymbol{k}_1 \cdot (\boldsymbol{k}_2 \times \boldsymbol{k}_3)$. This can be thought of as the *handedness* or helicity of the tetrahedron. The triple product is positive for a right-handed configuration and negative for a left-handed configuration. Under a parity transform, the sign of the triple product changes, so a parity transformation interchanges the two helicities.

We can isolate the trispectrum for only right-handed configurations, and obtain the right-handed trispectrum $T_{\rm R}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)$, and similarly for the left-handed trispectrum $T_{\rm L}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)$. Let $\pi_{123} \equiv \mathrm{sgn}\big(\boldsymbol{k}_1 \cdot (\boldsymbol{k}_2 \times \boldsymbol{k}_3)\big)$, then

$$T(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) = \frac{1+\pi_{123}}{2} T_{\rm R}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) + \frac{1-\pi_{123}}{2} T_{\rm L}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)\,. \tag{2}$$

The right- and left-handed parts of the trispectrum are not necessarily equal. If they are different, this indicates that the trispectrum has a parity-odd component. For the modes of a real field, a parity transformation is complex conjugation,

$$\Phi(-\boldsymbol{k}) = \Phi(\boldsymbol{k})^*\,, \tag{3}$$

so the parity-even component of the trispectrum is its real part and is proportional to the sum of $T_{\rm R} + T_{\rm L}$. The parity-odd component is its imaginary part and is proportional to the difference $T_{\rm R} - T_{\rm L}$:

$$T(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) = \frac{1}{2}\Big(T_{\rm R}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) + T_{\rm L}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)\Big) + \frac{\pi_{123}}{2}\Big(T_{\rm R}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) - T_{\rm L}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)\Big) \tag{4}$$

$$= T_+(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) + i T_-(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)\,. \tag{5}$$

The relationship between the right-/left-trispectra and the parity-even/parity-odd trispectra are,

$$T_{\rm R/L} = T_+ \pm i\pi_{123} T_-\,, \tag{6}$$

$$T_\pm = \frac{1}{2}(-i\pi_{123})^{(1\mp1)/2}\big(T_{\rm R} \pm T_{\rm L}\big)\,. \tag{7}$$

Due to isotropy, the only vectors we can use to form a parity-odd structure are $\boldsymbol{k}_1$, $\boldsymbol{k}_2$, and $\boldsymbol{k}_3$, so $T_-(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)$ must be proportional to the triple product $\boldsymbol{k}_1 \cdot (\boldsymbol{k}_2 \times \boldsymbol{k}_3)$. We parametrize the shape of the imaginary trispectrum as (Coulton, Philcox & Villaescusa-Navarro 2024)

$$T_-(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) = \boldsymbol{k}_1 \cdot (\boldsymbol{k}_2 \times \boldsymbol{k}_3)\, \tau_-(k_1, k_2, k_3, k_4, K, \tilde{K})\,. \tag{8}$$

The left-hand side of equation (1) is totally symmetric under the interchange of any of the wave vectors, and the triple product in equation (8) is totally antisymmetric. These symmetries require that $\tau_-(k_1, k_2, k_3, k_4, K, \tilde{K})$ is totally antisymmetric under the interchange of any two of the four wave vectors. The trispectrum cannot depend on one diagonal and not the other since $K \to \tilde{K}$ under the interchange of $\boldsymbol{k}_2$ and $\boldsymbol{k}_4$. Also, under the interchange of $\boldsymbol{k}_2$ and $\boldsymbol{k}_3$, K becomes $|\boldsymbol{k}_1 + \boldsymbol{k}_3|$, which is not independent, since

$$|\boldsymbol{k}_1 + \boldsymbol{k}_3|^2 = -K^2 - \tilde{K}^2 + k_1^2 + k_2^2 + k_3^2 + k_4^2\,. \tag{9}$$

This parametrization is not unique. We could choose any six geometrically independent quantities of the tetrahedra to parametrize the continuous degrees of freedom. For example, we could reparametrize by replacing K with the angle between $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$. Similarly, we can choose either parity or helicity to parametrize the discrete degree of freedom.

3 POP SPECTRA

CMB and LSS observables are consistent with cosmic structure originating from a single primordial scalar potential, $\Phi(\boldsymbol{x})$. In this section, we construct POP-spectrum-like observables, which we call POP spectra. These are compressions of the 6D parity-odd trispectrum down to a 1D power spectrum of composite fields. First, we construct the *vector POP spectrum*, defined below in equation (14), by cross-correlating a vector field with a pseudovector field, both derived as composite fields quadratic in Φ. Next, we construct the *scalar POP spectrum*, defined below in equation (24), by cross-correlating a scalar field with a pseudoscalar field. The latter is formed as a composite field that is cubic in Φ.

3.1 Vector POP spectrum

Our goal in this subsection is to construct composite vector and pseudovector fields out of a scalar field such that the cross-power spectrum between them is a compressed estimator of the parity-odd trispectrum. Since a vector is parity odd and a pseudovector is parity even, their cross-correlation will be a POP spectrum. Obtaining a vector field from a scalar is trivial: take its gradient, $\nabla\Phi(\boldsymbol{x})$. Obtaining a pseudovector is non-trivial. One possibility is to convolve the field with a filtering or smoothing function $f_a(\boldsymbol{x})$:

$$\Phi_a(\boldsymbol{x}) = \int \mathrm{d}^3y\; f_a(\boldsymbol{x} - \boldsymbol{y})\Phi(\boldsymbol{y}) \tag{10}$$

$$= \int_{\boldsymbol{k}} f_a(\boldsymbol{k})\Phi(\boldsymbol{k}) e^{i\boldsymbol{k}\cdot\boldsymbol{x}}\,, \tag{11}$$

where the integral notation is defined in Table 1. For two distinct filtering functions, $f_a(\boldsymbol{x})$ and $f_b(\boldsymbol{x})$, the composite vector field $\boldsymbol{V}_{ab}(\boldsymbol{x}) \equiv$

Table 1. Here, we list our Fourier-space integral notation and conventions.

Notation	Definition
$\int_{\boldsymbol{k}}$	$\int \frac{\mathrm{d}^3k}{(2\pi)^3}$
$\int_{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_n}$	$\int_{\boldsymbol{q}_1}\cdots\int_{\boldsymbol{q}_n}$
$\int_{q}$	$\int_0^\infty \frac{q^2\mathrm{d}q}{2\pi^2}$
$\int_{q_1,\ldots,q_n}$	$\int_{q_1}\cdots\int_{q_n}$

The bold symbols indicate wave vectors and the integral is over all of Fourier space. The non-bold symbols indicate the wave vector magnitude.

$\Phi_a(\boldsymbol{x})\nabla\Phi_b(\boldsymbol{x})$ is not the gradient of a scalar, so it also has a curl component. Therefore, we can carry out a Helmholtz decomposition,

$$\boldsymbol{V}_{ab}(\boldsymbol{x}) = \nabla\phi_{ab}(\boldsymbol{x}) - \nabla\times\boldsymbol{A}_{ab}(\boldsymbol{x})\,. \tag{12}$$

Since $\boldsymbol{V}_{ab}$ and the spatial derivatives are parity odd, $\boldsymbol{A}_{ab}$ is a pseudovector. Notice that if the two fields, Φ_a and Φ_b, were identical, $\boldsymbol{A}_{ab}$ would vanish. Although one of these can be the original, unfiltered field. We impose that its divergence vanishes because any possible divergence would not contribute to $\boldsymbol{V}_{ab}(\boldsymbol{x})$. Then we solve for its modes by taking the curl of $\boldsymbol{V}_{ab}(\boldsymbol{x})$ and solving the resulting Poisson equation,

$$\boldsymbol{A}_{ab}(\boldsymbol{k}) = -\frac{1}{k^2}i\boldsymbol{k}\times\boldsymbol{V}_{ab}(\boldsymbol{k})\,. \tag{13}$$

In general, we are free to choose four distinct filter functions: $f_a(\boldsymbol{x})$, $f_b(\boldsymbol{x})$, $f_c(\boldsymbol{x})$, and $f_d(\boldsymbol{x})$. From the first two, we define the composite vector field $\boldsymbol{V}_{ab}(\boldsymbol{x})$. From the second two, we define the composite vector field $\boldsymbol{V}_{cd}(\boldsymbol{x})$ and then isolate the pseudovector field $\boldsymbol{A}_{cd}(\boldsymbol{x})$ from it. The dot product $\boldsymbol{V}_{ab}(\boldsymbol{x})\cdot\boldsymbol{A}_{cd}(\boldsymbol{x})$ is parity odd. As a result, we can define our first POP spectrum as

$$\langle\boldsymbol{V}_{ab}(\boldsymbol{k})\cdot\boldsymbol{A}_{cd}(\boldsymbol{k}')\rangle = (2\pi)^3\delta_{\rm D}^{(3)}\left(\boldsymbol{k}+\boldsymbol{k}'\right)P_{\rm vector}(k)\,. \tag{14}$$

Explicitly, we can express the power spectrum as

$$P_{\rm vector}(k) = \frac{1}{\mathcal{N}_{\rm vector}(k)}\int_{\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_3,\boldsymbol{q}_4}\Big[(2\pi)^6\delta_{\rm D}^{(3)}\left(\boldsymbol{q}_1+\boldsymbol{q}_2-\boldsymbol{k}\right)\,\delta_{\rm D}^{(3)}\left(\boldsymbol{q}_3+\boldsymbol{q}_4+\boldsymbol{k}\right)\,f_a(\boldsymbol{q}_1)f_b(\boldsymbol{q}_2)f_c(\boldsymbol{q}_3)f_d(\boldsymbol{q}_4)$$
$$\times\,\left[\boldsymbol{k}\cdot(\boldsymbol{q}_1\times\boldsymbol{q}_3)\right]^2\,\tau_-(q_1,q_2,q_3,q_4,k,|\boldsymbol{q}_1+\boldsymbol{q}_4|)\Big]\,. \tag{15}$$

The wave vector integrals are defined in Table 1. The normalization factor $\mathcal{N}_{\rm vector}(k)$ is somewhat arbitrary, but if chosen poorly, the estimator will depend strongly on the geometry and resolution of the region where we sample $\Phi(\boldsymbol{x})$. A sensible choice is

$$\mathcal{N}_{\rm vector}(k) = \int_{\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_3,\boldsymbol{q}_4}(2\pi)^6\delta_{\rm D}^{(3)}\left(\boldsymbol{q}_1+\boldsymbol{q}_2-\boldsymbol{k}\right)\,\delta_{\rm D}^{(3)}\left(\boldsymbol{q}_3+\boldsymbol{q}_4+\boldsymbol{k}\right)\,f_a(\boldsymbol{q}_1)f_b(\boldsymbol{q}_2)f_c(\boldsymbol{q}_3)f_d(\boldsymbol{q}_4)\,\left[\boldsymbol{k}\cdot(\boldsymbol{q}_1\times\boldsymbol{q}_3)\right]^2\,, \tag{16}$$

so that equation (15) is a weighted average over the parity-odd trispectrum shape $\tau_-(q_1,q_2,q_3,q_4,k,|\boldsymbol{q}_1+\boldsymbol{q}_4|)$.

