Peak-background split, renormalization, and galaxy clustering

Fabian Schmidt, Donghui Jeong, and Vincent Desjacques

Phys. Rev. D **88**, 023515 — Published 12 July 2013

DOI: 10.1103/PhysRevD.88.023515

Peak-Background Split, Renormalization, and Galaxy Clustering

Fabian Schmidt,[1,2] Donghui Jeong,[3] and Vincent Desjacques[4]

[1] *Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA*
[2] *Einstein fellow*
[3] *Department of Physics and Astronomy, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21210, USA*
[4] *Département de Physique Théorique and Center for Astroparticle Physics (CAP) Université de Genève, 24 quai Ernest Ansermet, CH-1211 Genève, Switzerland*

We present a derivation of two-point correlations of general tracers in the peak-background split (PBS) framework by way of a rigorous definition of the PBS argument. Our expressions only depend on connected matter correlators and "renormalized" bias parameters with clear physical interpretation, and are independent of any coarse-graining scale. This result should be contrasted with the naive expression derived from a local bias expansion of the tracer number density with respect to the matter density perturbation δ_L coarse-grained on a scale R_L. In the latter case, the predicted tracer correlation function receives contributions of order $\langle\delta_L^n\rangle$ at each perturbative order n, whereas, in our formalism, these are absorbed in the PBS bias parameters at all orders. Further, this approach naturally predicts both a scale-dependent bias $\propto k^2$ such as found for peaks of the density field, and the scale-dependent bias induced by primordial non-Gaussianity in the initial conditions. The only assumption made about the tracers is that their abundance at a given position depends solely on the matter distribution within a finite region around that position.

PACS numbers: 98.80.-k, 98.65.-r,98.62.Py, 95.35.+d
Keywords: cosmology; large-scale structure; non-Gaussianity

I. INTRODUCTION

The clustering of tracers of the large-scale structure (LSS) in the Universe, such as galaxies, clusters, or the Lyman-α forest, is one of the most important probes of the origin and evolution of cosmological perturbations. On sufficiently large scales, correlations are weak and one should hope that a perturbative approach will allow us to cleanly connect observations with the predictions of cosmological models (such as the standard cold dark matter scenario with adiabatic Gaussian initial conditions). In case of the matter density perturbations, these perturbative approaches are well developed (see [1] for a review). On the other hand, the formation of the tracers we actually observe necessarily involves highly nonlinear, small-scale mechanisms which cannot be described perturbatively. In order to relate observations to theory, we thus need an effective description which involves unknown "bias parameters". These parameters in general need to be determined observationally, and we would like as many parameters as necessary to accurately describe tracer correlations down to some minimum scale, but not more, in order to retain the maximum amount of cosmological information. Which bias parameters need to be included, and to what order, is still an open problem.

The simplest and most well-known bias expansion is the local expansion in terms of density [2–4],

$$\delta_{h,L}(\mathbf{x}) = c_0 + c_1\delta_L(\mathbf{x}) + \frac{c_2}{2}\delta_L^2(\mathbf{x}) + \ldots, \qquad (1)$$

where c_n are the bias parameters, $\delta_{h,L}$ is the fractional tracer density perturbation, and δ_L is the corresponding matter density perturbation. Here, both the tracer and the matter density field are understood as coarse-grained on some scale R_L, so that this expansion can be interpreted as a counts-in-cells relation, and the c_n as "scatter-plot bias parameters". This perturbative description is commonly assumed to be valid on large scales [5, 6], i.e. only if $\sigma_L^2 = \langle\delta_L^2\rangle \ll 1$, although the series could actually converge even for $\sigma_L \gtrsim 1$ if the condition $|c_{n+1}/c_n| \approx$ const. is satisfied in the limit $n \to \infty$; however, the larger σ_L, the more terms need to be included to obtain a converged expression. As can easily be seen, the correlation function predicted by Eq. (1) depends on the coarse-graining scale through the variance $\sigma_L^2 \equiv \langle\delta_L^2\rangle$, and all higher moments of the density field.

In contrast to counts-in-cells studies [2], where a specific scale R_L is singled out, we expect that no additional smoothing scale R_L should enter the calculation of correlation functions on a scale r (unless we directly link R_L to r, as done in [7]). Thus, we need to absorb the R_L-dependent pieces into renormalized bias parameters b_N, as proposed for the first time by McDonald [8] and tested against simulations in Jeong and Komatsu [9]. More recently, this approach has been pursued to higher order in the multipoint propagator framework [10] by Matsubara [11]. In accordance with renormalization theory, the expression for tracer correlators in terms of the parameters b_N should be R_L-independent, and convergent in terms of matter correlators at the separation at which we measure tracer correlations (rather than zero-lag correlators such as $\langle\delta_L^2\rangle$). In the process of renormalization, we have to introduce a new bias parameter b_N at each order in δ_L which needs to be determined from observations. Similar conclusions hold when adding other, non-local quantities to the expansion Eq. (1), such as derivatives of δ

or the tidal tensor. With in principle arbitrarily many free parameters, the clustering of LSS tracers loses all cosmological constraining power. Moreover, treating the b_N as mere nuisance parameters (as in [8, 12]) precludes us from using any information contained in the b_N on the parent halos and formation history of the tracers.

There is thus strong motivation to try to associate physical meaning with the renormalized bias parameters. This would allow models of, say, galaxy formation to provide sensible priors on the allowed range of the parameters, or to predict connections between different bias parameters. For example, there is a well-motivated connection between the linear bias with respect to density of a tracer, and the parameter that quantifies the scale-dependent bias induced by primordial non-Gaussianity [13, 14]. In this case, the connection is crucial, as without it one would not be able to constrain primordial non-Gaussianity from the scale-dependent bias.

Our goal in this paper is to explicitly derive the physical meaning of the renormalized bias parameters b_N, and their exact relation to observables such as the correlation function to any order. Physically, the renormalized bias parameters quantify the response of the mean abundance of tracers to a change in the background matter density $\overline{\rho}$ of the Universe (at fixed proper time since the Big Bang), i.e.

$$b_N = \frac{\overline{\rho}^N}{\bar{n}} \frac{\partial^N \bar{n}}{\partial \overline{\rho}^N} \,. \tag{2}$$

This definition applies to tracers of any nature; in the considerably more restrictive case of universal mass functions, where the abundance of tracers depends only on $\nu_c = \delta_c/\sigma(M)$, a fractional change D in the background density is equivalent to a change in the parameter $\delta_c \to \delta_c - D$, leading to the well-known peak-background split (PBS) bias parameters [15], Eq. (49) in Sec. II C. Even though Eq. (2) is much more general, we will still refer to the parameters b_N as "PBS biases" for convenience, although it is important to keep the distinction in mind.

We will show that the bias parameters defined through Eq. (2), and its generalization to non-local biases, are the coefficients multiplying powers of the matter correlation function $\xi_L(r)$ in the expansion of the tracer correlation function. Specficially, for a Gaussian density field we obtain

$$\xi_h(r) = \sum_{N=1}^{\infty} \frac{b_N^2}{N!} \left[\xi_L(r)\right]^N \,. \tag{3}$$

We also show that the b_N agree with the bias parameters identified through a direct calculation of the correlations of thresholded regions and peaks of the density field. However, it is important to note the difference in philosophy between each derivation: thresholded regions and peaks constitute "microscopic" models of tracers, where the relation between matter and tracer density is explicitly specified on all scales (this is also the approach taken in Matsubara [11], although the description is kept general there). On the other hand, here we treat local biasing (and its generalization to non-local quantities) as effective description on sufficiently large scales, independently of the microscopic physics. In the language of field theory, the former approaches constitute specific "UV-complete" theories, while the approach presented here is an "effective theory" of biasing. Specifically, R_L serves the role of a UV cutoff whose precise value should not impact correlations on scales of observable interest.

The exact relation between the parameters b_N and tracer correlations provides a rigorous framework in which further modeling assumptions for any given tracer, for example from the excursion set, peak model, or halo occupation distribution, can be embedded — both in order to tighten constraints on cosmological parameters, and in order to infer the physics of the formation of the tracers (as also pointed out in [16], the deviations from the peak-background split predictions found in [17] are due to the fact that the authors assumed a universal mass function of halos, not due to the inaccuracy of the peak-background split argument itself). Along the way, we also show that renormalization removes the zero-lag matter correlators from the expression for the tracer correlation function at all orders, as required. We also show that the same bias parameters describe both the tracer auto- and the cross-correlation with matter. The expression of tracer correlations in terms of renormalized bias parameters and "no-zero-lag" matter correlators such as the correlation function is manifestly convergent as long as these matter correlators are small. This is in close analogy to the expression for tracer correlations in terms of resummed bias parameters in [11]. Note that the treatment in [11] is in Fourier space, while we work in real space here, for which we believe that the physical assumptions and arguments are more clear.

The renormalization approach proposed in [8] provides us with another extremely valuable tool: in describing the tracer density in terms of the "bare" bias parameters c_n, we coarse-grain both tracer and matter fields on some scale R_L. The requirement that the final expression for observable tracer correlations be independent of R_L provides us with a quantitative estimate of the limits of the ansatz Eq. (1). Once an R_L-dependence is found, we are guided to find an additional dependence of the tracer density on a "regional" property of the matter density field which absorbs the dependence on R_L through renormalization. We show this explicitly in two cases: a bias with respect to $\nabla^2\delta$, previously found specifically for peaks of the density field, which absorbs the R_L-dependence induced by the smoothing of the matter correlation function over the coarse-graining scale; and a bias with respect to the variance of the small-scale density field, which has to be introduced in the presence of primordial non-Gaussianity. In the latter case, the renormalization absorbs the $c_1 c_2 \langle \delta_L(1) \delta_L^2(2) \rangle$ term into an overall scale-dependent bias coefficient, similarly as shown in [18]. However, in our approach we find a clear

physical interpretation of this coefficient, which preserves the connection to the Gaussian bias derived in [13, 14].