If we choose isotropic filtering functions that depend only on the magnitudes of the $\boldsymbol{q}_i$, the vector POP spectrum is a real-valued quantity that is non-vanishing only if the parity-odd trispectrum is non-vanishing. With isotropic filtering functions, the parity-even part of the trispectrum does not contribute to the imaginary part of the POP spectrum. To see this, we note that the parity-even trispectrum is invariant under $\boldsymbol{q}_i\to-\boldsymbol{q}_i$ while the triple product receives a minus sign. Therefore, the integrand is antisymmetric and the integral vanishes.[1]

The vector-pseudovector construction compresses the trispectrum down to a power spectrum by correlating two fields of order Φ^2. The modes of these fields represent triangles of wave vectors satisfying $\boldsymbol{k}+\boldsymbol{q}_1+\boldsymbol{q}_2=0$ and $\boldsymbol{k}'+\boldsymbol{q}_3+\boldsymbol{q}_4=0$. Computing the power spectrum enforces that the two triangles connect along a side of equal length, k, which imposed $\boldsymbol{k}+\boldsymbol{k}'=0$. Thus, the two triangles form a tetrahedron (shown in Fig. 1) as the trispectrum requires. The vector POP spectrum is a weighted average over all trispectrum configurations while holding fixed the diagonal side length k, which is the argument of this POP spectrum. For this reason, the vector POP carries crucial information about the soft limit where the diagonal approaches zero.

If the trispectrum has no explicit dependence on the diagonals and the filtering functions depend only on the magnitudes of the wave vectors in Fourier space, we can compute the angular integrals in equations (15) and (16) analytically. Then the expression for the vector POP

[1]The sign of the $\boldsymbol{k}$ in the Dirac delta functions does not matter, since $P_{\rm vector}(k)$ is a function of only the magnitude of $\boldsymbol{k}$.

spectrum simplifies (as detailed in Appendix A) to

$$P_{\rm vector}(k) = \frac{\pi^4}{2\mathcal{N}_{\rm vector}(k)} \int\limits_{q_1,q_2,q_3,q_4} \Big[\frac{q_1}{q_2}\,\Theta\big(\sin^2\theta_{12}\big)\sin^2\theta_{12}\,\frac{q_3}{q_4}\,\Theta\big(\sin^2\theta_{34}\big)\sin^2\theta_{34}\,f_a(q_1)f_b(q_2)f_c(q_3)f_d(q_4) \times\ \tau_-(q_1,q_2,q_3,q_4)\Big]\,, \tag{17}$$

and

$$\mathcal{N}_{\rm vector}(k) = \frac{\pi^4}{2} \int\limits_{q_1,q_2,q_3,q_4} \frac{q_1}{q_2}\,\Theta\big(\sin^2\theta_{12}\big)\sin^2\theta_{12}\,\frac{q_3}{q_4}\,\Theta\big(\sin^2\theta_{34}\big)\sin^2\theta_{34}\,f_a(q_1)f_b(q_2)f_c(q_3)f_d(q_4)\,. \tag{18}$$

The radial Fourier integrals are defined in Table 1. Here, $\Theta(x)$ is the Heaviside function,

$$\Theta(x) = \begin{cases} 0, & \text{if } x < 0 \\ \frac{1}{2}, & \text{if } x = 0 \\ 1, & \text{if } x > 0 \end{cases}\,, \tag{19}$$

and θ_{12} the angle between $\boldsymbol{q}_1$ and $\boldsymbol{k}$ in the first triangle, so

$$\sin^2\theta_{12} = 1 - \left(\frac{q_1^2 + k^2 - q_2^2}{2q_1k}\right)^2\,, \tag{20}$$

and similarly for θ_{34}, with $1, 2 \to 3, 4$. Notice that if q_i, q_j, and k fail to satisfy the triangle inequalities $|q_j - q_j| \le k \le q_i + q_j$, then the right-hand sides in equation (20) would be negative and the corresponding sine would be purely imaginary. Thus, the Heaviside functions in equations (17) and (18) enforce the triangle inequalities.

As a final remark, note that the pseudovector defined in equation (13) is not unique. For example, one could alternatively take the following curl,

$$\boldsymbol{B}_{ab}(\boldsymbol{x}) = \boldsymbol{\nabla}\Phi_a(\boldsymbol{x}) \times \boldsymbol{\nabla}\Phi_b(\boldsymbol{x})\,. \tag{21}$$

Using this instead of $\mathbf{A}_{ab}$ simply multiplies the vector POP spectrum $P_{\rm vector}(k)$ by a factor of $-k^2$. Instead of applying separate filtering functions to each field to the cross-product does not vanish, we could define an antisymmetric convolution kernel. Using convolution kernels is similar to the skew spectrum (Schmittfull et al. 2015), and is the method used for clustering fossil estimators (Jeong & Kamionkowski 2012).[2] These are better suited to model-specific optimal estimators, whereas in this work we are taking a more model-independent approach.

3.2 Scalar POP spectrum

We can construct another POP spectrum by cross-correlating a scalar field with a pseudoscalar field. The construction of a pseudoscalar, similar to that of the pseudovector from the previous subsection, is non-trivial and requires convolving the scalar field with filtering functions. In this case, we first construct the pseudovector given in equation (21). We have the freedom to further smooth this pseudovector by another filtering function, with modes

$$\boldsymbol{B}_{bcd}(\mathbf{Q}) = -f_d(\mathbf{Q}) \int\limits_{\boldsymbol{q}_3,\boldsymbol{q}_4} (2\pi)^3\delta_{\rm D}^{(3)}\left(\boldsymbol{q}_3 + \boldsymbol{q}_4 - \mathbf{Q}\right) f_b(\boldsymbol{q}_3)f_c(\boldsymbol{q}_4)(\boldsymbol{q}_3 \times \boldsymbol{q}_4)\Phi(\boldsymbol{q}_3)\Phi(\boldsymbol{q}_4)\,. \tag{22}$$

Then, we construct the triple product field

$$\Psi_{abcd}(\boldsymbol{x}) = \boldsymbol{\nabla}\Phi_a(\boldsymbol{x}) \cdot \boldsymbol{B}_{bcd}(\boldsymbol{x})\,, \tag{23}$$

which is a pseudoscalar. Its cross-power spectrum with $\Phi(\boldsymbol{k})$ is a POP spectrum,

$$\langle\Phi(\boldsymbol{k})\Psi_{abcd}(\boldsymbol{k}')\rangle = (2\pi)^3\delta_{\rm D}^{(3)}\left(\boldsymbol{k} + \boldsymbol{k}'\right) P_{\rm scalar}(k)\,. \tag{24}$$

The scalar POP spectrum is given by

$$P_{\rm scalar}(k) = \frac{1}{\mathcal{N}_{\rm scalar}(k)} \int\limits_{\boldsymbol{q}_2,\boldsymbol{q}_3,\boldsymbol{q}_4,\boldsymbol{Q}} \Big[(2\pi)^6\delta_{\rm D}^{(3)}\left(\boldsymbol{k} + \boldsymbol{q}_2 + \boldsymbol{Q}\right)\,\delta_{\rm D}^{(3)}\left(\boldsymbol{q}_3 + \boldsymbol{q}_4 - \boldsymbol{Q}\right)\,f_a(\boldsymbol{q}_2)f_b(\boldsymbol{q}_3)f_c(\boldsymbol{q}_4)f_d(\boldsymbol{Q}) \times\ \left[\boldsymbol{Q}\cdot(\boldsymbol{q}_2 \times \boldsymbol{q}_4)\right]^2\,\tau_-(k, q_2, q_3, q_4, Q, |\boldsymbol{q}_2 + \boldsymbol{q}_3|)\Big]\,, \tag{25}$$

[2]There are conditions under which the vector-mode clustering fossil estimator and our vector POP spectrum exactly coincide (up to the choice of normalization and inverse-variance-weighting). Referring to equation (1) of Jeong & Kamionkowski (2012), if their mode coupling kernels for right-/left-handed vector modes are the identical and have the form $f_{\rm R/L}(\boldsymbol{k}_1, \boldsymbol{k}_2) = f_a(k_1)f_b(k_2) - f_b(k_1)f_a(k_2)$, then this is the same construction as our vector POP spectrum in the case where $f_c(k) = f_a(k)$ and $f_d(k) = f_b(k)$. There is not yet an equivalent or analogous estimator in the clustering fossil approach for the scalar POP spectrum that we present in the next section.

with normalization

$$\mathcal{N}_{\rm scalar}(k) = \int\limits_{\boldsymbol{q}_2,\boldsymbol{q}_3,\boldsymbol{q}_3,\boldsymbol{Q}} (2\pi)^6 \delta_{\rm D}^{(3)}\left(\boldsymbol{k}+\boldsymbol{q}_2+\boldsymbol{Q}\right)\, \delta_{\rm D}^{(3)}\left(\boldsymbol{q}_3+\boldsymbol{q}_4-\boldsymbol{Q}\right)\, f_a(\boldsymbol{q}_2) f_b(\boldsymbol{q}_3) f_c(\boldsymbol{q}_4) f_d(\boldsymbol{Q}) \left[\boldsymbol{Q}\cdot(\boldsymbol{q}_2\times\boldsymbol{q}_4)\right]^2 . \tag{26}$$

As with the vector-pseudovector construction, we choose the normalization so that the POP spectrum is a weighted average over the imaginary trispectrum. In this case, we have constructed the compressed four-point statistic by cross-correlating a pseudoscalar of order Φ^3 with the original potential Φ. The resulting power spectrum is an average over the trispectrum tetrahedra, fixing the magnitude of one of the wave vector sides. The average integrates over the diagonal $\mathbf{Q}$ this time (corresponding to the k in equations (14)–(18), and the scalar POP spectrum is a function of only the fixed wave vector magnitude k, a side length of the tetrahedron. This property makes the scalar POP particularly useful for probing the soft limit where one of the side lengths approaches zero.

The side length that we label k here was labelled q_1 for the vector POP spectrum in equations (15)–(18), so in the following equations a subscript 1 will refer to this side length. Again, if the trispectrum does not depend on the diagonals and the filtering functions depend only on the magnitudes of the wave vectors, we can compute the angular integrals analytically, and the expression simplifies (as detailed in Appendix A) to

$$P_{\rm scalar}(k) = \frac{\pi^4}{2k\mathcal{N}_{\rm scalar}(k)} \int\limits_{q_2,q_3,q_3,Q} \Big[q_2\,\Theta\left(\sin^2\theta_{21}\right)\sin^2\theta_{21}\,\frac{q_4}{q_3}\,\Theta\left(\sin^2\theta_{43}\right)\sin^2\theta_{43}\, f_a(q_2) f_b(q_3) f_c(q_4) f_d(Q) \times\ \tau_-(k,q_2,q_3,q_4)\Big] , \tag{27}$$

and

$$\mathcal{N}_{\rm scalar}(k) = \frac{\pi^4}{2k} \int\limits_{q_2,q_3,q_3,Q} q_2\,\Theta\left(\sin^2\theta_{21}\right)\sin^2\theta_{21}\,\frac{q_4}{q_3}\,\Theta\left(\sin^2\theta_{43}\right)\sin^2\theta_{43}\, f_a(q_2) f_b(q_3) f_c(q_4) f_d(Q) . \tag{28}$$

Here, we evaluate $\sin^2\theta_{21}$ and $\sin^2\theta_{43}$ using

$$\sin^2\theta_{21} = 1 - \left(\frac{q_2^2+Q^2-k^2}{2q_2Q}\right)^2 , \tag{29}$$

$$\sin^2\theta_{43} = 1 - \left(\frac{q_4^2+Q^2-q_3^2}{2q_4Q}\right)^2 . \tag{30}$$

3.3 Constructing the estimators

To construct the vector POP spectrum from a scalar field $\Phi(\boldsymbol{x})$ on a discrete grid, we start by taking its Fourier transform, $\Phi(\boldsymbol{q})$. Then, we make four copies of these modes and rescale them by the filter functions $\{f_a(\boldsymbol{q}), f_b(\boldsymbol{q}), f_c(\boldsymbol{q}), f_d(\boldsymbol{q})\}$, obtaining $\{\Phi_a(\boldsymbol{q}), \Phi_b(\boldsymbol{q}), \Phi_c(\boldsymbol{q}), \Phi_d(\boldsymbol{q})\}$. Multiplying the modes of $\Phi_b(\boldsymbol{q})$ and $\Phi_d(\boldsymbol{q})$ by $i\boldsymbol{q}$ and then inverse Fourier transforming yields the modes of $\nabla\Phi_b(\boldsymbol{x})$ and $\nabla\Phi_d(\boldsymbol{x})$. We also inverse Fourier transform the other two fields, Φ_a and Φ_c and, now in position space, form the products $\boldsymbol{V}_{ab}(\boldsymbol{x}) = \Phi_a(\boldsymbol{x})\nabla\Phi_b(\boldsymbol{x})$ and $\boldsymbol{V}_{cd}(\boldsymbol{x}) = \Phi_c(\boldsymbol{x})\nabla\Phi_d(\boldsymbol{x})$. Next, we Fourier transform both of these and, from $\boldsymbol{V}_{cd}(\boldsymbol{q})$, solve for the modes of $\boldsymbol{A}_{cd}(\boldsymbol{q})$ using equation (13). The unbinned vector POP spectrum estimator is

$$P_{\rm vector}(k) = \frac{1}{V_{\rm box}N_{\rm vector}(k)} \sum_{|\boldsymbol{q}|=k} \boldsymbol{V}_{ab}(\boldsymbol{q})\cdot\boldsymbol{A}^*_{cd}(\boldsymbol{q}) , \tag{31}$$

where the sum is over all modes with equal wave vector magnitudes on the grid.

The normalization is computed by initializing four grids of modes corresponding to the filter functions: $f_a(\boldsymbol{q})$, $f_b(\boldsymbol{q})$, $f_c(\boldsymbol{q})$, and $f_d(\boldsymbol{q})$. From here, obtaining the normalization corresponding to equation (16) is somewhat complicated. When constructing the power spectrum, we get one triple product from the estimator and another from the parity-odd trispectrum. Overall, the power spectrum involves the square of the triple product. To get a squared triple product in the normalization factor, we need to compute the Hessian of the field $f_a(\boldsymbol{x})$ and construct the composite tensor field with components

$$F^{ij}_{ab}(\boldsymbol{x}) = f_a(\boldsymbol{x})\nabla^i\nabla^j f_b(\boldsymbol{x}) . \tag{32}$$

After Fourier transforming the Hessian of the filter function fields, we take the curl of one of the two components,

$$J^{ij}_{ab}(\boldsymbol{q}) \equiv \epsilon^{imn} i k^m F^{nj}_{ab}(\boldsymbol{q}) , \tag{33}$$

where repeated indices imply summation. Following the same steps for $f_c(\boldsymbol{x})$ and $f_d(\boldsymbol{x})$, the desired normalization is

$$N_{\rm vector}(k) = -\sum_{|\boldsymbol{q}|=k} J^{ij}_{ab}(\boldsymbol{q})\left[J^{ij}_{cd}(\boldsymbol{q})\right]^\dagger . \tag{34}$$

The dagger † denotes the Hermitian conjugate of the matrix. We then bin the estimator in wave number bins of width Δk,

$$P_{\rm vector}(k|\Delta k) = \frac{1}{N_{\rm vector}(k|\Delta k)} \sum_{q\in k} P_{\rm vector}(q) N_{\rm vector}(q)\,, \tag{35}$$

where

$$N_{\rm vector}(k|\Delta k) = \sum_{q\in k} N_{\rm vector}(q)\,. \tag{36}$$

Here, the sum over $q \in k$ means summing over all $q \in [k - \Delta k/2, k + \Delta k/2)$.

The computation of the scalar POP spectrum follows similar steps. In this case, we compute the gradients $\nabla\Phi_a(\boldsymbol{x})$, $\nabla\Phi_b(\boldsymbol{x})$ and $\nabla\Phi_c(\boldsymbol{x})$. We use the latter two to form $\boldsymbol{B}_{bc}(\boldsymbol{x})$ according to equation (21). We Fourier transform this and then multiply by the filtering function modes $f_d(\boldsymbol{q})$, resulting in the modes of $\boldsymbol{B}_{bcd}(\boldsymbol{q})$. We then inverse Fourier transform this, and its dot product with $\nabla\Phi_a(\boldsymbol{x})$ yields $\Psi_{abcd}(\boldsymbol{x})$. The unbinned scalar POP spectrum is

$$P_{\rm scalar}(k) = \frac{1}{V_{\rm box} N_{\rm scalar}(k)} \sum_{|\boldsymbol{q}|=k} \Phi(\boldsymbol{q})\,[\Psi_{abcd}(\boldsymbol{q})]^*\,. \tag{37}$$

Constructing the normalization is again non-trivial due to the squared triple product factor. In this case, we require the Hessians of the field $f_a(\boldsymbol{x})$, $f_b(\boldsymbol{x})$, and $f_c(\boldsymbol{x})$. From the latter two, we construct the tensor field $F^{ij}_{bc}(\boldsymbol{x})$ as in equation (32). By Fourier transforming all components and multiplying the modes by the filter function $f_d(\boldsymbol{q})$, we obtain $F^{ij}_{bcd}(\boldsymbol{q}) \equiv f_d(\boldsymbol{q}) F^{ij}_{bc}(\boldsymbol{q})$. Inverse Fourier transforming this field and multiplying by the Hessian of $f_a(\boldsymbol{x})$ gives

$$G^{ij}_{abcd}(\boldsymbol{x}) = \epsilon^{imn}\epsilon^{jpq} F^{mp}_{bcd}(\boldsymbol{x}) \nabla^n \nabla^q f_a(\boldsymbol{x})\,. \tag{38}$$

Fourier transforming this, the scalar POP spectrum normalization is

$$N_{\rm scalar}(k) = -\sum_{|\boldsymbol{q}|=k} k^i k^j G^{ij}_{abcd}(\boldsymbol{q})\,. \tag{39}$$

The binned estimator is

$$P_{\rm scalar}(k|\Delta k) = \frac{1}{N_{\rm scalar}(k|\Delta k)} \sum_{q\in k} P_{\rm scalar}(q) N_{\rm scalar}(q)\,, \tag{40}$$

where

$$N_{\rm scalar}(k|\Delta k) = \sum_{q\in k} N_{\rm scalar}(q)\,. \tag{41}$$

If we choose filter functions that only select wave vector magnitude shells of width Δk, then both POP spectra coincide and equal the binned trispectrum estimator in equation (A30). This fact illustrates the computational advantage of our POP spectra. Evaluating all the bins of one POP spectrum is computationally equivalent to constructing a 1D slice of the full trispectrum. The tradeoff is that we have significantly compressed the information of the full trispectrum. However, the presence of two estimators, each carrying complementary information from the parity-odd trispectrum, partially compensates for this reduction in information. We also have the freedom to choose the filter functions, which can be optimized. The extent of information loss could be further mitigated by considering extensions and generalizations of the POP spectra, which we explore in Section 5.

The algorithms that we presented above for constructing these estimators are efficient since the most costly operations are the Discrete Fourier Transforms (DFTs), which scale as $N \log N$, with N the number of grid points. All other operations are point-wise multiplications, which scale as N. Ignoring the normalization factors, the vector POP spectrum requires 15 and the scalar POP spectrum requires 17 3D DFTs. Constructing the estimators without the normalization factor takes about 6 s for the vector and 8 s for the scalar POP spectrum running on a single node with 128 cores. The normalization factors are more expensive to compute, but these only need to be computed once and can be saved and reused since they are not data-dependent.

4 VALIDATION OF THE ESTIMATORS

4.1 Random parity-violating realizations

To validate our POP estimators, we generate a set of random primordial curvature fields that include a specific parity-violating trispectrum. We start by generating the modes of a Gaussian random field primordial curvature perturbation $\zeta_{\rm G}(\boldsymbol{k})$,

$$\zeta_{\rm G}(\boldsymbol{k}) = \sqrt{\frac{V_{\rm Box}\,\pi^2 A_{\rm s}}{k^3}\left(\frac{k}{k_{\rm p}}\right)^{n_{\rm s}-1}}\,\big[N_1(\boldsymbol{k}) + i N_2(\boldsymbol{k})\big], \tag{42}$$

where N_1 and N_2 are drawn from a standard normal distribution, $A_{\rm s}$ is the amplitude of the primordial power spectrum, $n_{\rm s}$ is the spectral tilt, and $k_{\rm p} = 0.05\ {\rm Mpc}^{-1}$ is the pivot scale. We then transform this Gaussian random field into a non-Gaussian field in real space as (Coulton et al.

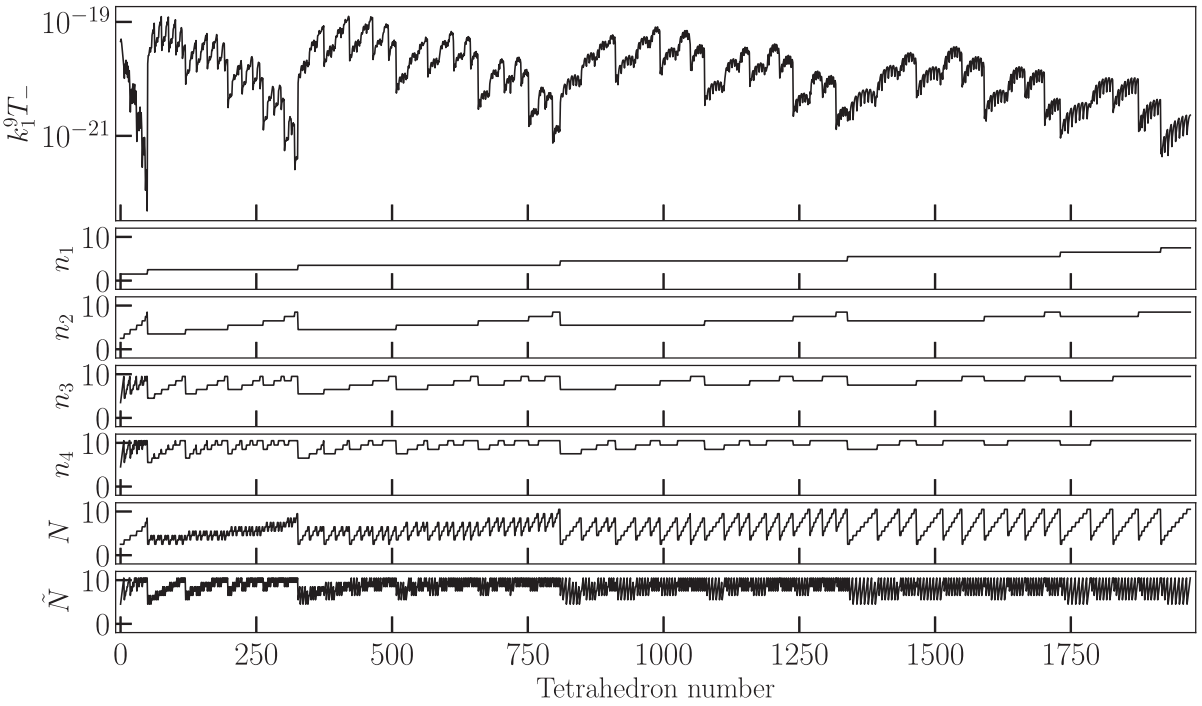

Figure 2. Here we show the trispectrum in the thin bin limit (top panel) and the binning scheme (lower six panels). The shape of the trispectrum template from equation (45) appears in the top panel. The bottom six panels indicate the binning scheme with $n_i = k_i/k_{\rm F}$, $N = K/k_{\rm F}$, and $\tilde{N} = \tilde{K}/k_{\rm F}$. Here, $k_{\rm F}$ is the fundamental mode of a box interpreted as having length 4 Gpc h^{-1}. We impose $k_1 < k_2 < k_3 < k_4$. We have used 10 k-bins of width $k_{\rm F}$.