In order to achieve these results efficiently, we adopt several simplifications. First, we work in real space rather than Fourier space, since this is where the coarse-graining and "separation of scales" is defined. We thus do not address effects such as stochasticity and exclusion, which are restricted to very small physical separations (while they in general affect Fourier-space correlations at all k). Furthermore, our approach is primarily intended as being applied in Lagrangian space. However, we will not restrict the treatment to a linear or Gaussian density field. Thus, our results are applicable to biasing with respect to a nonlinearly evolved matter density field as well, though, in that case, further non-local biases should in general be considered [19–21]. We also assume that the tracer abundance solely depends on the total matter distribution, rather than the baryon and cold dark matter density separately; the approach can easily be generalized to deal with the two-component fluid. Finally, we restrict ourselves to two-point correlations of tracers.

Implicitly, we will work in synchronous-comoving gauge throughout. That is, all comoving observers on a constant-t slice share the same proper time (at linear order), and thus see a Universe of equal age. For a different choice of time slicing, tracer correlations in general receive contributions from the different evolutionary stage of different regions as well [22] (see also [23–26]).

The outline of the paper is as follows. In Sec. II, we describe the basic approach assuming purely local biasing as in Eq. (1) (although we never actually use this relation) and Gaussian initial conditions, and derive the expression for tracer correlations in terms of PBS bias parameters. In Sec. III, we show how a bias with respect to $\nabla^2\delta$ naturally appears in this approach. Finally, in Sec. IV we consider the case of non-Gaussian initial conditions. We conclude in Sec. V. The appendix contains the derivation of the key result of Sec. II for general non-Gaussian initial conditions, the extension of the treatment in Sec. III to higher order, and detailed derivations of some relations used in the main text.

II. PEAK-BACKGROUND SPLIT AND TRACER CORRELATIONS

Consider a filter function $W_L(\mathbf{x})$ of characteristic size R_L, normalized to unity in 3D space. In the following we assume that the filter function is isotropic, $W_L = W_L(|\mathbf{x}|)$. This is a natural assumption since any anisotropy would correspond to introducing preferred directions. We define the filtered (coarse-grained) density field δ_L in terms of the full density field $\delta(\mathbf{x})$ through

$$\delta_L(\mathbf{x}) \equiv \int d^3\mathbf{y}\, W_L(\mathbf{x}-\mathbf{y})\delta(\mathbf{y})\,, \tag{4}$$

where the subscript L refers to the coarse-graining scale R_L (indicated as circles in Fig. 1). We can think of $\delta_L(\mathbf{x})$ as the average density within a region $\mathcal{U}$ of size R_L centered on $\mathbf{x}$. Note that while we primarily think of $\delta(\mathbf{x})$ as being the Lagrangian density field, many of our results will not make any assumptions about the statistics of $\delta(\mathbf{x})$ (e.g., Gaussianity). The small-scale density field, which we will consider in Sec. IV, is then defined as $\delta_s(\mathbf{x}) = \delta(\mathbf{x}) - \delta_L(\mathbf{x})$ [Eq. (86)]. The number of tracers (orange dots in Fig. 1) within this region $\mathcal{U}$ is simply given by the weighted sum, i.e., the discretized analog of Eq. (4),

$$\hat{n}_h(\mathbf{x}) = \sum_i W_L(\mathbf{x}_i - \mathbf{x}), \tag{5}$$

where the sum runs over all tracers in the (idealized) sample, and $\mathbf{x}_i$ is the position of tracer i. The expectation value $\langle\hat{n}_h\rangle$, estimated by averaging over N different regions $\mathcal{U}$ and letting $N\to\infty$, is equal to the cosmic mean of the abundance of tracers. It can be measured (with some uncertainty) either observationally or in simulations for any given tracer.

We can now implicitly define a function $F_{h,L}(\delta_L;\mathbf{x})$ through

$$\hat{n}_h(\mathbf{x}) = F_{h,L}(\delta_L(\mathbf{x});\mathbf{x}). \tag{6}$$

The dependence of $F_{h,L}$ on $\mathbf{x}$ denotes any departure, or "scatter", of the tracer number density from a deterministic relation $\hat{n}_h(\mathbf{x}) = \hat{n}_h[\delta_L(\mathbf{x})]$; by definition, this scatter is equivalent to the dependence of $\hat{n}_h$ on the small-scale fluctuations δ_s in the given region. The key assumption we will make below is that the correlation of this scatter with large-scale perturbations (in particular on the scales we are measuring correlation functions) is negligible. Then, the scatter will add noise to the measurement, but will not contribute to the expectation value of correlation functions on large scales [27]. If this assumption breaks down, it is straightforward to include other properties of the density field as arguments of $F_{h,L}$, which will be the subject of the following sections. The PBS argument we will apply below will allow us to derive the statistics of the tracer without any explicit knowledge of the function $F_{h,L}$. As indicated by the notation, the function $F_{h,L}$ will depend on R_L.

We can formally expand Eq. (6) in a Taylor series,

$$\hat{n}_h(\mathbf{x}) = \sum_{n=0}^{\infty}\frac{1}{n!}F^{(n)}_{h,L}(0;\mathbf{x})\,[\delta_L(\mathbf{x})]^n, \tag{7}$$

where $F^{(n)}_{h,L}(0;\mathbf{x})$ denotes the n-derivative of $F_{h,L}(\delta_L,\mathbf{x})$ with respect to δ_L evaluated at position $\mathbf{x}$ and at $\delta_L = 0$. We now take the expectation value of Eq. (7) in order to obtain an expression for the mean number density of tracers, our first observable. Our assumption of negligible correlation between δ_L and the scatter encoded in the explicit $\mathbf{x}$-dependence of $F_{h,L}$ implies that the two factors in each term of Eq. (7) are independent random variables (cf. the Poisson clustering model [4]):

$$\langle F^{(n)}_{h,L}(0;\mathbf{x})\,[\delta_L(\mathbf{x})]^n\rangle = \langle F^{(n)}_{h,L}(0;\mathbf{x})\rangle\,\langle[\delta_L(\mathbf{x})]^n\rangle \tag{8}$$

We then obtain

$$\langle \hat{n}_h(\mathbf{x})\rangle = \sum_n \frac{1}{n!}\left\langle F^{(n)}_{h,L}(0;\mathbf{x})\,[\delta_L(\mathbf{x})]^n\right\rangle \quad (9)$$

$$= \sum_n \frac{1}{n!}\left\langle F^{(n)}_{h,L}(0;\mathbf{x})\right\rangle \langle\delta^n_L\rangle \quad (10)$$

$$= \langle F_{h,L}(0)\rangle\left(1+\frac{c_2}{2}\sigma^2_L+\frac{c_3}{6}\langle\delta^3_L\rangle+\ldots\right), \quad (11)$$

where we have defined

$$c_n \equiv \frac{1}{\langle F_{h,L}(0)\rangle}\left\langle F^{(n)}_{h,L}(0)\right\rangle, \quad (12)$$

dropping the argument $\mathbf{x}$ since the ensemble average is independent of location due to homogeneity. The c_n depend on R_L, but for clarity we will not indicate this dependence explicitly. Further, as in Sec. I,

$$\sigma^2_L \equiv \langle\delta^2_L\rangle = \int \frac{d^3k}{(2\pi)^3}|\tilde{W}_L(k)|^2 P(k). \quad (13)$$

Note that by definition, $\langle\delta_L\rangle = 0$, and $c_0 = 1$. In the limit that $R_L \to \infty$ so that $\sigma_L \to 0$, we see that $\langle\hat{n}_h\rangle = \langle F_{h,L}(0)\rangle$, the expectation value of the function $F_{h,L}$ at the background density. For finite values of R_L however, $\langle\hat{n}_h\rangle$ receives contributions from the variance σ^2_L and higher order moments of the density field coarse-grained with W_L. This just says that for finite regions $\mathcal{U}$, $\langle F_{h,L}(0)\rangle$ does not give the cosmic mean of the tracer abundance, $\langle\hat{n}_h\rangle$. This is commonly phrased as a non-zero zeroth order bias parameter given by $\langle\hat{n}_h\rangle/\langle F_{h,L}(0)\rangle$ so that the cosmic mean is recovered upon taking the ensemble average.

A. Correlations

We now turn to the correlation function ξ_h of tracers. If we measure the correlation function at separation r, we clearly need $R_L < r$ in order to avoid large effects of the coarse graining. However, as discussed in Sec. I, the precise value of the coarse graining scale should not have an effect on the final expression for the correlation function. We will deal with the effects of the coarse-graining on the tracer correlation function in Sec. III.

In terms of the coarse-grained densities $\hat{n}_h$, the simplest estimator for ξ_h can be written as

$$\hat{\xi}_h(r) = \sum_{r<|\mathbf{x}_i-\mathbf{x}_j|<r+\Delta r}\frac{\hat{n}_h(\mathbf{x}_i)\hat{n}_h(\mathbf{x}_j)}{\bar{n}^2_h} - 1. \quad (14)$$

Here, $\bar{n}_h$ is the mean observed abundance of tracers, which can be defined as $\bar{n}_h = N^{-1}\sum_i \hat{n}_h(\mathbf{x}_i)$, where the sum runs over a large number N of random locations. We now let Δr go to zero, and take the expectation value of Eq. (14):

$$\langle\hat{\xi}_h(r)\rangle = \frac{\langle\hat{n}_h(1)\hat{n}_h(2)\rangle}{\langle\hat{n}_h\rangle^2} - 1, \quad (15)$$

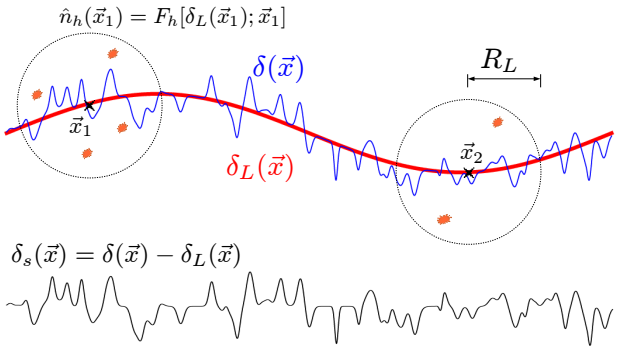

FIG. 1: Sketch of the separation of the density field (blue, thin line) into large-scale part δ_L (red, thick line) and a small scale part δ_s (Eq. (86) in Sec. IV; thin black line below), via an arbitrary coarse-graining scale R_L. The tracer density coarse-grained on scale R_L (circles) is described by the function $F_{h,L}(\delta_L;\mathbf{x})$ [Eq. (6)], where the explicit dependence on $\mathbf{x}$ encodes the scatter around the mean relation with δ_L, which is assumed to be uncorrelated with δ_L.