2024)

$$\zeta(\boldsymbol{x}) = \zeta_{\rm G}(\boldsymbol{x}) + g_- \nabla\zeta_{\rm G}^{[\alpha]}(\boldsymbol{x}) \cdot \left[\nabla\zeta_{\rm G}^{[\beta]}(\boldsymbol{x}) \times \nabla\zeta_{\rm G}^{[\gamma]}(\boldsymbol{x})\right] . \tag{43}$$

The coefficient g_- controls the amplitude of the primordial parity-odd trispectrum, and the modes of the fields in the triple product are

$$\zeta_{\rm G}^{[\alpha]}(\boldsymbol{k}) = k^\alpha \zeta_{\rm G}(\boldsymbol{k}) . \tag{44}$$

The leading order imaginary part of the trispectrum for this template is given by

$$T_-(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) = g_- \, \boldsymbol{k}_1 \cdot (\boldsymbol{k}_2 \times \boldsymbol{k}_3) \left(2\pi^2 A_{\rm s}\right)^3 \left(k_1^{\alpha-4+n_{\rm s}} k_2^{\beta-4+n_{\rm s}} k_3^{\gamma-4+n_{\rm s}} k_4^0 \mp 23 \text{ signed permutations}\right) . \tag{45}$$

The 24 terms in the parenthesis are the permutations of $\{k_1, k_2, k_3, k_4\}$. Even permutations (with an even number of transpositions) get a positive sign, and odd permutations get a negative sign in the sum. We have included the k_4^0 factor to emphasize that every term has one of the four wave vector magnitudes raised to the exponent zero.

The trispectrum is scale invariant if $\alpha + \beta + \gamma = -3$. To preserve the large-scale power spectrum we also restrict these exponents to be less than or equal to zero. We choose $\alpha = -2$, $\beta = -1$, and $\gamma = 0$, which is the same template used in Coulton et al. (2024). For simplicity, we choose $A_{\rm s} = 2 \times 10^{-9}$, $n_{\rm s} = 1$, and $g_- = \pm 10^6$. In Fig. 2, we display the shape of this trispectrum.

Choosing a binning scheme where $k_1 < k_2 < k_3 < k_4$, the trispectrum shape function is dominated by the terms

$$\tau_-(k_1, k_2, k_3, k_4) \simeq \frac{g_-(2\pi^2 A_s)^3}{k_1^5 k_2^4} \left(\frac{1}{k_3^3} - \frac{1}{k_4^3}\right) . \tag{46}$$

This shape peaks when the wave vectors $\boldsymbol{k}_1$, $\boldsymbol{k}_2$, and $\boldsymbol{k}_3$ are parallel so that $k_4 = k_1 + k_2 + k_3$. The diagonals for this configuration are $K = k_1 + k_2$ and $\tilde{K} = k_4 - k_1$. However, this configuration is insensitive to parity, since the triple product vanishes when any two wave vectors are colinear. The trispectrum peaks on a tetrahedron shape that deviates from colinearity with diagonals $K < k_1 + k_2$ and $\tilde{K} > k_4 - k_1$ that maximize the product of the right-hand side of equation (46) and the triple product $\boldsymbol{k}_1 \cdot (\boldsymbol{k}_2 \times \boldsymbol{k}_3)$. This peak analysis is true for any scale-invariant trispectrum of the form in equation (45), although the particular diagonals that maximize the trispectrum depend on the values of the template exponents. These parity-odd trispectrum templates diverge in the soft limit $k_1 \to 0$.

Since we are dealing with a scale-invariant primordial potential, the box length is arbitrary, but we interpret it as $L_{\rm Box} = 4$ Gpc h^{-1} so that the scales involved are relevant for CMB and LSS. We generate 64 random pairs of these non-Gaussian potential fields on a grid of size $N_{\rm grid} = 512^3$. The pairs have the same underlying Gaussian realization but opposite signs for their non-linear terms. Subtracting the POP spectra measured on these pairs suppresses cosmic variance from the purely Gaussian part of the fields. As can be seen in Fig. 3, the cosmic

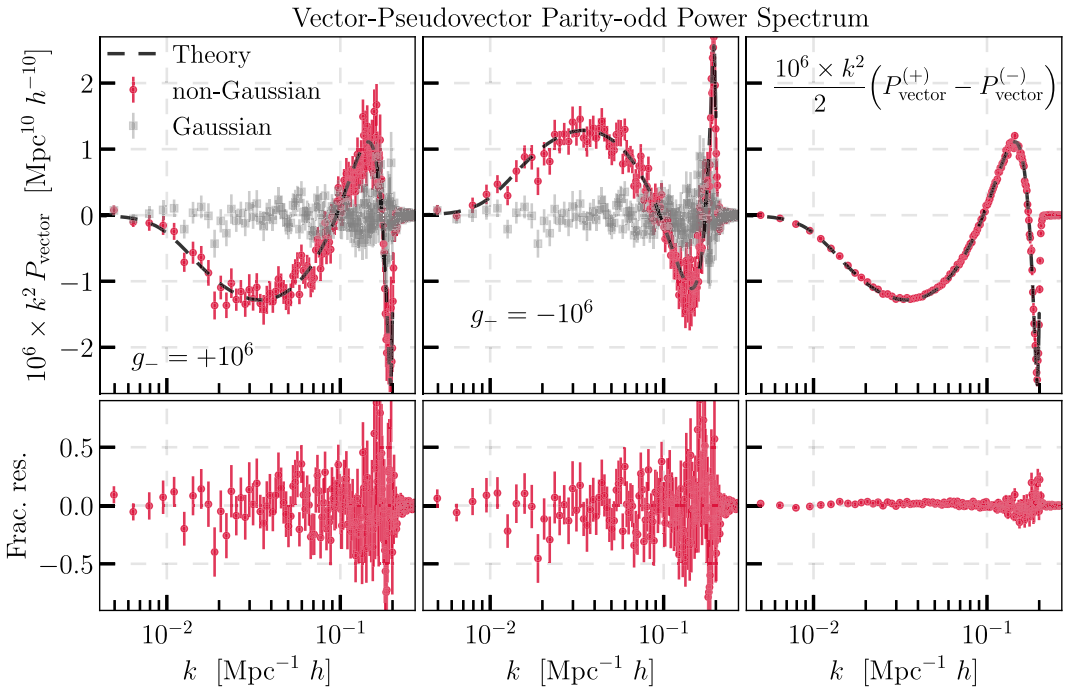

Figure 3. The vector POP spectrum of the primordial potential defined in equation (43). In the left panels, the amplitude of the primordial trispectrum is $g_- = 10^6$, while in the middle panels $g_- = -10^6$. We used the same underlying Gaussian realizations for both signs of g_-. The right panels show the variance-suppressed estimator, taking the difference between results with positive and negative g_-. The dashed curves are the expected signal, computed from equations (17) and (18). The bottom row shows fractional residuals of the simulated data with respect to the semi-analytical calculation. Square data points show the estimator on a Gaussian random field. These scatter around zero and are consistent with a vanishing parity-violating signal. All data points have error bars obtained by bootstrap averaging over the 64 simulations, resampling with replacement 10^5 times. In the right-most panel, the sample variance cancellation is not perfect. This is because the POP spectrum involves four weakly non-Gaussian fields.

variance is not perfectly cancelled. This is because the POP spectrum is a compression of the four-point correlator of a weakly non-Gaussian field.

There is a considerable amount of freedom in our estimators since we can choose any set of filtering functions. For the vector POP spectrum, $f_a(k)$ must be distinct from $f_b(k)$, and $f_c(k)$ must be distinct from $f_d(k)$, otherwise $P_{\rm vector}(k)$ vanishes identically. For the scalar POP spectrum, $f_a(k)$, $f_b(k)$, and $f_c(k)$ must all be distinct for $P_{\rm scalar}(k)$ not to vanish. Here, we choose

$$f_a(k) = k^2\Theta(k - k_{\rm min})\Theta(k_{\rm max} - k)\,, \tag{47}$$

$$f_c(k) = k^{-2}\Theta(k - k_{\rm min})\Theta(k_{\rm max} - k)\,, \tag{48}$$

$$f_b(k) = f_d(k) = \Theta(k - k_{\rm min})\Theta(k_{\rm max} - k)\,. \tag{49}$$

Each filtering function involves the same scale cuts, $k_{\rm min} = 5 \times 10^{-3}$ Mpc^{-1} h and $k_{\rm max} = 2 \times 10^{-1}$ Mpc^{-1} h, which selects a thick spherical shell of modes. As long as the fundamental mode in the box is less than $k_{\rm min}$, and the Nyquist mode is greater than $k_{\rm max}$, the box contains the full range of modes and the estimator should be independent of the box size and resolution. However, the discreteness of the grid is noticeable on large scales if $k_{\rm min}$ is very close to the fundamental mode.

The autocorrelation of the non-linear term in equation (43) will introduce corrections to the power spectrum. These corrections are of the order $\sim g_-^2 A_{\rm s}^3$ and are subdominant compared to the Gaussian power spectrum for our template at the scales we consider. The non-Gaussian corrections to the power spectrum grow at small scales and increase the power of the Nyquist mode, $k = 0.4$ Mpc^{-1} h, in our simulated boxes by about 1 per cent. After imposing the scale cuts in our POP spectra the corrections to the power spectrum are far below percent level for the modes we analyse. These corrections to the power spectrum can be removed by further rescaling the non-Gaussian field (Coulton et al. 2024). However, corrections to the power spectrum could be a physical effect. The (inflationary) mechanisms that generate the primordial trispectrum can also affect the shape of the power spectrum. This template does not introduce a primordial bispectrum because all contributions would involve an odd number of Gaussian fields, resulting in a vanishing expectation value.

We have omitted many complications of observational survey data in this analysis. Cosmological surveys target observables that are biased tracers of the underlying matter field, which has undergone non-linear evolution. Accurate modelling will require the linear matter transfer function, the bias expansion, and possibly the non-linearity of gravitational clustering depending on the scales considered. Surveys

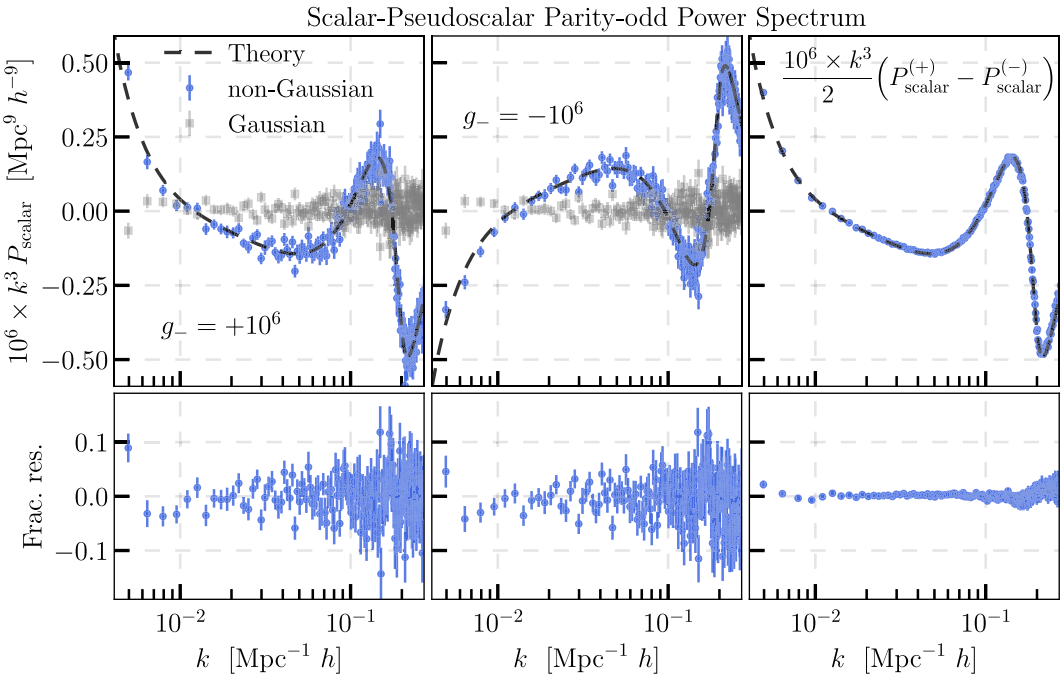

Figure 4. Same as Fig. 3 but for the scalar POP spectrum. In this case, the signal grows in the limit $k \to 0$, showing sensitivity to the soft limit of the trispectrum as one of the tetrahedron wave vector side lengths becomes small. In contrast, Fig. 3 demonstrates decreasing sensitivity in the diagonal soft limit of the vector POP spectrum because the specific trispectrum template that we simulate does not peak in this limit. The behaviour of the residual in the right-most panel is again (as in Fig. 3) due to the additional sample variance contributed by the weakly non-Gaussian nature of the fields.

observe objects in redshift space where isotropy does not hold. Observational analyses must also model the survey mask, selection function, and shot noise. We will investigate the impact of these aspects of observational data on the POP spectra in future work.