where '1' and '2' stand for two arbitrary locations separated by a distance r. Following the reasoning above, the derivatives of $F_{h,L}$ with respect to δ_L at the two locations separated by r are independent random variables,

$$\langle F^{(n)}_{h,L}(0;\mathbf{x}_1)\,F^{(m)}_{h,L}(0;\mathbf{x}_2)\rangle = \langle F^{(n)}_{h,L}(0;\mathbf{x}_1)\rangle\,\langle F^{(m)}_{h,L}(0;\mathbf{x}_2)\rangle. \quad (16)$$

Using Eq. (7), we can then write Eq. (15) in terms of the statistics of the coarse-grained density field δ_L, and the coefficients c_n:

$$\langle\hat{\xi}_h(r)\rangle = \frac{\sum_{n,m=0}^{\infty}\frac{c_n c_m}{n!m!}\langle\delta^n_L(1)\delta^m_L(2)\rangle}{\sum_{n,m=0}^{\infty}\frac{c_n c_m}{n!m!}\langle\delta^n_L\rangle\langle\delta^m_L\rangle} - 1. \quad (17)$$

Similarly, we can define the tracer-matter cross-correlation function (assuming that δ_L is observable somehow) and obtain its expectation value:

$$\hat{\xi}_{hm}(r) = \sum_{r<|\mathbf{x}_i-\mathbf{x}_j|<r+\Delta r}\frac{\hat{n}_h(\mathbf{x}_i)\hat{\delta}_L(\mathbf{x}_j)}{\bar{n}_h} \quad (18)$$

$$\langle\hat{\xi}_{hm}(r)\rangle = \frac{\langle\hat{n}_h(1)\delta_L(2)\rangle}{\langle\hat{n}_h\rangle} = \frac{\sum_{n=1}^{\infty}\frac{c_n}{n!}\langle\delta^n_L(1)\delta_L(2)\rangle}{\sum_{n=0}^{\infty}\frac{c_n}{n!}\langle\delta^n_L\rangle}. \quad (19)$$

These expressions involve sums over moments of the density field, which contain disconnected pieces such as $\langle\delta^2_L\rangle$, multiplied by the bare bias parameters c_n. The convergence of this sum of R_L-dependent coefficients multiplying R_L-dependent disconnected moments is apparently only assured in the general case if σ^2_L, and higher moments of the density field, are much less than one, since

terms of order $\langle\delta_L^n\rangle$ appear at arbitrarily high n. On the other hand, a physically reasonable perturbative bias model for $\xi_h(r)$, $\xi_{hm}(r)$ should converge as long as the *connected* matter correlators are much less than one, independently of the choice of the fictitious coarse-graining scale. Thus, our goal is to reorder the sum in Eq. (17) and Eq. (19) into a sum of R_L-independent coefficients b_N multiplying only powers of connected matter correlators. This can be seen as a renormalization of the "bare" coefficients c_n into "renormalized" bias parameters b_N. We will see below that they have clear physical significance.

We can expand the correlators into connected parts (cumulants) as follows. The correlator appearing in the cross-correlation Eq. (19) can be written for any statistical field as

$$\langle\delta_L^n(1)\delta_L(2)\rangle = \sum_{N=0}^{n}\binom{n}{N}\langle\delta_L^{n-N}\rangle\langle\delta_L^N(1)\delta_L(2)\rangle_c, \quad (20)$$

where the subscript c denotes connected correlators. Thus, Eq. (19) can also be written as

$$\langle\hat{\xi}_{hm}(r)\rangle = \frac{1}{\mathcal{N}}\sum_{n=1}^{\infty}\frac{c_n}{n!}\sum_{N=0}^{n}\binom{n}{N}\langle\delta_L^{n-N}\rangle\langle\delta_L^N(1)\delta_L(2)\rangle_c, \quad (21)$$

where we have defined

$$\mathcal{N} \equiv \sum_{n=0}^{\infty}\frac{c_n}{n!}\langle\delta_L^n\rangle. \quad (22)$$

The density field correlator in the auto-correlation Eq. (17) is more complicated. As shown in App. A,

$$\langle\delta_1^n\delta_2^m\rangle = \sum_{k=0}^{n}\sum_{l=0}^{m}\binom{n}{k}\langle\delta_1^k\rangle\binom{m}{l}\langle\delta_2^l\rangle\langle\delta_1^{n-k}\delta_2^{m-l}\rangle_{\rm nzl}\,, \quad (23)$$

where $\langle\cdot\rangle_{\rm nzl}$ denotes a disconnected correlator that, when expanded into cumulants, does not contain any zero-lag pieces, i.e. no factors that asymptote to a constant as $r\to\infty$ (see Eq. (A12) in App. A for a mathematical expression of this definition). We will restrict to the Gaussian case here, and postpone the discussion of non-Gaussian density fields to Sec. IV and App. A. The no-zero-lag requirement entails that $\delta_L(1)$ and $\delta_L(2)$ have to appear in equal powers in "nzl" correlators. Defining

$$\xi_L(r) \equiv \langle\delta_L(1)\delta_L(2)\rangle, \quad (24)$$

we obtain

$$\langle\delta_L^n(1)\delta_L^m(2)\rangle_{\rm nzl} \overset{\rm Gaussian}{=} n!\,[\xi_L(r)]^n\;\delta_{nm}. \quad (25)$$

The factor $n!$ comes about because there are $n!$ ways to contract $2n$ factors of δ_L (n of $\delta_L(1)$ and $\delta_L(2)$ each) into a product of n correlation functions. Further,

$$\langle\delta_L^n\rangle = (n-1)!!\,\sigma_L^n \quad (26)$$

for n even, and zero for n odd. This yields

$$\langle\delta_L^n(1)\delta_L^m(2)\rangle \overset{\rm Gaussian}{=} \sum_{k:\ n-k,\,m-k\ \text{even}}^{\min(n,m)}(n-k-1)!!(m-k-1)!! \times\binom{n}{k}\binom{m}{k}k!\,\sigma_L^{n+m-2k}[\xi_L(r)]^k. \quad (27)$$

B. PBS and bias parameters

So far, all we have done is divide the universe into fictitious regions, and describe the number density of tracers in regions in terms of a function $F_{h,L}$ [Eq. (6)]. We were then able to formally express the correlations of tracers in terms of the statistics of the matter density and the derivatives of the function $F_{h,L}$, all of which depend on the coarse-graining scale R_L.

We now turn to the peak-background split argument, and the definition of the PBS bias parameters with respect to density. The argument can be summarized as follows: if the description of the clustering of tracers solely through their dependence on δ_L is sufficient, then the expected abundance of tracers in a region $\mathcal{U}$ characterized by a coarse-grained overdensity $\delta_L = D$ is sufficiently well approximated by the average abundance of tracers $\langle\hat{n}_h\rangle$ in a fictitious Universe with modified background density

$$\overline{\rho}' = \overline{\rho}(1+D)\,, \quad (28)$$

where $\overline{\rho}$ is the actual background density.

The advantage of this approach is that we only need a prediction for $\langle\hat{n}_h\rangle$ as function of the background density $\overline{\rho}'$ to calculate the statistics of tracers; no knowledge of the function $F_{h,L}$ is necessary. Note also that this is directly connected to the derivation of bias (linear bias in that case) in the relativistic context presented in [22]. Specifically, we are working in the synchronous gauge where all space-time points on an equal-coordinate-time hypersurface share the same cosmic age. Correspondingly, when calculating $\langle\hat{n}_h\rangle$ for varying $\overline{\rho}'$ it is crucial to keep the age of the Universe fixed.

Thus, we now consider the case where we perturb the background density by $\Delta\overline{\rho} = D\overline{\rho}$, where D is an infinitesimal parameter. Thus, in a region with overdensity δ_L the matter density is perturbed to

$$\rho_L = \overline{\rho}(1+\delta_L) \to \overline{\rho}(1+\delta_L) + \Delta\overline{\rho} = \overline{\rho}(1+\delta_L+D) \quad (29)$$

Note that we add a fixed amount of *uniform* matter density everywhere; we do not rescale the local matter density ρ by $1+D$, which would also amplify the fluctuations δ. We can obtain the average number density of tracers $\langle\hat{n}_h\rangle$ (more precisely, the expectation value of the estimated mean number density in some volume) in such a Universe from the expansion in terms of coarse-grained

δ_L, Eq. (11):

$$\langle \hat{n}_h \rangle|_D = \langle F_{h,L}(0) \rangle \sum_{n=0}^{\infty} \frac{c_n}{n!} \langle (\delta_L + D)^n \rangle, \qquad (30)$$

where $F_{h,L}$ and c_n both refer to the Universe with background density $\overline{\rho}$, i.e. $D = 0$.

Let us now define the *peak-background split bias parameters* b_N ($N \geq 1$):

$$b_N \equiv \frac{1}{\langle \hat{n}_h \rangle|_{D=0}} \frac{\partial^N \langle \hat{n}_h \rangle|_D}{\partial D^N}\Big|_{D=0}. \qquad (31)$$

Using Eq. (28), we can also write this as

$$b_N = \frac{\overline{\rho}^N}{\langle \hat{n}_h \rangle} \frac{\partial^N \langle \hat{n}_h \rangle}{\partial \overline{\rho}^N}\,, \qquad (32)$$

where the derivatives are evaluated at the fiducial value of $\overline{\rho}$. It is worth emphasizing the difference between these bias parameters and the c_n defined in the last section: the b_N quantify the response of the cosmic mean abundance of tracers to a change in the background density of the Universe; specifically, they do not make any reference to the regions $\mathcal{U}$, or the scale R_L. The c_n on the other hand quantify the average response of the abundance of tracers within a region $\mathcal{U}$ to a change in the average density δ_L within that region, evaluated at $\delta_L = 0$; they thus necessarily depend on the filter function W_L and scale R_L. Further, there is no "scatter" in the defining relation Eq. (31) for the PBS biases, although the values for b_N measured in reality will clearly have a finite error bar as we can only approximate this relation within a finite volume.