4.2 COMPARISON WITH ANALYTICAL COMPUTATION

In Fig. 3, we display the vector POP spectrum compared with the theoretical expectation obtained by numerically evaluating the integrals in equations (17) and (18). The left panel shows the results for 64 simulations with $g_- = +10^6$. The middle panel shows results for the same underlying Gaussian random fields but with $g_- = -10^6$. We significantly reduce cosmic variance by computing the difference between these two estimators, as illustrated in the right panel of Fig. 3. We estimated the error bars by bootstrap averaging over the estimators constructed for the 64 realizations, resampling 10^5 times with replacement.

We obtained the semi-analytic result in Fig. 3 by evaluating equation (17) on a discretized lattice for the radial integrals over all q_i and $k = |\boldsymbol{q}_1 + \boldsymbol{q}_2|$. Specifically, for a given bin, we construct a 3D grid of values for $\{q_1, q_2, k\}$, with q_1 and q_2 spaced uniformly between $k_{\rm min}$ and $k_{\rm max}$ and k spanning the width of the power spectrum bin. We then mask out regions that violate the triangle inequalities. Next, we construct a second grid for $\{q_3, q_4, k\}$, masking the non-triangular regions. Finally, we integrate equation (17) term by term. For each permutation term in the trispectrum of equation (45), we construct the integrand for q_1 and q_2, and compute the discretized integral using Romberg integration with $2^8 + 1$ sample points. We do the same for q_3 and q_4, except we include explicit factors of k appearing in equation (45) in this grid. We then multiply the results from the two grids. Finally, we compute the discretized integral of the diagonal k using Romberg integration with $2^7 + 1$ sample points. The procedure for integrating equation (18) is the same, omitting the factors from the primordial trispectrum template shape. We parallelize this calculation by computing multiple bins simultaneously. From Fig. 3, we see excellent agreement with the semi-analytic calculation of the expected signal. This agreement demonstrates that we accurately recover the expected signal from the trispectrum shape injected into the data.

In Fig. 4, we display similar results for the scalar POP spectrum. The theoretical calculation of equations (27) and (28) differs from the vector case in that k in the triangle $\{k, q_2, Q\}$ is integrated over the estimator's bin width spanning the range between $k_{\rm min}$ and $k_{\rm max}$ and the filter function $f_d(Q)$ applies to the diagonal Q. Again, we find excellent agreement between our simulated data and the expected shape of $P_{\rm scalar}(k)$. The bias at low k is due to the discreteness of the grid, compared with the continuous-limit integrals computed in the semi-analytic prediction. This bias is not noticeable for the vector estimator in Fig. 3 due to the signal vanishing at low k.

As a final remark note that, unlike the vector POP spectrum, the scalar POP spectrum increases as k approaches zero. This occurs because the argument k of the scalar POP spectrum corresponds to the side length of the tetrahedron. Our specific trispectrum template, equation (45), diverges as one side length goes to zero, so the scalar POP spectrum also diverges in this limit. If we chose a template that peaks in the soft limit where a diagonal approaches zero, the vector POP spectrum would increase as its argument approaches zero. We could also construct a

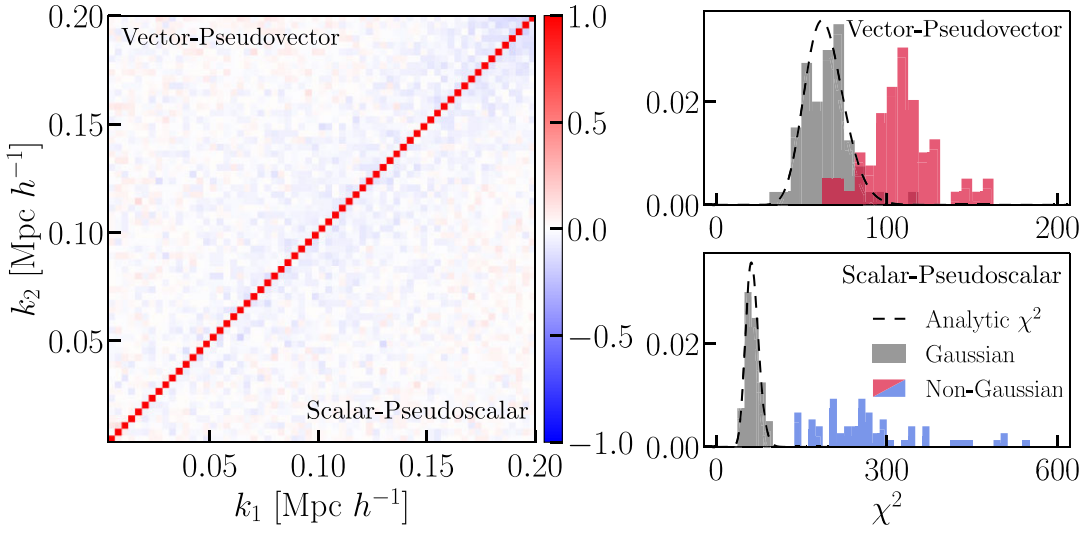

Figure 5. On the left panel, we display the correlation matrix, equation (50), for the vector-pseudovector (upper triangle) and scalar-pseudoscalar (lower triangle) estimators. The correlation matrix is dominated by the diagonal, with off-diagonal contributions less than 10 per cent. On the right panels, we display histograms of the χ^2 values from 64 random non-Gaussian fields compared with the distribution from Gaussian realizations. The peak of the dashed analytical χ^2 curve coincides with the number of bins analysed.

template that diverges in the soft side length and soft diagonal limits. Then the vector and scalar POP spectra would increase as their arguments go to zero. This demonstrates that the $k \to 0$ limits of our POP spectra encode crucial and complementary information about the soft limits of the trispectrum.

4.3 Correlation matrix and χ^2 test

The POP spectra integrate over many configurations to obtain a power spectrum from a trispectrum. For this reason, we may worry that the covariance matrices of the POP spectra have complicated structures with large off-diagonal contributions. To test this, we generate 1024 realizations of random Gaussian fields, construct the vector and scalar POP spectra estimators, and compute the covariance matrix for each,

$$C(k_i, k_j) = \langle P(k_i)P(k_j)\rangle - \langle P(k_i)\rangle\langle P(k_j)\rangle\,, \tag{50}$$

where k_i indicates the $i^{\rm th}$ wave number bin, and P can be either $P_{\rm vector}$(k) or $P_{\rm scalar}$(k). In Fig. 5 we display the correlation matrices,

$$R(k_i, k_j) = \frac{C(k_i, k_j)}{\sqrt{C(k_i, k_i)C(k_j, k_j)}}\,, \tag{51}$$

for the vector POP spectrum in the upper-right triangle and the scalar POP spectrum in the lower-left triangle. We find the correlation matrix is dominated by its diagonal for both estimators, with off-diagonal contributions less than 10 per cent.

We assess the sensitivity of our compressed estimators by computing the χ^2 from the inverse of the covariance matrix,

$$\chi^2 = \sum_{ij} P(k_i)C^{-1}(k_i, k_j)P(k_j)\,. \tag{52}$$

We compute the χ^2 values for all 64 of our parity-violating non-Gaussian fields with positive amplitude for the primordial trispectrum and their underlying Gaussian fields. Fig. 5 displays histograms of the χ^2 values. The scalar POP spectrum exhibits greater sensitivity for detecting primordial parity violation than the vector POP spectrum. We can understand this from the form of primordial non-Gaussianity in equation (43) that leads to the shape of the trispectrum in equation (45). This shape peaks as k_1, the argument of the scalar POP spectrum, goes to zero. We see this in the low-k behaviour in Fig. 4.

The sensitivity depends on the choice of filtering kernels and the shape of the primordial signal. Different parity-odd trispectrum templates will generally yield different χ^2 distributions with more or less sensitivity. In principle, we can find the optimal filter functions to maximize the signal-to-noise ratio of the POP power spectra for a given shape of the underlying primordial trispectrum, as was done for the CMB lensing estimator (Hu 2001), but we do not explore that in this work. Since the vector and scalar POP spectra contain complementary information, it would be best to do a joint analysis of both simultaneously. This joint analysis requires the joint covariance, which will have off-diagonal contributions between the vector and scalar POP spectra bins. We leave the analysis of the joint covariance matrix to future work.

5 EXTENSIONS AND GENERALIZATIONS

The examples of POP spectra explored above are the simplest examples of a more general class of constructions. In this section, we outline some of these generalizations.

5.1 Connection to the polarization basis

Before exploring some generalizations of our POP spectra, we first write down an alternative and equivalent definition of the vector POP spectrum based on polarization vectors, which will be useful later. As described in Section 2, parity violation is closely related to right- and left-handed helicities. Here we will demonstrate the relationship between our vector POP spectrum and these helicities. To do so, let us take two vector fields $\boldsymbol{V}_{ab}$ and $\boldsymbol{V}_{cd}$. From these, we can cross-correlate their Cartesian components as

$$\langle V^i_{ab}(\boldsymbol{k}) V^j_{cd}(\boldsymbol{k}')\rangle = (2\pi)^3 \delta^{(3)}_{\rm D}(\boldsymbol{k}+\boldsymbol{k}') P^{ij}(k)\,. \tag{53}$$

Assuming isotropy, we can decompose this power spectrum matrix into its trace, its traceless symmetric, and its antisymmetric parts,

$$P^{ij}(k) = \frac{1}{3}\delta^{ij} P_\parallel(k) + \left(\frac{k^i k^j}{k^2} - \frac{1}{3}\delta^{ij}\right) P_\perp(k) + i\epsilon^{ijk}\frac{k^k}{k} P_-(k)\,. \tag{54}$$

Assuming parity invariance, only the first two terms are present. These are the longitudinal and transverse parts of the parity-even power spectrum matrix. The final term is the parity-odd part of the power spectrum matrix.