The b_N are closely related to the resummed bias propagators defined in [11] [see Eqs. (83)–(84) there], while the bare bias parameters c_n correspond to the bare propagators [Eqs. (1)–(2) in that paper].

Using Eq. (30), we can derive an algebraic relation between b_N and c_n:

$$b_N = \frac{1}{\mathcal{N}} \sum_{n=N}^{\infty} \frac{c_n}{n!} \frac{n!}{(n-N)!} \langle \delta_L^{n-N} \rangle. \qquad (33)$$

By reordering the sum in Eq. (21), we immediately see that the tracer-matter cross-correlation in terms of the PBS bias parameters is given by

$$\langle \hat{\xi}_{hm}(r) \rangle = \sum_{N=1}^{\infty} \frac{b_N}{N!} \langle \delta_L^N(1) \delta_L(2) \rangle_c. \qquad (34)$$

This is the well-known bias expansion of the tracer-matter cross-correlation function. However, note that the matter correlators appearing here are the *connected* correlators. In particular, in the case of a Gaussian density field we obtain

$$\langle \hat{\xi}_{hm}(r) \rangle = b_1 \xi_L(r), \qquad (35)$$

i.e. the tracer-matter cross-correlation function is simply proportional to the linear matter correlation function (the same conclusion was reached by [28], who only considered Gaussian density fields). Similarly, using Eq. (33) and Eq. (27) we can re-express the auto-correlation [Eq. (17)] as

$$\langle \hat{\xi}_h(r) \rangle = \sum_{N,M=1}^{\infty} \frac{b_N}{N!} \frac{b_M}{M!} \langle \delta_L^N(1) \delta_L^M(2) \rangle_{\rm nzl}\,. \qquad (36)$$

This relation straightforwardly generalizes to the cross-correlation between two different tracers h_1, h_2, yielding

$$\langle \hat{\xi}_{h_1 h_2}(r) \rangle = \sum_{N,M=1}^{\infty} \frac{b_N^{(1)}}{N!} \frac{b_M^{(2)}}{M!} \langle \delta_L^N(1) \delta_L^M(2) \rangle_{\rm nzl}\,,$$

where $b_N^{(i)}$ denotes the PBS bias prameter for tracer h_i [Eq. (34) is of course a special case of this, with $b_1 = 1$, $b_{N>1} = 0$ for matter].

Even though Eq. (27) assumes a Gaussian density field, Eq. (36) is in fact valid for a general non-Gaussian density field. The proof in this case requires somewhat more effort and is given in App. A. We point out that the derivation of Eq. (36) is equivalent to the renormalization of multi-point propagators [10, 11, 29], and valid for general statistical fields. Note that there is no b_0; the expressions Eqs. (34)–(36) only involve terms with b_N for $N \geq 1$. In the Gaussian case, Eq. (36) further simplifies to

$$\langle \hat{\xi}_h(r) \rangle = \sum_{N=1}^{\infty} \frac{b_N^2}{N!} \left[\xi_L(r)\right]^N. \qquad (37)$$

Eqs. (34)–(37) achieve the desired result: an expansion of the tracer correlation function in terms of R_L-independent bias parameters which multiply powers of the matter correlation function $\xi_L(r)$ (or, more generally, no-zero-lag correlators). The series in Eqs. (36)–(37) have a convergence radius set solely by the values of the b_N and the amplitude of the matter correlation function at scale r, which is what we expect from a physical bias expansion.[1] On the other hand, in the bare bias expansion, Eq. (17), terms of order σ_L^n appear at every successive higher order, suggesting that we need to coarse-grain the density field on quasi-linear scales in order for the expansion to be perturbatively valid. This of course would be disastrous for any sharp features in the correlation function; for example, choosing a coarse-graining scale of $R_L = 30 - 50\,{\rm Mpc}/h$ would erase the baryon acoustic oscillation (BAO) feature at $r = 150\,{\rm Mpc}/h$. In the

[1] If $b_N^2/N!$ grows faster with N than an exponential, e.g, if $b_N \sim N^{aN}$ with $a > 1/2$, then the series does not converge for any r. In this case, our approach does not make a prediction for the clustering of tracers in the large-scale limit. Through Eq. (31), this essentially means that $\langle \hat{n}_h \rangle$ is defined only for one exact value of $\overline{\rho}$ and nowhere else, which is clearly not a physical behavior.

PBS bias expansion, there is no need to choose a coarse-graining scale this large. Rather, the validity of the the result Eqs. (34)–(37) is determined by the requirement that any residual dependence on R_L be negligible.

Another important property of this expansion is that the same PBS bias parameters describe both the matter-tracer cross-correlation and the tracer auto-correlation, which is what we expect from a physical bias expansion (the corresponding statement in Fourier space is complicated by the small-scale effects such as stochasticity, shot noise and halo exclusion which contribute to the tracer power spectrum at all k).

One crucial advantage of this approach is that we have an indicator for when the underlying assumptions break down: if evaluation of Eq. (36) shows that the result is in fact R_L-dependent, then we know that the underlying assumptions, in particular the description $\hat{n}_h(\mathbf{x}) = F_{h,L}(\delta_L(\mathbf{x}))$, break down. One then has to identify the physical reason for this R_L-dependence, and add a dependence of the tracer number density on additional parameters which will absorb ("renormalize") the R_L-dependence. We will see two important examples of this in Sec. III and Sec. IV. First however, we will derive the PBS biases for the widely considered case of universal mass functions, and then illustrate the approach on a concrete example of biased tracer: regions above threshold.

C. PBS biases for universal mass functions

The mean abundance of tracers such as dark matter halos of some mass M_* is often parametrized in the form

$$\bar{n}_h = \overline{\rho}\, f(\nu_c)\, J_* \tag{38}$$

$$\nu_c \equiv \frac{\delta_c}{\sigma_*}; \quad J_* \equiv \frac{d\ln\sigma_*}{d\ln R_*}, \tag{39}$$

where σ_* is the variance of the linear matter density field on scale R_*, R_* is related to the mass M_* through $M_* = 4\pi/3\,\overline{\rho} R_*^3$, and δ_c is the linearly extrapolated threshold for collapse. Further, $f(\nu_c)$ is in general an arbitrary function of ν_c. The Jacobian J_* is present in order to convert from an interval in σ_* to a mass interval. Eq. (38) is referred to as "universal mass function" and was originally motivated by the excursion set formalism [30]. It is a special case of a more general description of mean tracer abundance we will consider in Sec. IV E.

In order to derive the bias parameters Eq. (31), we need to know how $\bar{n}_h$ changes under a change in the background density of the Universe [Eq. (28)]. Since we work in the Lagrangian picture, we will ignore the trivial dependence through the $\overline{\rho}$ prefactor in Eq. (38). The threshold δ_c is defined as the fractional overdensity a region must have to collapse[2] at a fixed proper time t_0. In an Einstein-de Sitter Universe, a spherical perturbation with a mean initial fractional overdensity $\delta_c \approx 1.686$, i.e. with $\rho(< R, t) = [1 + a(t)\delta_c]\,\overline{\rho}(t)$ average interior density for $a(t) \ll 1$, collapses at $a = 1$. The same reasoning also holds for more general expansion histories, where δ_c assumes other values. Since the evolution of such a perturbation is independent of the external Universe (by Birkhoff's theorem), a perturbation of the same physical density ρ_c will collapse at the same proper time in a Universe with perturbed background density $\overline{\rho}' = \overline{\rho}(1 + D)$ as well. The significance $\nu_c = \delta_c/\sigma_* = (\rho_c - \overline{\rho})/\delta\rho_{\rm RMS}$ quantifies how rare fluctuations above a physical density threshold $\rho_c = (1 + \delta_c)\overline{\rho}$ are given the RMS fluctuation amplitude $\delta\rho_{\rm RMS} = \sigma_*\overline{\rho}$. Clearly, if we add a uniform matter density component $D\,\overline{\rho}$, the critical overdensity changes to

$$\rho_c - \overline{\rho}' = (1 + \delta_c)\overline{\rho} - (1 + D)\overline{\rho} = (\delta_c - D)\overline{\rho}\,. \tag{40}$$

Thus, the significance is modified to

$$\nu_c' = \frac{\rho_c - \overline{\rho}'}{\sigma_*\overline{\rho}} = \frac{\delta_c - D}{\sigma_*}\,. \tag{41}$$

For a mass function of the form Eq. (38), changing the background density is thus equivalent to changing $\delta_c \to \delta_c - D$. Eq. (38) and Eq. (31) thus immediately yield

$$b_N = \frac{(-1)^N}{\langle\hat{n}_h\rangle}\frac{\partial^N\langle\hat{n}_h\rangle}{\partial\delta_c^N} = \frac{(-1)^N}{\sigma_*^N}\frac{1}{f(\nu_c)}\frac{d^N f(\nu_c)}{d\nu_c^N}. \tag{42}$$

This is the widely known expression for the peak-background split bias parameters, which in our approach is a special case of Eq. (31).

D. Application to regions above threshold

We now turn to a simple example of tracer for which an *exact* expression of the tracer correlations is known. Precisely, we define our tracer to be a region where the density field δ is above a fixed threshold $\delta_c = \nu_c\sigma_*$, where σ_* is the RMS fluctuation of the density field and ν_c an arbitrary fixed parameter. The density field can be thought of as smoothed on some scale r_*; however, since for our purposes this scale is irrelevant, we will not make this smoothing explicit in our notation in order to avoid confusion. Note also that we do not make any assumption about ν_c, such as the "high peak limit". In the present case, unlike peaks of the density field, the tracer population is not a (countable) point set. Rather, tracers cover a finite volume, and the tracer number can be defined as a continuous field

$$N_h(\mathbf{x}) = \Theta(\delta(\mathbf{x}) - \nu_c\sigma_*). \tag{43}$$

[2] Since General Relativity is scale-free, this threshold is independent of the size and enclosed mass of the perturbation.

The number density of tracers as defined earlier in this section is then given by

$$\hat{n}_h(\mathbf{x}) = \int d^3\mathbf{y} N_h(\mathbf{y}) W_L(\mathbf{x}-\mathbf{y}). \tag{44}$$

We first review the exact approach to the clustering of such "tracers", and then investigate the PBS prediction. Note that essentially all these results have already been derived in [11, 31–36]. However, we review it here in light of the discussion presented above.

1. Exact calculation for a Gaussian density field

The mean "number density" $\langle \hat{n}_h \rangle$ of the tracers defined above is simply given by the fraction of the total volume that is above the threshold $\nu_c \sigma_*$,

$$\langle \hat{n}_h \rangle = P_1(\nu_c), \tag{45}$$

where $P_1(\nu_c)$ is the probability that the density field at a random location is larger than $\nu_c \sigma_*$. If the underlying density field follows Gaussian statistics with variance σ_*, P_1 is given by

$$P_1(\nu_c) = \frac{1}{\sqrt{2\pi}} \int_{\nu_c}^{\infty} dx e^{-x^2/2} = \frac{1}{2} \mathrm{erfc}\left(\frac{\nu_c}{\sqrt{2}}\right). \tag{46}$$

The exact expression for the two-point correlation function of our "tracers", $\xi_h(r)$, is then given by the probability of finding two peaks P_2 separated by r, relative to the random probability (Sec. III B of [31]) through

$$\begin{aligned}\xi_h(r) &= \frac{P_2(\nu_c; r)}{[P_1(\nu_c)]^2} - 1 \\ &= \frac{2}{\pi}\left[\mathrm{erfc}\left(\frac{\nu_c}{\sqrt{2}}\right)\right]^{-2} \\ &\quad \times \sum_{N=1}^{\infty} \frac{\left[\xi(r)\right]^N}{N! \sigma_*^{2N}} \left[H_{N-1}(\nu_c)\right]^2 e^{-\nu_c^2}.\end{aligned} \tag{47}$$

Here, $\xi(r)$ is the two-point correlation function of the underlying density field smoothed on the scale R_*, so that $\xi(0) = \sigma_*^2$.

2. Peak-background split calculation

Eq. (36) gives the tracer correlation function in terms of the PBS bias parameters and powers of $\xi_L(r)$. Note however that given the explicit relation between $\hat{n}_h$ and δ through Eq. (43) in this simple example, the division into regions $\mathcal{U}$ is a purely conceptual device here, and we can always set R_L to be equal to or smaller than the smoothing scale adopted in the thresholding approach. Hence, we will drop the subscript R_L below.

The PBS bias parameters are defined by Eq. (31). Given our definition of tracers as regions where $\delta(\mathbf{x}) > \delta_c$, or equivalently, $\rho(\mathbf{x}) > \overline{\rho}(1+\delta_c)$, we see that a fractional change D in the background density is equivalent to a change in the threshold δ_c:

$$\overline{\rho} \to \overline{\rho}(1+D) \;\Leftrightarrow\; \delta_c \to \delta_c - D. \tag{48}$$

Hence, the PBS bias parameters are given by

$$b_N = \frac{(-1)^N}{\langle \hat{n}_h \rangle} \frac{\partial^N \langle \hat{n}_h \rangle}{\partial \delta_c^N} = \frac{(-1)^N}{\sigma_*^N} \frac{1}{P_1(\nu_c)} \frac{d^N P_1}{d\nu_c^N}. \tag{49}$$

This of course can also be derived by noting that the "abundance" of regions above treshold Eq. (45) is a special case of universal mass functions (Eq. (38) without the Jacobian factor which is irrelevant for the b_N), so that Eq. (42) applies. By using the generating function of the (probabilists') Hermite polynomial

$$H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n}\left(e^{-x^2/2}\right) \tag{50}$$

we calculate the n-th derivative of P_1 ($n \geq 1$) as

$$\begin{aligned}\frac{d^n P_1}{d\nu_c^n} &= \frac{d^{(n-1)}}{d\nu_c^{(n-1)}}\left(-\frac{e^{-\nu_c^2/2}}{\sqrt{2\pi}}\right) \\ &= (-1)^n H_{n-1}(\nu_c) \frac{e^{-\nu_c^2/2}}{\sqrt{2\pi}}.\end{aligned} \tag{51}$$

Thus, we can explicitly write the PBS bias parameters for our tracers:

$$b_N = \sqrt{\frac{2}{\pi}} \left[\mathrm{erfc}\left(\frac{\nu_c}{\sqrt{2}}\right)\right]^{-1} \frac{e^{-\nu_c^2/2}}{\sigma_*^N} H_{N-1}(\nu_c). \tag{52}$$

Inserting this into Eq. (37), we immediately obtain the PBS prediction for the correlation of thresholded regions:

$$\xi_h(r) = \frac{2}{\pi}\left[\mathrm{erfc}\left(\frac{\nu_c}{\sqrt{2}}\right)\right]^{-2} e^{-\nu_c^2} \sum_{N=1}^{\infty} \frac{H_{N-1}(\nu_c)}{N!\,\sigma_*^{2N}} \left[\xi(r)\right]^N. \tag{53}$$

We see that this agrees with the direct calculation, Eq. (47). A mathematically similar derivation was presented in [31]. The difference is that in [31], we used Eq. (49) and Eq. (47) to *infer* the general relation Eq. (37). Here, we are simply illustrating how the independently derived Eq. (37) applies to the case of thresholded regions, a case where we know explicitly the function $F_{h,L}(\delta_L)$. As proven in the previous section, Eq. (37) and the much more general Eq. (36) apply to *any* tracer as long as the dependence of the tracer density on other quantities apart from the matter density can be neglected.

One alternative to the simple local bias expansion Eq. (1) in the context of thresholded regions is to expand Eq. (43) in terms of Hermite polynomials, as done in [11, 32, 36] (see also [3, 4]):

$$\Theta(\nu - \nu_c) = \sum_{n=0}^{\infty} a_n(\nu_c) H_n(\nu) \tag{54}$$

where $\nu(\mathbf{x}) \equiv \delta(\mathbf{x})/\sigma_*$, and

$$a_n(\nu_c) = \begin{cases} \frac{1}{2}\mathrm{erfc}\left(\frac{1}{\sqrt{2}}\nu_c\right) & n=0 \\ \frac{1}{n!}\frac{1}{\sqrt{2\pi}}e^{-\nu_c^2/2}H_{n-1}(\nu_c) & n\geq 1 \end{cases}. \tag{55}$$

The bias parameters they obtain are exactly equal to our renormalized PBS bias parameters, Eq. (52). The reason for this is that, in the Gaussian case, the Hermite expansion ensures that no disconnected pieces remain in the correlation function expression Eq. (17). Specifically, denoting $\nu_i = \nu(\mathbf{x}_i)$, we have

$$\begin{aligned}\langle H_n(\nu_1)H_m(\nu_2)\rangle &= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\nu_1\, H_n(\nu_1)\int_{-\infty}^{\infty} d\nu_2\, H_m(\nu_2) \\ &\quad\times \exp\left(\frac{\xi(r_{12})}{\sigma_*^2}\frac{\partial^2}{\partial\nu_1\partial\nu_2}\right)e^{-(\nu_1^2+\nu_2^2)/2} \\ &= \frac{1}{2\pi}\sum_{N=0}^{\infty}\frac{1}{N!}\left(\frac{\xi(r_{12})}{\sigma_*^2}\right)^N\int d\nu_1 H_n(\nu_1)H_N(\nu_1)e^{-\nu_1^2/2} \\ &\quad\times\int d\nu_2 H_m(\nu_2)H_N(\nu_2)e^{-\nu_2^2/2} \\ &= n!\left(\frac{\xi(r_{12})}{\sigma_*^2}\right)^n\delta_{nm}. \end{aligned} \tag{56}$$

The Hermite expansion is thus an elegant way of directly obtaining renormalized bias parameters in the case of thresholding in a Gaussian density field. However, the additional contributions obtained in the non-Gaussian case spoil this renormalization of all zero-lag terms, as we will see in Sec. IV.

III. SMOOTHED CORRELATION FUNCTION AND CURVATURE BIAS

Above we explained that the expression of the tracer correlation in terms of PBS bias parameters and connected matter correlators [Eq. (36) and Eq. (37) for the general and Gaussian case, respectively] should be numerically insensitive to the value of the coarse-graining scale R_L. Further, a significant R_L-dependence indicates the break-down of our assumption that the tracer density is a function only of the local coarse-grained density.

For $r \gg R_L$, and as long as $\xi(r)$ is smooth (e.g., a power law), the smoothed version of $\xi(r)$, $\xi_L(r)$, will not differ significantly from $\xi(r)$. However, if $\xi(r)$ has some features on a scale $\delta r \ll r$, such as the BAO feature with $\delta r \sim 20\,\mathrm{Mpc}/h$, then the condition for R_L-independence becomes much more restrictive: $R_L \ll \delta r$. We now show how the R_L-dependence induced through $\xi_L(r)$ can be cured.