One can choose left-/right-handed polarization vectors and form a complete basis $\{\boldsymbol{e}_\parallel(\boldsymbol{k}), \boldsymbol{e}_{\rm R}(\boldsymbol{k}), \boldsymbol{e}_{\rm L}(\boldsymbol{k})\}$ in Fourier space,

$$\boldsymbol{V}_{ab}(\boldsymbol{k}) = V_{ab,\parallel}(\boldsymbol{k})\boldsymbol{e}_\parallel(\boldsymbol{k}) + V_{ab,{\rm R}}(\boldsymbol{k})\boldsymbol{e}_{\rm R}(\boldsymbol{k}) + V_{ab,{\rm L}}(\boldsymbol{k})\boldsymbol{e}_{\rm L}(\boldsymbol{k})\,. \tag{55}$$

The transverse polarization vectors are defined as eigenvectors of the curl operator,

$$i\boldsymbol{k}\times\boldsymbol{e}_{\rm R/L}(\boldsymbol{k}) \equiv \pm k\,\boldsymbol{e}_{\rm R/L}(\boldsymbol{k})\,, \tag{56}$$

and $\boldsymbol{e}_\parallel(\boldsymbol{k}) \equiv i\boldsymbol{k}/k$. These quantities transform into one another after a parity transformation, as $\boldsymbol{e}_{\rm R/L}(-\boldsymbol{k}) = \pm i\boldsymbol{e}_{\rm L/R}(\boldsymbol{k})$. Therefore, one can define the eigenvectors of the parity operator as

$$\boldsymbol{e}_\pm(\boldsymbol{k}) = \frac{1}{\sqrt{2}}\left(\boldsymbol{e}_{\rm R}(\boldsymbol{k}) \pm i\boldsymbol{e}_{\rm L}(\boldsymbol{k})\right). \tag{57}$$

Using this alternative basis, the vector field is decomposed as

$$\boldsymbol{V}_{ab}(\boldsymbol{k}) = V_{ab,\parallel}(\boldsymbol{k})\boldsymbol{e}_\parallel(\boldsymbol{k}) + V_{ab,-}(\boldsymbol{k})\boldsymbol{e}_+(\boldsymbol{k}) + V_{ab,+}(\boldsymbol{k})\boldsymbol{e}_-(\boldsymbol{k})\,. \tag{58}$$

The $\pm$ labels of the coefficients are the opposite of the basis labels because the vector must be parity odd, so the coefficients transform with the opposite sign compared to the transverse basis vectors. We can then project the power spectrum matrix onto the polarization vectors,

$$P_{\rm RR/LL}(k) = P^{ij}(k) e^i_{\rm R/L}(\boldsymbol{k}) e^j_{\rm R/L}(-\boldsymbol{k})\,, \tag{59}$$

or

$$P_{+-}(k) = P^{ij}(k) e^i_+(\boldsymbol{k}) e^j_-(-\boldsymbol{k})\,. \tag{60}$$

It is straightforward to show that

$$P_-(k) = -\frac{k}{2} P_{\rm vector}(k) \tag{61}$$

$$= \frac{1}{2}\Big(P_{\rm RR}(k) - P_{\rm LL}(k)\Big) \tag{62}$$

$$= P_{+-}(k)\,, \tag{63}$$

where $P_{\rm vector}(k)$ is the vector POP spectrum.

5.2 Higher-order derivatives: angular dependence

The vector and scalar POP spectra that we have considered thus far are minimal in the sense that they involve the minimum number of spatial derivatives required to form the parity-odd triple product. We are free to add any number of spatial derivatives to these constructions, forming higher-rank tensor POP spectra. For example, we consider adding an additional derivative to the vector and pseudovector fields from the vector POP spectrum, as

$$H^{ij}_{\rm 1ab}(\boldsymbol{x}) = \nabla^i \Phi_{\rm a}(\boldsymbol{x}) \nabla^j \Phi_{\rm b}(\boldsymbol{x})\,, \tag{64}$$

$$H^{ij}_{\rm 2cd}(\boldsymbol{x}) = \epsilon^{ikl}\nabla^j\nabla^k \Phi_{\rm c}(\boldsymbol{x}) \nabla^l \Phi_{\rm d}(\boldsymbol{x})\,. \tag{65}$$

The POP spectrum formed by contracting the indices of these two fields $\langle H^{ij}_{1ab}(\boldsymbol{k}) H^{ij}_{2cd}(\boldsymbol{k}')\rangle$ has a similar form to the vector POP spectrum in equation (15), the difference being that the integrand would now contain an additional factor of $\boldsymbol{q}_1 \cdot \boldsymbol{q}_3$. Therefore, this POP tensor spectrum probes a different form of angular dependence in the parity-odd trispectrum.

We can also define higher-rank tensor POP spectra by contracting $\langle H^{li}_{1ab}(\boldsymbol{k}) H^{lj}_{2cd}(\boldsymbol{k}')\rangle$. We can then project this onto the basis tensors

$$e^{ij}_{\mathrm{R/L}}(\boldsymbol{k}) = e^{i}_{\mathrm{R/L}}(\boldsymbol{k}) e^{j}_{\mathrm{R/L}}(\boldsymbol{k})\,, \tag{66}$$

where $e^{i}_{\mathrm{R/L}}(\boldsymbol{k})$ are components of the polarization vectors, defined in equation (56). This decomposition gives the right-handed and left-handed components of the transverse tensor spectra, and the difference between these is a POP spectrum. This kind of construction can be easily generalized to higher-derivative and higher-rank tensor POP spectra. Some specific examples of these constructions can be found in Jeong & Kamionkowski (2012) and Shim et al. (2024). Similar constructions for the scalar POP spectra are also possible.

5.3 Higher-order constructions: parity-odd bispectra and beyond

The scalar bispectrum is defined by the three-point correlation function in Fourier space through

$$\langle \Phi(\boldsymbol{k}_1)\Phi(\boldsymbol{k}_2)\Phi(\boldsymbol{k}_3)\rangle = (2\pi)^3 \delta^{(3)}_{\mathrm{D}}(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3) B(k_1, k_2, k_3)\,. \tag{67}$$

The Dirac delta function enforces statistical translational invariance and makes the bispectrum a function over 2D, closed triangles. Since in three dimensions, we can rotate a parity-transformed 2D triangle back to its initial configuration, the bispectrum is insensitive to parity under statistical isotropy.

We construct the parity-odd bispectrum that compresses the parity-odd four-point function in a similar way to the vector POP spectra,

$$\langle \boldsymbol{A}_{ab}(\boldsymbol{k}_1) \cdot i\boldsymbol{k}_2 \Phi(\boldsymbol{k}_2)\Phi(\boldsymbol{k}_3)\rangle = (2\pi)^3 \delta^{(3)}_{\mathrm{D}}(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3) B_{\mathrm{vector}}(k_1, k_2, k_3)\,. \tag{68}$$

In this case, we have used the gradient of the scalar rather than the composite vector field. With the composite vector field, this construction would involve five scalar fields corresponding to a compression of the parity-odd five-point function. Similarly, the pseudoscalar requires a product of three scalar fields for its constructions, so the parity-odd bispectrum of one composite pseudoscalar with two scalar fields compresses the parity-odd five-point statistics. Additionally, we can correlate the vector-pseudovector inner product with two pseudoscalar fields or correlate three pseudoscalar fields. These constructions are of order six in the underlying scalar field, compressing the parity-odd six-point function. Thus, equation (68) is the only form for a parity-odd compression of the trispectrum to a bispectrum.

6 CONCLUSION

Exploring parity violation on large cosmological scales will provide crucial insights into the fundamental symmetries governing the origins and evolution of the Universe. The search for a parity-odd signal within galaxy surveys presents a key challenge in observational cosmology, especially in light of recent hints of a parity-odd signal in the four-point correlation function of galaxies (Philcox 2022; Hou et al. 2023b), both of which used the isotropic basis functions of Cahn & Slepian (2023), the method proposed in Cahn et al. (2023), and the covariance template of Hou et al. (2022). Direct measurements of four-point correlations pose several technical challenges related to the high dimensionality of the data vector and the resulting size of the covariance matrix, making it difficult to interpret the signal and assess its statistical significance. Motivated by these challenges, we developed a novel set of statistics for parity violation.

The new observables are two-point correlators calculated between composite fields derived via non-linear transformations of the original scalar field. These constructions compress the trispectrum information into a power spectrum, resulting in a significant simplification compared to the full four-point correlation function. The advantages of this approach are two-fold. First, the computation of a two-point function is faster because of the significant dimensionality reduction of the data vector; our computation on a 512^3 grid takes fewer than 10 s for each spectrum on 128 cores. This reduction facilitates a more efficient and interpretable analysis of the underlying parity-odd signal and its covariance. Secondly, formulating the estimators in Fourier space enables efficient characterization of the scale dependence of the signal and its soft limits.

We defined two sets of estimators: one based on correlating a vector and a pseudovector field and another correlating a scalar and a pseudoscalar field, both derived from the original scalar field. These two estimators are distinct compressions of the trispectrum and thus carry complementary information on the parity-odd signal. In particular, they are sensitive to different soft limits of the trispectrum. To empirically validate the effectiveness of these estimators, we tested them on a set of mock simulations with an injected parity-odd four-point correlation function. On these mocks, we compared the result of the estimator with semi-analytical theoretical calculations. The agreement between the data and our theoretical calculations demonstrates that our estimators robustly capture the injected parity-odd signal.

Beyond validation, we investigated the sensitivity of our new estimators by computing the correlation matrix on Gaussian realizations. We found that different bins of our estimator are predominantly uncorrelated. Furthermore, a χ^2-squared test on simulated parity-violating mocks highlighted the substantial statistical significance of both estimators, with the scalar variant exhibiting enhanced sensitivity compared to its vector counterpart. This sensitivity is specific to the choice of the primordial parity-violating trispectrum template and filtering functions. In particular, we did not attempt to optimize the filtering functions by maximizing signal-to-noise. Such an optimization could dramatically increase the significance. We plan to explore this in future work.

Applying our estimator to realistic survey data requires several developments. Surveys observe galaxies within a specific survey mask and selection function. We must incorporate these into our estimator to remove the spurious clustering signals they would otherwise produce. Moreover, galaxies are discrete tracers, and this results in shot noise that should be characterized and subtracted from the POP spectra. We leave these developments for future work. After overcoming these challenges, we expect the POP estimators will emerge as an essential tool in the search for parity violation in current and future CMB and galaxy surveys.

ACKNOWLEDGEMENTS

We thank Toshiki Kurita, Fabian Schmidt, Azadeh Moradinezhad Dizgah, Robert Cahn, Donghui Jeong, and Ue-Li Pen for helpful discussions. AC acknowledges funding support from the Initiative Physique des Infinis (IPI), a research training program of the Idex SUPER at Sorbonne Université. JH has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement no. 101025187. This work was supported in part by the Excellence Cluster ORIGINS, which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy: Grant No. EXC-2094–390783311. This work has also received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 101007633. The Kavli IPMU is supported, in part, by World Premier International Research Center Initiative (WPI), MEXT, Japan. ZS thanks the Max Planck Institute for Astrophysics for hospitality during some of the period during which this work was performed.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author. The software used to generate and analyse the data, and the analysis results data in Figs 3, 4, and 5 are publicly available at https://github.com/dsjamieson/parity_violating_templates.