For a general isotropic filter function (see Eq. (B1) in App. B) the effect of smoothing on the correlation function $\xi(r)$ can be perturbatively described through

$$\begin{aligned}\xi_L(r) &= \int\frac{d^3k}{(2\pi)^3}|\tilde{W}_L(k)|^2P(k)e^{i\mathbf{k}\cdot\mathbf{r}} \\ &= \int\frac{d^3k}{(2\pi)^3}\left(1-2R_L^2k^2+\mathcal{O}(k^4)\right)P(k)e^{i\mathbf{k}\cdot\mathbf{r}} \\ &= \xi(r)+2R_L^2\nabla^2\xi(r)+\mathcal{O}(\nabla^4\xi(r)),\end{aligned} \tag{57}$$

by suitable definition of the parameter R_L. In App. B, we give the general expansion of $\tilde{W}_L(k)$ [Eq. (B2)] and the expansion of $\xi_L(r)$ in terms of derivatives of $\xi(r)$ [Eq. (B3)]. Thus, if $R_L^2\nabla^2\xi$ is comparable to ξ, our requirement of R_L-independence does not hold. In our approach, this signals a breakdown of the underlying assumption that tracer statistics can be described purely by their dependence on the local matter density. Instead, let us assume that the local number density also depends on the coarse-grained Laplacian of the density field, i.e. the curvature:

$$\hat{n}_h(\mathbf{x}) = F_{h,L}(\delta_L(\mathbf{x});\nabla^2\delta_L(\mathbf{x});\mathbf{x}). \tag{58}$$

The Laplacian is the lowest order term in derivatives of δ_L, because a dependence on the gradient of δ_L would imply a preferred direction[3]. In general, we now have to perform a bivariate expansion of the function $F_{h,L}$ in δ_L and $\nabla^2\delta_L$. Let us for now restrict to lowest order to keep the treatment clear, and consider the Gaussian case. The expansion to higher orders in derivatives of δ_L is described in App. B. The tracer auto-correlation becomes

$$\begin{aligned}\xi_h(r) &= c_1^2\langle\delta_L(1)\delta_L(2)\rangle + 2c_1c_{\nabla^2\delta}\langle\delta_L(1)\nabla^2\delta_L(2)\rangle \\ &\quad+\mathcal{O}(\nabla^4\xi) \\ &= c_1^2\left[\xi(r)+2R_L^2\nabla^2\xi(r)\right]+2c_1c_{\nabla^2\delta}\nabla^2\xi(r).\end{aligned} \tag{59}$$

Here, we have defined

$$c_{\nabla^2\delta} = \frac{1}{\langle F_{h,L}(0)\rangle}\left\langle\frac{\partial F_{h,L}}{\partial(\nabla^2\delta_L)}\Big|_{\delta_L=0,\nabla^2\delta_L=0}\right\rangle. \tag{60}$$

Eq. (59) is again phrased in terms of (in general) disconnected matter correlators and R_L-dependent bare bias parameters. We now need to introduce a R_L-independent PBS bias parameter for $\nabla^2\delta$ as defined in Sec. II B for the density itself. We would like a transformation where the Laplacian of the density perturbation shifts by a constant:

$$\nabla^2\delta_\alpha(\mathbf{x}) = \nabla^2\delta(\mathbf{x}) + \frac{\alpha}{\ell^2}, \tag{61}$$

where α is a dimensionless small parameter, and we have added a length scale ℓ. This corresponds to

$$\delta_\alpha(\mathbf{x}) = \delta(\mathbf{x}) + \frac{\alpha}{6\ell^2}\left(\mathbf{x}^2+\mathbf{A}\cdot\mathbf{x}+C\right), \tag{62}$$

[3] Terms such as $(\boldsymbol{\nabla}\delta_L)^2$ could however appear at second and higher order.

where $\mathbf{A}$ and C are constants and the center of the region considered is chosen as the origin. We are not interested in adding a gradient to the density field and hence set $\mathbf{A} = 0$. Note that Eq. (62) is only defined for a region of finite size (e.g., a simulation box), so that $\langle n_h \rangle$ in the following is to be considered as an ensemble average over many such finite regions. We will also set $C = 0$ so that $\delta(\mathbf{0})$, at the center of the region considered, is unchanged (a constant shift in δ such as described by C just corresponds to the density bias transformation of Sec. II B). Thus,

$$\delta_\alpha(\mathbf{x}) = \delta(\mathbf{x}) + \frac{\alpha}{6\ell^2}\mathbf{x}^2. \tag{63}$$

We can now defined a (renormalized) PBS bias parameter through

$$b_{\nabla^2\delta} = \frac{\ell^2}{\langle \hat{n}_h \rangle} \frac{\partial \langle \hat{n}_h(\mathbf{0}) \rangle}{\partial \alpha}\Big|_{\alpha=0}. \tag{64}$$

Defined in this way, the scale ℓ will disappear out of the final expression for the tracer correlation function (note that $b_{\nabla^2\delta}$ has dimension length squared). In order to derive the relation between $b_{\nabla^2\delta}$ and the c_n, we need the behavior of both $\delta_L(\mathbf{x})$ and $\nabla^2\delta_L(\mathbf{x})$ under the transformation Eq. (63). The latter is immediately obtained as

$$\nabla^2\delta_{L,\alpha}(\mathbf{x}) = \nabla^2\delta_L(\mathbf{x}) + \frac{\alpha}{\ell^2}. \tag{65}$$

We are interested in the change of δ_L near the origin (the center of the region $\mathcal{U}$). In analogy with Eq. (57), we obtain

$$\begin{aligned}\delta_{L,\alpha}(\mathbf{0}) &= \int d^3\mathbf{y}\, W_L(\mathbf{y}) \left[\delta(\mathbf{y}) + \frac{1}{6}\frac{\alpha}{\ell^2}\mathbf{y}^2\right] \\ &= \delta_L(\mathbf{0}) + \alpha\frac{R_L^2}{\ell^2}.\end{aligned} \tag{66}$$

Thus, Eq. (58) yields at lowest order

$$\begin{aligned}b_{\nabla^2\delta} &= \frac{\ell^2}{F_{h,L}(0)} \\ &\quad\times \left(\frac{\partial F_{h,L}(\delta_L,0)}{\partial\delta_L}\frac{\partial\delta_L}{\partial\alpha} + \frac{\partial F_{h,L}(0,\nabla^2\delta_L)}{\partial(\nabla^2\delta_L)}\frac{\partial(\nabla^2\delta_L)}{\partial\alpha}\right) \\ &= c_1 R_L^2 + c_{\nabla^2\delta}.\end{aligned} \tag{67}$$

Now, we can write the generalization of our previous result, Eq. (37), at lowest order, assuming we absorb the smoothing effect on $\xi(r)$, as

$$\xi_h(r) = b_1^2\xi(r) + 2b_1 b_{\nabla^2\delta}\nabla^2\xi(r), \tag{68}$$

where the factor of 2 comes from the two permutations when writing down all mixed no-zero-lag correlators between δ and $\nabla^2\delta$. Using Eq. (67), this is equal to

$$\xi_h(r) = c_1^2\xi(r) + 2\left[c_1^2 R_L^2 + c_1 c_{\nabla^2\delta}\right]\nabla^2\xi(r) + \mathcal{O}(\nabla^4\xi(r)),$$

exactly matching the result of Eq. (59). Thus, by introducing a dependence of the tracer density on the Laplacian of the density field, and a corresponding PBS bias parameter, we are able to absorb the effects of the coarse-graining on the correlation function. Moreover, in App. B we show that this continues to arbitrary powers of derivatives, and that the PBS bias parameters can *entirely* absorb the smoothing effect on ξ_L [Eq. (B16)]. Specifically, up to order $\nabla^4\xi(r)$, we obtain

$$\begin{aligned}\xi_h(r) = b_1^2\xi(r) &+ 2b_1 b_{\nabla^2\delta}\nabla^2\xi(r) \\ &+ \left[(b_{\nabla^2\delta})^2 + b_1 b_{\nabla^4\delta}\right]\nabla^4\xi(r).\end{aligned} \tag{69}$$

There are two important implications of this result. First, our approach, which does not make any assumptions on the tracers themselves, generically predicts the existence of a bias with respect to $\nabla^2\delta$, which in k-space corresponds to a scale-dependent bias $\propto k^2$. One can interpret this as the statement that the tracer density is in general not a truly local function of the matter density, but depends on the matter distribution within a finite region whose characteristic scale is given by $\sqrt{b_{\nabla^2\delta}}$.

Second, the definition of the PBS bias parameter $b_{\nabla^2\delta}$ [Eq. (64)] has a clear physical interpretation: it corresponds to the response of the tracer number density to a uniform shift in the curvature of the density field (and shifts of higher derivatives in the general case, Eq. (B4)). Below we will show how this bias can be evaluated for an analytical example, peaks of the density field. A quantitative test on N-body simulations will be the subject of future work.

Beyond linear order in the matter correlation function, one in principle has to expand the tracer density in a multivariate bias expansion of $\nabla^{2n}\delta_L$ ($n \geq 0$). Although we have only shown that this expansion removes the R_L-dependence contained in $\xi_L(r)$ at linear order in ξ, we expect this to be the case for higher orders as well. In practice, the suppression of $b_{\nabla^{2n}\delta}\nabla^{2n}\xi$ compared to ξ ensures that one only needs to keep a finite number of terms.

A. Connection with the peak model

In the peak model, we identify large-scale structure tracers with discrete peaks of the density field above some threshold δ_c. These distinct peaks constitute a point set (in contrast to the regions above threshold considered in Sec. II D). For a Gaussian density field, it is possible to calculate the two-point correlation of these peaks exactly [37, 38]. It is well known that peaks exhibit a scale-dependent bias in Fourier space $\propto k^2$ which is equivalent to a bias with respect to $\nabla^2\delta$ [37–39]. In this section, we show how the approach outlined in the previous section relates to this model.

As in Sec. II D, we smooth the density field on a scale R_* and define the local significance $\nu(\mathbf{x}) = \delta(\mathbf{x})/\sigma_*$. Note that the smoothing scale here is physical, unlike

the fictitious coarse-graining scale adopted in the renormalization approach described above. Apart from R_*, which is usually identified with the Lagrangian radius of the halos considered and is irrelevant for the discussion here, the peak model by itself does not involve any other coarse-graining scale. We begin by outlining the derivation of the mean number density of peaks, which follows App. A of [40], but will use the slightly different notation of [37, 38].

The location $\mathbf{x}_p$ of a peak is defined through constraints on the overdensity δ (or equivalently ν), its gradient $\boldsymbol{\eta} \equiv \boldsymbol{\nabla}\delta$, and its Hessian $\zeta_{ij} \equiv \partial_i\partial_j\delta$:

$$\begin{aligned} \nu(\mathbf{x}_p) &> \nu_c\sigma_* \\ \boldsymbol{\eta}(\mathbf{x}_p) &= 0 \\ \lambda_1(\mathbf{x}),\, \lambda_2(\mathbf{x}), \lambda_3(\mathbf{x}) &> 0, \end{aligned} \tag{70}$$

where $\nu_c = \delta_c/\sigma_*$ is the scaled threshold, and λ_i are the eigenvalues of ζ_{ij}. Let $\mathbf{V}$ denote the 10-component vector consisting of $\nu, \boldsymbol{\eta}$, and the six independent components of ζ_{ij}. The differential (in terms of ν) mean number density of peaks is then given by

$$\begin{aligned} \bar{n}_{\rm pk}(\nu_c) \propto & \int \delta_D(\nu-\nu_c)\delta_D(\boldsymbol{\eta})\Theta(\lambda_1)\Theta(\lambda_2)\Theta(\lambda_3) \\ & \times e^{-Q(\mathbf{V})} d^{10}\mathbf{V} \\ Q(\mathbf{V}) = & \frac{1}{2}\mathbf{V}^T M^{-1}\mathbf{V}, \end{aligned} \tag{71}$$

where M is the covariance matrix of $\mathbf{V}$. The most difficult part of the calculation is deriving the integration region and measure of $\mathbf{V}$, but we will not need to deal with this explicitly as we are only interested in how $\bar{n}_{\rm pk}$ transforms under a change in $\ln\bar{\rho}$ and under the transformation Eq. (61). Following Bardeen et al. [40] we introduce new variables u, y, z (u is their x), where in particular

$$u \equiv -\frac{\mathrm{Tr}\,\zeta_{ij}}{\sigma_2} = -\frac{\nabla^2\delta}{\sigma_2}, \quad \sigma_2^2 = \langle(\nabla^2\delta)^2\rangle, \tag{72}$$

corresponds to minus the curvature of the density field scaled to unit variance. The log-likelihood Q then becomes

$$2Q = \nu^2 + \frac{(u-u_*)^2}{1-\gamma^2} + 2Q_+, \tag{73}$$

where

$$u_* = \gamma\nu, \quad \text{and} \quad \gamma = \langle\nu u\rangle \tag{74}$$

is a scaled spectral moment quantifying the correlation between ν and u. $2Q_+$ is the log-likelihood of y, z, $\boldsymbol{\eta}$, and the other components of ζ_{ij}, and is independent of ν and u.

Bardeen et al. [40] then find for the mean differential peak number density

$$\bar{n}_{\rm pk} = \frac{1}{(2\pi)^2 R_1^3}\, G_0(\gamma, \gamma\nu_c) e^{-\nu_c^2/2}\,, \tag{75}$$

where R_1 is the characteristic radius of a peak [Eq. (19) in [38]], and

$$G_0(\gamma, u_*) = \int_0^\infty du\, f(u) \frac{\exp\left(-\frac{(u-u_*)^2}{2(1-\gamma^2)}\right)}{\sqrt{2\pi(1-\gamma^2)}}. \tag{76}$$

Here, $f(u)$ is a function encoding the integral over y and z which accounts for the asphericity of the peak profile [40].

Going back to Eq. (73), we see that adding a uniform density component corresponds to a change in the threshold as described in Sec. II D, whereas the other variables u, y, z are not affected. Thus, the density bias parameters are given by

$$b_N = \frac{(-1)^N}{\bar{n}_{\rm pk}} \frac{\partial^N \bar{n}_{\rm pk}}{\partial(\delta_c)^N}, \tag{77}$$

where δ_c is the height of the density threshold. These PBS bias parameters precisely agree with the coefficients of $\xi(r)$ and $[\xi(r)]^2$ derived from an explicit computation of the peak 2-point correlation function [38]. The relation Eq. (77) was already pointed out in [15].

More interesting in this context is the derivation of the bias with respect to $\nabla^2\delta$. We consider the transformation Eq. (61) in a region of finite size ℓ, so that the effect on the density and ν is negligible if $\alpha \ll 1$. Thus, all components of the vector $\mathbf{V}$ are unaffected with the exception of u. Instead of following a Gaussian of mean zero and variance of 1, u is now Gaussian-distributed around (recall the minus sign in the definition of u)

$$\langle u\rangle_\alpha = -\frac{\alpha}{\sigma_2\ell^2}, \tag{78}$$

with unit variance. Thus, the log-likelihood Eq. (73) changes to

$$2Q(\alpha) = \nu^2 + \frac{1}{1-\gamma^2}\left(u + \frac{\alpha}{\sigma_2\ell^2} - u_*\right)^2 + 2Q_+. \tag{79}$$

Clearly, this is equivalent to changing

$$u_* \to u_* - \frac{\alpha}{\sigma_2\ell^2}. \tag{80}$$

The remainder of the calculation of $\bar{n}_{\rm pk}$ follows through as before, since it is independent of the distribution of u. We thus obtain for the renormalized curvature bias [Eq. (64)] in the peak model

$$\begin{aligned} b_{\nabla^2\delta} &= \frac{\ell^2}{\bar{n}_{\rm pk}} \frac{\partial\bar{n}_{\rm pk}}{\partial\alpha}\Big|_{\alpha=0} = -\frac{1}{\bar{n}_{\rm pk}\sigma_2} \frac{\partial\bar{n}_{\rm pk}}{\partial u_*}\Big|_{u_*=\gamma\nu_c} \\ &= -\frac{1}{G_0(\gamma,\gamma\nu_c)\sigma_2} \frac{\partial G_0(\gamma,u_*)}{\partial u_*}\Big|_{u_*=\gamma\nu_c} \\ &= -(G_0\sigma_2)^{-1}\int_0^\infty du \left(\frac{u-\gamma\nu_c}{1-\gamma^2}\right) f(u) \frac{e^{-\frac{(u-\gamma\nu_c)^2}{2(1-\gamma^2)}}}{\sqrt{2\pi(1-\gamma^2)}} \\ &= -\frac{1}{\sigma_2}\left(\frac{\bar{u}-\gamma\nu_c}{1-\gamma^2}\right), \end{aligned} \tag{81}$$

where $\bar{u}$ is the mean *peak* curvature (i.e. the integral of u times the integrand of G_0, normalized by G_0 and evaluated at $\gamma\nu_c$). Eq. (81) is precisely the scale-dependent bias parameter b_{01} found in [37, 38], except for a minus sign which arises from the fact that these authors defined b_{01} as the linear bias associated with $\sigma_2 u = -\nabla^2\delta$. The biases with respect to higher powers of $\nabla^2\delta$ are obtained by generalizing Eq. (81) to

$$b_{(\nabla^2\delta)^N} = \frac{(-1)^N}{G_0(\gamma, u_*)\sigma_2^N}\frac{\partial^N G_0(\gamma,\gamma\nu_c)}{\partial u_*^N}\Big|_{u_*=\gamma\nu_c}. \quad (82)$$

In particular, we find for the PBS bias with respect $(\nabla^2\delta)^2$

$$\begin{aligned} b_{(\nabla^2\delta)^2} &= (G_0\sigma_2^2)^{-1}\int_0^\infty du\,\frac{(u-\gamma\nu_c)^2-(1-\gamma^2)}{(1-\gamma^2)^2}\,f(u) \\ &\quad\times\frac{e^{-\frac{(u-\gamma\nu_c)^2}{2(1-\gamma^2)}}}{\sqrt{2\pi(1-\gamma^2)}} \\ &= \frac{1}{\sigma_2^2}\left[\frac{\langle(u-\gamma\nu_c)^2\rangle_{\rm pk}}{(1-\gamma^2)^2}-\frac{1}{1-\gamma^2}\right], \end{aligned} \quad (83)$$

in agreement with b_{02} as derived in [41]. Note that by construction, the peak model does not predict any biasing with respect to higher than second derivatives of the density field ($b_{\nabla^{2n}\delta} = 0$ for $n > 1$).

Desjacques et al. [38] were also able to derive the scale-dependent bias $b_{01} = -b_{\nabla^2\delta}$ from a peak-background split calculation. However, they employed a conditional mass function, i.e. the number density of peaks given a (spherical) overdensity on a much larger scale $R_B \gg R_*$, whereas here we derived all PBS bias parameters from the unconditional mass function (this is also the approach taken by [41]). Nevertheless, all these treatments are based on a similar reasoning, i.e. a long-wavelength perturbation shifts the mean curvature (in the case of [38], the shift is correlated with the density), and lead to the same final result. Note that the peak-background split approach can be generalized to derive full expressions for the peak correlation functions [41], which also includes dependencies on more general rotational invariants, such as $(\vec{\nabla}\delta)^2$ and $[(\partial_i\partial_j - \delta_{ij}\nabla^2/3)\delta]^2$. These are present in the clustering of peaks, even though it is not necessary to introduce them in order to cure the R_L-dependence of ξ_h induced by smoothing, as we have seen here and in App. B.

It is important to note that the peak model is fundamentally different from local bias expansions in the sense that it only involves a physical smoothing scale R_*, and no further coarse-graining. Nevertheless, the term $\propto b_{01}\nabla^2\delta$ in $\xi_{\rm pk}$ is equivalent to the generic $b_{\nabla^2\delta}$ term in ξ_h. Namely, $b_{01}\nabla^2\delta$ restores the contrast of the baryon acoustic oscillation, otherwise smeared out by the filtering in $b_{10}^2\xi_*$ [37]. Moreover, in general b_{01} can be greater than $\sim R_*^2$ so that the contrast of the BAO in the peak correlation $\xi_{\rm pk}(r)$ can even be enhanced relative to that in the *unsmoothed* mass correlation $\xi(r)$ (see Fig. 5 in [37]). However, subsequent gravitational evolution suppresses most of this scale-dependence (expected to be at the few percent level at the time of collapse [38]). We expect this to be a general feature of a $\nabla^2\delta$ bias specified in Lagrangian space.

IV. NON-GAUSSIAN CASE

We now return to the case of a tracer which can be sufficiently well described by density bias, i.e. we neglect the curvature bias corrections, but consider the case of initial conditions that are non-Gaussian. Specifically, we will first focus on the case of local primordial non-Gaussianity, which has been shown to lead to a large modification of clustering on large scales; general shapes of non-Gaussianity will be considered in Sec. IV D.