REFERENCES

Adame A. G. et al., 2024, preprint (arXiv:2404.03002)
Aghamousa A. et al., 2016, preprint (arXiv:1611.00036)
Amendola L. et al., 2018, Living Rev. Rel., 21, 2
Bartolo N., Komatsu E., Matarrese S., Riotto A., 2004, Phys. Rept., 402, 103
Cahn R. N., Slepian Z., 2023, J. Phys. A, 56, 325204
Cahn R. N., Slepian Z., Hou J., 2023, Phys. Rev. Lett., 130, 201002
Coulton W. R., Philcox O. H. E., Villaescusa-Navarro F., 2024, Phys. Rev. D, 109, 23531
Doré O. et al., 2014, preprint (arXiv:1412.4872)
Garwin R. L., Lederman L. M., Weinrich M., 1957, Phys. Rev., 105, 1415
Gualdi D., Verde L., 2022, J. Cosmol. Astropart. Phys., 9, 50
Gualdi D., Gil-Marín H. e., Verde L., 2021, J. Cosmol. Astropart. Phys., 7, 8
Hou J., Cahn R. N., Philcox O. H. E., Slepian Z., 2022, Phys. Rev. D, 106, 43515
Hou J., Moradinezhad Dizgah A., Hahn C., Massara E., 2023a, J. Cosmol. Astropart. Phys., 3, 45
Hou J., Slepian Z., Cahn R. N., 2023b, MNRAS, 522, 5701
Hu W., 2001, Phys. Rev. D, 64, 83005
Ivezić v. et al., 2019, ApJ, 873, 111
Jeong D., Kamionkowski M., 2012, Phys. Rev. Lett., 108, 251301
Lee T. D., Yang C.-N., 1956, Phys. Rev., 104, 254
Moradinezhad Dizgah A., Lee H., Schmittfull M., Dvorkin C., 2020, J. Cosmol. Astropart. Phys., 4, 11
Munshi D., Lee H., Dvorkin C., McEwen J. D., 2022, J. Cosmol. Astropart. Phys., 11, 20
Philcox O. H. E., 2022, Phys. Rev. D, 106, 63501
Philcox O. H. E., 2023, Phys. Rev. Lett., 131, 181001
Philcox O. H. E., Ereza J., 2024, preprint (arXiv:2401.09523)
Philcox O. H. E., Shiraishi M., 2024, Phys. Rev. D, 109, 63522
Sabiu C. G., Hoyle B., Kim J., Li X.-D., 2019, ApJS, 242, 29
Schmittfull M., Baldauf T., Seljak U., 2015, Phys. Rev. D, 91, 43530
Schmittfull M., Moradinezhad Dizgah A., 2021, J. Cosmol. Astropart. Phys., 3, 20
Shim J., Pen U.-L., Yu H.-R., Okumura T., 2024, preprint (arXiv:2406.06080)
Shiraishi M., 2016, Phys. Rev. D, 94, 83503
Slepian Z., Eisenstein D. J., 2015, MNRAS, 454, 4142
Slepian Z. et al., 2017, MNRAS, 468, 1070
Sunseri J., Slepian Z., Portillo S., Hou J., Kahraman S., Finkbeiner D. P., 2023, RAS Techniques Instruments, 2, 62
Takada M. et al., 2014, Publ. Astron. Soc. Jap., 66, R1
Wang Y. et al., 2022, ApJ, 928, 1
Wu C. S., Ambler E., Hayward R. W., Hoppes D. D., Hudson R. P., 1957, Phys. Rev., 105, 1413
Xu X., Padmanabhan N., Eisenstein D. J., Mehta K. T., Cuesta A. J., 2012, MNRAS, 427, 2146

APPENDIX A: THE BINNED PARITY-ODD TRISPECTRUM AND THE BINNED POP SPECTRA

Measuring the parity-odd trispectrum from data requires choosing a binning scheme for the tetrahedral configurations of wave vectors. To form a bin of nearby tetrahedra, we specify the side lengths k_1, k_2, $k_3 k_4$, and the diagonal lengths K and $\tilde{K}$. We then select the set of wave vectors satisfying

$$\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{K} = 0\,, \tag{A1}$$

$$\boldsymbol{k}_3 + \boldsymbol{k}_4 - \boldsymbol{K} = 0\,, \tag{A2}$$

$$\boldsymbol{k}_1 + \boldsymbol{k}_4 + \tilde{\boldsymbol{K}} = 0\,. \tag{A3}$$

We define the 6D binning scheme through the integral

$$\int_{\mathbf{bin}} \equiv \int_{\boldsymbol{q}_1\in k_1}\int_{\boldsymbol{q}_2\in k_2}\int_{\boldsymbol{q}_3\in k_2}\int_{\boldsymbol{q}_4\in k_4} \Theta\left(4\left(|\boldsymbol{q}_1+\boldsymbol{q}_2|-K\right)^2-\Delta k^2\right)\Theta\left(4\left(|\boldsymbol{q}_1+\boldsymbol{q}_4|-\tilde{K}\right)^2-\Delta k^2\right). \tag{A4}$$

The integral notation is defined in Table A1. The Heaviside functions enforce $K - \Delta k/2 \leq |\boldsymbol{q}_1 + \boldsymbol{q}_2| \leq K + \Delta k/2$ and $\tilde{K} - \Delta k/2 \leq |\boldsymbol{q}_1 + \boldsymbol{q}_4| \leq \tilde{K} + \Delta k/2$, which defines the binning scheme for the diagonals.

We isolate the binned parity-odd trispectrum by integrating the product of the triple product and the four-point expectation value over the bin,

$$\bar{\tau}_-(k_1,k_2,k_3,k_4,K,\tilde{K}) = \mathcal{N}_{\tau_-}^{-1}\int_{\mathbf{bin}} -i\mathbf{q}_1\cdot(\mathbf{q}_2\times\mathbf{q}_3)\left\langle\Phi(\mathbf{q}_1)\Phi(\mathbf{q}_2)\Phi(\mathbf{q}_3)\Phi(\mathbf{q}_4)\right\rangle \tag{A5}$$

$$= \mathcal{N}_{\tau_-}^{-1}\int_{\mathbf{bin}} (2\pi)^3\delta_{\mathrm{D}}^{(3)}(\mathbf{q}_1+\mathbf{q}_2+\mathbf{q}_3+\mathbf{q}_4)\left[\mathbf{q}_1\cdot(\mathbf{q}_2\times\mathbf{q}_3)\right]^2\tau_-\left(q_1,q_2,q_3,q_4,|\mathbf{q}_1+\mathbf{q}_2|,|\mathbf{q}_1+\mathbf{q}_4|\right). \tag{A6}$$

Here, $\mathcal{N}_{\tau_-}$ is a normalization factor, which we will define later. The minus sign in the first line is for convenience, and we could absorb it into the normalization factor. We split the Dirac delta function as follows:

$$(2\pi)^3\delta_{\mathrm{D}}^{(3)}(\mathbf{q}_1+\mathbf{q}_2+\mathbf{q}_3+\mathbf{q}_4) = (2\pi)^9\int_{\mathbf{Q},\tilde{\mathbf{Q}}}\delta_{\mathrm{D}}^{(3)}(\mathbf{q}_1+\mathbf{q}_2+\mathbf{Q})\,\delta_{\mathrm{D}}^{(3)}(\mathbf{q}_3+\mathbf{q}_4-\mathbf{Q})\,\delta_{\mathrm{D}}^{(3)}\left(\mathbf{q}_1+\mathbf{q}_4+\tilde{\mathbf{Q}}\right). \tag{A7}$$

The right-hand side of this expression imposes the condition that the tetrahedron consists of two triangles, $\{\mathbf{q}_1, \mathbf{q}_2, -\mathbf{Q}\}$ and $\{\mathbf{q}_3, \mathbf{q}_4, \mathbf{Q}\}$, which are joined along their common side of length Q. The vectors $\mathbf{q}_1$ and $\mathbf{q}_4$ are additionally constrained to form a triangle $\{\mathbf{q}_1, \mathbf{q}_4, -\tilde{\mathbf{Q}}\}$. After substituting equation (A7) into equation (A6), the Heaviside functions limit the integrations over Q and $\tilde{Q}$. Then, with the integral notation in Table A1, the binned parity-odd trispectrum is given by

$$\bar{\tau}_-(k_1,k_2,k_3,k_4,K,\tilde{K}) = \mathcal{N}_{\tau_-}^{-1}\int_{6D\text{-}\mathbf{bin}}\Big[(2\pi)^9\delta_{\mathrm{D}}^{(3)}\left(\boldsymbol{q}_1+\boldsymbol{q}_2+\boldsymbol{Q}\right)\delta_{\mathrm{D}}^{(3)}\left(\boldsymbol{q}_3+\boldsymbol{q}_4-\boldsymbol{Q}\right)\delta_{\mathrm{D}}^{(3)}\left(\boldsymbol{q}_1+\boldsymbol{q}_4+\tilde{\boldsymbol{Q}}\right)$$
$$\times\ \left[\boldsymbol{q}_1\cdot(\boldsymbol{q}_2\times\boldsymbol{q}_3)\right]^2\tau_-(q_1,q_2,q_3,q_4,Q,\tilde{Q})\Big]. \tag{A8}$$

Table A1. Fourier-space integral and harmonic-space summation notation.

Notation	Definition
$\int_{q\in k}$	$\int_{k-\Delta k/2}^{k+\Delta k/2}\frac{k^2\mathrm{d}k}{2\pi^2}$
$\int_{\hat{q}}$	$\int\frac{\mathrm{d}\Omega_{\hat{q}}}{4\pi}$
$\int_{\boldsymbol{q}\in k}$	$\int_{\hat{q}}\int_{q\in k}$
$\int_{6D\text{-}\mathbf{bin}}$	$\int_{\boldsymbol{q}_1\in k_1}\int_{\boldsymbol{q}_2\in k_2}\int_{\boldsymbol{q}_3\in k_2}\int_{\boldsymbol{q}_4\in k_4}\int_{\boldsymbol{Q}\in K}\int_{\tilde{\boldsymbol{Q}}\in\tilde{K}}$
$\int_{5D\text{-}\mathbf{bin}}$	$\int_{\boldsymbol{q}_1\in k_1}\int_{\boldsymbol{q}_2\in k_2}\int_{\boldsymbol{q}_3\in k_3}\int_{\boldsymbol{q}_4\in k_4}\int_{\boldsymbol{Q}\in K}$
$\int_{5D\text{-bin}}$	$\int_{q_1\in k_1}\int_{q_2\in k_2}\int_{q_3\in k_3}\int_{q_4\in k_4}\int_{Q\in K}$
$\sum_{L_i M_i}$	$\sum_{L_1=0}^{\infty}\sum_{M_1=-L_1}^{L_1}\sum_{L_2=0}^{\infty}\sum_{M_2=-L_2}^{L_2}\sum_{L_3=0}^{\infty}\sum_{M_3=-L_3}^{L_3}$

The notation $q \in k$ indicates a wave vector magnitude belonging to a bin centered on k with width Δk . The integral over $\mathrm{d}\Omega_{\hat{q}}$ is the solid angle integral over the azimuthal and polar angles associated with the wave vector $\boldsymbol{q}$ in polar coordinates.

Unfortunately, the full 6D trispectrum is difficult to analyse directly. The first two Dirac delta functions factor into integrals over two triangles, but the third couples these together, making the whole integral not factorizable. For this reason, we extend the $\tilde{Q}$ bin to be unrestricted, $\tilde{Q} \in [0, \infty)$, which fully averages over the second diagonal. An alternative approach is to implement isotropic basis functions, which discretize the angular dependence through spherical harmonic expansions (Cahn & Slepian 2023), but we will not pursue that here.

The 5D binned trispectrum, with notation from Table A1, is

$$\bar{\tau}_-(k_1, k_2, k_3, k_4, K) = \mathcal{N}_{\tau_-}^{-1} \int_{5D\text{-bin}} (2\pi)^6 \delta_{\rm D}^{(3)}\left(\boldsymbol{q}_1 + \boldsymbol{q}_2 + \boldsymbol{Q}\right) \delta_{\rm D}^{(3)}\left(\boldsymbol{q}_3 + \boldsymbol{q}_4 - \boldsymbol{Q}\right) \left[\boldsymbol{q}_1 \cdot (\boldsymbol{q}_2 \times \boldsymbol{q}_3)\right]^2 \tau_-\left(q_1, q_2, q_3, q_4, Q, |\boldsymbol{q}_1 + \boldsymbol{q}_4|\right), \tag{A9}$$

The third Dirac delta function in equation (A8) enforced $\tilde{Q} = |\boldsymbol{q}_1 + \boldsymbol{q}_4|$. We choose the somewhat arbitrary normalization factor $\mathcal{N}_{\tau_-}$ to be

$$\mathcal{N}_{\tau_-} = \int_{5D\text{-bin}} (2\pi)^9 \delta_{\rm D}^{(3)}\left(\boldsymbol{q}_1 + \boldsymbol{q}_2 + \boldsymbol{Q}\right) \delta_{\rm D}^{(3)}\left(\boldsymbol{q}_3 + \boldsymbol{q}_4 - \boldsymbol{Q}\right) \left[\boldsymbol{q}_1 \cdot (\boldsymbol{q}_2 \times \boldsymbol{q}_3)\right]^2, \tag{A10}$$

so equation (A9) is the weighted average of $\tau_-(q_1, q_2, q_3, q_4, Q, |\mathbf{q}_1 + \mathbf{q}_4|)$ within the bin.

Expressing both Dirac delta functions as the Fourier transform of plane waves enables us to compute most angular integrals analytically. The calculation simplifies greatly by noticing that we can rewrite the triple product as $\boldsymbol{q}_1 \cdot (\boldsymbol{q}_2 \times \boldsymbol{q}_3) = -\boldsymbol{Q} \cdot (\boldsymbol{q}_1 \times \boldsymbol{q}_3)$, which is more symmetric between the two triangles forming the tetrahedron. However, if the trispectrum depends on the diagonals, the angular dependence on $\boldsymbol{q}_1 \cdot \boldsymbol{q}_4$ may prevent us from analytically computing all angular integrals. From here on, we will restrict to the case where the trispectrum has no explicit dependence on the diagonals.

The spherical harmonic expansion of the squared triple product is

$$[\mathbf{Q} \cdot (\mathbf{q}_1 \times \mathbf{q}_3)]^2 = -6(4\pi)^3 (Q q_1 q_3)^2 \sum_{L_i, M_i} D^{M_1 M_2 M_3}_{L_1 L_2 L_3} Y_{L_1 M_1}(\hat{q}_1) Y_{L_2 M_2}(\hat{q}_3) Y_{L_3 M_3}(\hat{Q}), \tag{A11}$$

where the coefficients are

$$D^{M_1 M_2 M_3}_{L_1 L_2 L_3} = \begin{pmatrix} L_1 & L_2 & L_3 \\ M_1 & M_2 & M_3 \end{pmatrix} \begin{Bmatrix} L_1 & L_2 & L_3 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{Bmatrix} \prod_{i=1}^{3} \sqrt{\frac{2L_i + 1}{4\pi}} \begin{pmatrix} 1 & 1 & L_i \\ 0 & 0 & 0 \end{pmatrix}. \tag{A12}$$

Here, the 2 by 3 matrices are Wigner 3-j symbols, and the 3 by 3 matrix is a Wigner 9-j symbol. The form of the coefficients in the product restricts $L_i \in \{0, 2\}$.

The trispectrum estimator is an integral over tetrahedra formed by two triangles: $\{\mathbf{q}_1, \mathbf{q}_2, -\mathbf{Q}\}$ and $\{\mathbf{q}_3, \mathbf{q}_4, \mathbf{Q}\}$. By taking one vector from each triangle and the common vector shared by both triangles in the triple product, we preserve the symmetry between the two pairs $(\mathbf{q}_1, \mathbf{q}_2)$ and $(\mathbf{q}_3, \mathbf{q}_4)$, greatly simplifying the calculation. Each triangle contributes a Dirac delta function and a Y_{LM} from a vector that is not $\pm\mathbf{Q}$. Expressing the Dirac delta functions as Fourier transforms of plane waves allows us to compute all angular integrals,

$$\int_{\mathbf{q}_1 \in k_1, \mathbf{q}_2 \in k_2} (2\pi)^3 \delta_{\rm D}^{(3)}(\mathbf{q}_1 + \mathbf{q}_2 - \mathbf{Q}) Y_{L_1 M_1}(\hat{q}_1) = Y_{L_1 M_1}(\hat{Q}) \int_{q_1 \in k_1, q_2 \in k_2} 4\pi \int_0^\infty \mathrm{d}x\, x^2\, j_{L_1}(q_1 x) j_0(q_2 x) j_{L_1}(Q x) \tag{A13}$$

$$= Y_{L_1 M_1}(\hat{Q}) \int_{q_1 \in k_1, q_2 \in k_2} I_{L_1 0 L_1}(q_1, q_2, Q), \tag{A14}$$

and similarly,

$$\int_{\mathbf{q}_3 \in k_3, \mathbf{q}_4 \in k_4} (2\pi)^3 \delta_{\rm D}^{(3)}(\mathbf{q}_3 + \mathbf{q}_4 + \mathbf{Q}) Y_{L_2 M_2}(\hat{q}_3) = (-1)^{L_2} Y_{L_2 M_2}(\hat{Q}) \int_{q_3 \in k_3, q_4 \in k_4} 4\pi \int_0^\infty \mathrm{d}x\, x^2\, j_{L_2}(q_3 x) j_0(q_4 x) j_{L_2}(Q x) \tag{A15}$$

$$= (-1)^{L_2} Y_{L_2 M_2}(\hat{Q}) \int_{q_3 \in k_3, q_4 \in k_4} I_{L_2 0 L_2}(q_3, q_4, Q). \tag{A16}$$

Since $L_2 \in \{0, 2\}$, we have $(-1)^{L_2} = 1$, so we can drop this factor. Both triangles contribute factors to the integrand that have the same form, illustrating the symmetric roles of the two triangles that form the tetrahedron. The triple-spherical Bessel integrals are given by

$$I_{L_1 0 L_1}(q_1, q_2, Q) \equiv 4\pi \int_0^\infty \mathrm{d}x\, x^2\, j_{L_1}(q_1 x) j_0(q_2 x) j_{L_1}(Q x) \tag{A17}$$

$$= \frac{\pi^2}{q_1 q_2 Q} \Theta\left(\sin^2\theta_{12}\right) P_{L_1}\left(\cos\theta_{12}\right). \tag{A18}$$

Here, θ_{12} is the angle between $\mathbf{q}_1$ and $\mathbf{Q}$. As discussed in the main text, the Heaviside function enforces the triangle inequalities that ensure $\mathbf{q}_1$, $\mathbf{q}_2$, and $-\mathbf{Q}$ can form a closed triangle. P_{L_1} is the $L_1^{\rm th}$ Legendre polynomial. The only relevant ones will be

$$P_0(x) = 1, \quad P_2(x) = \frac{1}{2}\left(3x^2 - 1\right). \tag{A19}$$

At this point, three $Y_{L_i M_i}$ products remain, all with the argument $\hat{Q}$. Integrating over $\hat{Q}$ gives Gaunt's integral (divided by 4π),[3]

$$G^{M_1 M_2 M_3}_{L_1 L_2 L_3} = \int_{\hat{Q}} Y_{L_1 M_1}(\hat{Q}) Y_{L_2 M_2}(\hat{Q}) Y_{L_3 M_3}(\hat{Q}) \tag{A20}$$

[3] https://dlmf.nist.gov/34.3, equation (34.3.22).

$$= \begin{pmatrix} L_1 L_2 L_3 \\ M_1 M_2 M_3 \end{pmatrix} \begin{pmatrix} L_1 L_2 L_3 \\ 000 \end{pmatrix} \prod_{i=1}^{3} \sqrt{\frac{2L_i + 1}{4\pi}} \,. \tag{A21}$$

The form of the trispectrum after computing all angular integrals is

$$\bar{\tau}_-(k_1, k_2, k_3, k_4, K) = -6\mathcal{N}_{\tau_-}^{-1} \int_{5D\text{-bin}} (Qq_1q_3)^2 \sum_{L_i M_i} D^{M_1M_2M_3}_{L_1L_2L_3} G^{M_1M_2M_3}_{L_1L_2L_3} I_{L_1 0 L_1}(q_1, q_2, Q) I_{L_2 0 L_2}(q_3, q_4, Q)\, \tau_-(q_1, q_2, q_3, q_4)\,, \tag{A22}$$

where the summation notation is defined in Table A1. Using the orthogonality relation for the 3-j symbol[4]

$$\sum_{M_1=-L_1}^{L_1} \sum_{M_2=-L_2}^{L_2} \sum_{M_3=-L_3}^{L_3} \begin{pmatrix} L_1 L_2 L_3 \\ M_1 M_2 M_3 \end{pmatrix}^2 = 1\,, \tag{A23}$$

we can perform the sums over all M_i. Since neither triple-spherical Bessel integral involves L_3, we can also sum over it, which gives

$$\bar{\tau}_-(k_1, k_2, k_3, k_4, K) = 6\mathcal{N}_{\tau_-}^{-1} \int_{5D\text{-bin}} (Qq_1q_3)^2 \sum_{L_1, L_2} H_{L_1L_2} I_{L_1 0 L_1}(q_1, q_2, Q) I_{L_2 0 L_2}(q_3, q_4, Q) \tau_-(q_1, q_2, q_3, q_4)\,, \tag{A24}$$

where the coefficients now simplify considerably,

$$H_{L_1L_2} = -\sum_{L_3} \begin{pmatrix} L_1 & L_2 & L_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{Bmatrix} L_1 & L_2 & L_3 \\ 1 & 1 & 1 \end{Bmatrix} \prod_{i=1}^{3} (2L_i + 1) \begin{pmatrix} 1 & 1 & L_i \\ 0 & 0 & 0 \end{pmatrix} \tag{A25}$$

$$= \frac{i^{L_1+L_2}}{27}\,. \tag{A26}$$

The two remaining sums factorize,

$$\bar{\tau}_-(k_1, k_2, k_3, k_4, K) = \frac{2}{9}\mathcal{N}_{\tau_-}^{-1} \int_{5D\text{-bin}} (Qq_1q_3)^2 \Big[\sum_{L_1} i^{L_1} I_{L_1 0 L_1}(q_1, q_2, Q)\Big] \Big[\sum_{L_2} i^{L_2} I_{L_2 0 L_2}(q_3, q_4, Q)\Big] \tau_-(q_1, q_2, q_3, q_4)\,. \tag{A27}$$

Each sum only has two terms, so we can sum them explicitly and expand the Legendre polynomials,

$$\bar{\tau}_-(k_1, k_2, k_3, k_4, K) = \frac{\pi^4}{2}\mathcal{N}_{\tau_-}^{-1} \int_{5D\text{-bin}} \frac{q_1q_3}{q_2q_4}\, \Theta\big(\sin^2\theta_{12}\big) \sin^2\theta_{12}\, \Theta\big(\sin^2\theta_{34}\big) \sin^2\theta_{34}\, \tau_-(q_1, q_2, q_3, q_4)\,, \tag{A28}$$

where we set $\tau_- = 1$ to obtain the normalization factor,

$$\mathcal{N}_{\tau_-} = \frac{\pi^4}{2} \int_{5D\text{-bin}} \frac{q_1q_3}{q_2q_4} \Theta\big(\sin^2\theta_{12}\big) \sin^2\theta_{12} \Theta\big(\sin^2\theta_{34}\big) \sin^2\theta_{34}\,. \tag{A29}$$

Instead of restricting the integration bounds, we could define filter functions with Heaviside functions that impose the binning. Then, the binned parity-odd trispectrum takes the form,

$$\bar{\tau}_-(k_1, k_2, k_3, k_4, K) = \frac{\pi^4}{2}\mathcal{N}_{\tau_-}^{-1} \int_{q_1, q_2, q_3, q_4, Q} \Big[\frac{q_1q_3}{q_2q_4}\, \Theta\big(\sin^2\theta_{12}\big) \sin^2\theta_{12}\, \Theta\big(\sin^2\theta_{34}\big) \sin^2\theta_{34}$$
$$\times\; f_a(q_1) f_b(q_2) f_c(q_3) f_d(q_4) f_e(Q)\, \tau_-(q_1, q_2, q_3, q_4)\Big]\,. \tag{A30}$$

By taking all but f_e to be generic filter functions, rather than just Heaviside functions selecting spherical shells, and taking the thin-bin limit of f_e, this has the same form as the vector POP spectrum in equation (17). Similarly, by taking all but f_a to be generic filter functions and taking the thin-bin limit of f_a, this has the same form as the scalar POP spectrum in equation (27). Thus, computing a POP spectrum is equivalent to computing a 1D subset of bins for the full 5D trispectrum.

[4] https://dlmf.nist.gov/34.3 equation (34.3.18).

This paper has been typeset from a TEX/LATEX file prepared by the author.