The derivation in Sec. II can be generalized to a general non-Gaussian density field, in which Eq. (36) formally retains its validity (see App. A). However, one can easily show that in general the resulting tracer correlation function depends on the coarse-graining scale R_L. At lowest order the tracer auto-correlation becomes

$$\xi_h(r) = b_1^2\xi_L(r) + b_1 b_2\langle\delta_L(1)\delta_L^2(2)\rangle + \mathcal{O}(\delta_L^4), \quad (84)$$

the second term being the leading non-Gaussian correction. For primordial non-Gaussianity of the local type, and in the limit $r \gg R_L$, the second correlator is given by (see Eq. (97))

$$\langle\delta_L(1)\delta_L^2(2)\rangle = 4f_{\rm NL}\sigma_L^2\xi_{\phi\delta}(r), \quad (85)$$

where $\xi_{\phi\delta}$ is the cross-correlation between the matter and primordial Bardeen potential ϕ. Given the appearance of σ_L^2, Eq. (84) is strongly R_L-dependent. This indicates that the description of the tracer density as a function of the matter density δ_L alone is insufficient even on large scales in the non-Gaussian case.

Instead, we need to include a dependence of the tracer density on the amplitude of small-scale fluctuations. This dependence is present regardless of the nature of the initial conditions; however, only in the non-Gaussian case are there large-scale modulations of the small-scale fluctuations, due to mode coupling, whereas in the Gaussian case we were able to neglect the small-scale fluctuations in the large-scale description. In general, one would imagine that the abundance of tracers depends on the amplitude of small-scale fluctuations on a range of scales. However, for simplicity we will parametrize the dependence through the variance of the density field on a single scale R_*. In the local model, which we will focus on here, this is sufficient in any case as all small-scale fluctuations are rescaled equally (in the large-scale limit), so that the value of R_* becomes irrelevant for the final result.

While we focus on primordial non-Gaussianity of the local type here, the extension to other types of non-Gaussianity is straightforward (see Sec. IV D). Furthermore, we only rely on the description of the density field

in terms of N-point functions, with the 3-point function being the lowest order non-Gaussian contribution which we focus on here. That is, we do not rely on a fictitious Gaussian field from which the non-Gaussian field is constructed. This is different than the approach taken in [16, 42, 43], where the separation of scales is typically applied in the fictitious Gaussian field, and an application in the physical non-Gaussian potential is not straightforward to implement [16].

We first define the small-scale density field as the local fluctuations around the coarse-grained field δ_L:

$$\begin{aligned}\delta_s(\mathbf{x}) &\equiv \delta_*(\mathbf{x}) - \delta_L(\mathbf{x}) \qquad (86)\\ &= \int d^3\mathbf{y}[W_*(\mathbf{x}-\mathbf{y}) - W_L(\mathbf{x}-\mathbf{y})]\delta(\mathbf{y})\\ &= \int \frac{d^3\mathbf{k}}{(2\pi)^3}\tilde{W}_s(k)\delta(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}},\\ \tilde{W}_s(k) &= \tilde{W}_*(k) - \tilde{W}_L(k). \qquad (87)\end{aligned}$$

In Fig. 1, δ_s is illustrated by the thin black line in the lower part of the figure. Note that as $k \to 0$, $\tilde{W}_s(k) \propto k^2$, i.e. the long-wavelength modes are filtered out as desired. This implies that the cross-correlation between $\delta_s(\mathbf{x})$ and the density field $\delta_R(\mathbf{x})$ smoothed on some scale R goes to zero as R becomes much larger than R_L:

$$\begin{aligned}\langle\delta_s(\mathbf{x})\delta_R(\mathbf{x})\rangle &= \int \frac{d^3k}{(2\pi)^3}\tilde{W}_s(k)\tilde{W}_R(k)P(k)\\ &\overset{R\gg R_L}{\to} 0, \qquad (88)\end{aligned}$$

and similarly for $\langle\delta_s(\mathbf{x}_1)\delta_L(\mathbf{x}_2)\rangle$ if $|\mathbf{x}_1 - \mathbf{x}_2| \gg R_L$. We further use the notation

$$\sigma_s^2 \equiv \langle\delta_s^2\rangle = \int \frac{d^3k}{(2\pi)^3}|\tilde{W}_s(k)|^2P(k). \qquad (89)$$

We quantify the dependence of the tracer abundance on the amplitude of small-scale fluctuations through

$$y_*(\mathbf{x}) \equiv \frac{1}{2}\left(\frac{\delta_s^2(\mathbf{x})}{\sigma_s^2} - 1\right), \qquad (90)$$

where the subscript $*$ refers to the smoothing scale R_*, $\langle y_*\rangle = 0$, and the factor of $1/2$ is included to obtain expressions which conform to standard convention later on. In the Gaussian case, $\xi_s(r) \to 0$ for $r \gg R_L$, so that the small-scale density field and y_* in particular have no large-scale correlations. In the non-Gaussian case however, y_* is in general correlated with long-wavelength perturbations. Note that $\langle\delta_s(1)\delta_L(2)\rangle$ vanishes by construction on large scales [Eq. (88)], so that it is natural to start the expansion with the leading term δ_s^2.

We now generalize Eq. (6) to explicitly include the dependence on y_*,

$$\hat{n}_h(\mathbf{x}) = F_{h,L}(\delta_L(\mathbf{x}), y_*(\mathbf{x}); \mathbf{x}). \qquad (91)$$

Although our approach here is formally similar to the bivariate local expansion in δ_L and ϕ_L adopted in [18, 26, 43], there is somewhat of a conceptual difference in that we expand $\hat{n}_h$ purely in terms of properties of the matter distribution. The effect of non-Gaussianity, and the fact that it derives from a potential ϕ, only enter through the expressions for the correlators between δ_L and y_* here. The nature of non-Gaussianity thus decouples from the description of the tracers (which only know about the matter density field) in this approach.

We can now repeat the derivation of Sec. II, including this additional dependence. All arguments about the residual scatter from the deterministic relation $\hat{n}_h(\mathbf{x}) = \hat{n}_h[\delta_L(\mathbf{x}), y_*(\mathbf{x})]$ and its negligible correlation with long-wavelength perturbations made in Sec. II also apply here. In fact, the dependence of $\hat{n}_h(\mathbf{x})$ on $y_*(\mathbf{x})$ is a source of uncorrelated scatter in the Gaussian case which becomes correlated with long-wavelength perturbations in the non-Gaussian case. This is another way of seeing why we need to introduce the dependence on y_* explicitly when dealing with large-scale non-Gaussianity. Taking the expectation value of Eq. (91), we obtain

$$\langle\hat{n}_h\rangle = \langle F_{h,L}(0)\rangle\sum_{n,m}\frac{c_{nm}}{n!m!}\langle\delta_L^n y_*^m\rangle, \qquad (92)$$

where we have defined bivariate "bare" bias parameters through

$$c_{nm} \equiv \frac{1}{\langle F_{h,L}(0)\rangle}\left\langle\frac{\partial^{n+m}F_{h,L}}{\partial\delta_L^n\partial y_*^m}\Big|_{\delta_L=0,y_*=0}\right\rangle. \qquad (93)$$

We then need expressions for the various cross-correlations of δ_L and y_*. In the following, we will restrict ourselves to the leading order terms, as the general expansion becomes lengthy.

A. Primordial non-Gaussianity of the local type

We will consider a density field derived from a Bardeen potential with non-Gaussianity of the local type. We will restrict our treatment to leading order in the non-linearity parameter $f_{\rm NL}$. At this order, the only relevant N-point function is the bispectrum,

$$\begin{aligned}B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) &= \mathcal{M}(k_1)\mathcal{M}(k_2)\mathcal{M}(k_3)B_\phi(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\\ B_\phi(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) &= 2f_{\rm NL}[P_\phi(k_1)P_\phi(k_2) + (2\text{ cyclic})]. \qquad (94)\end{aligned}$$

Here,

$$\mathcal{M}(k) = \frac{2}{3}\frac{k^2T(k)g(z)}{\Omega_m H_0^2(1+z)} \qquad (95)$$

is the relation in Fourier space between the density and the Bardeen potential ϕ,

$$\delta(\mathbf{k},z) = \mathcal{M}(k)\phi(\mathbf{k}), \qquad (96)$$

where $T(k)$ is the matter transfer function normalized to unity as $k \to 0$, and $g(z)$ is the linear growth rate of the

gravitational potential normalized to unity during the matter dominated epoch. Further, we define $\mathcal{M}_L(k) = \mathcal{M}(k)\tilde{W}_L(k)$, $\mathcal{M}_s(k) = \mathcal{M}(k)\tilde{W}_s(k)$, and so on. We can then derive the leading contributions in the large-scale limit. As shown in App. C,

$$\begin{aligned}\langle\delta_L(1)\delta_L^2(2)\rangle &= \int\frac{d^3k}{(2\pi)^3}e^{i\mathbf{k}\cdot\mathbf{r}}\mathcal{M}_L(k)\int\frac{d^3k_1}{(2\pi)^3}\int\frac{d^3k_2}{(2\pi)^3}\\ &\quad\times\mathcal{M}_L(k_1)\mathcal{M}_L(k_2)\langle\phi_{\mathbf{k}}\phi_{\mathbf{k}_1}\phi_{\mathbf{k}_2}\rangle\\ &= 4f_{\rm NL}\sigma_L^2\xi_{\phi\delta,L}(r),\end{aligned} \tag{97}$$

where $\xi_{\phi\delta,L}$ is the cross-correlation function between the density coarse-grained on scale R_L and the Bardeen potential ϕ, i.e.

$$\xi_{\phi\delta,L}(r) = \int\frac{d^3k}{(2\pi)^3}e^{i\mathbf{k}\cdot\mathbf{r}}\tilde{W}_L(k)\int\frac{d^3k_1}{(2\pi)^3}\langle\delta(\mathbf{k})\phi(\mathbf{k}_1)\rangle. \tag{98}$$

In deriving Eq. (97), we have expanded to lowest order in k/k_1 ("squeezed limit" of the bispectrum), with the next higher order being suppressed by $(k/k_1)^2$ in this limit. We will discuss this approximation in Sec. V.

Similarly, at leading order in $f_{\rm NL}$ (see App. C),

$$\begin{aligned}\langle\delta_L(1)y_*(2)\rangle &= \frac{1}{2}\Big\langle\delta_L(1)\frac{\delta_s^2(2)}{\sigma_s^2}\Big\rangle\\ &= \frac{1}{2\sigma_s^2}\int\frac{d^3k}{(2\pi)^3}e^{i\mathbf{k}\cdot\mathbf{r}}\mathcal{M}_L(k)\int\frac{d^3k_1}{(2\pi)^3}\int\frac{d^3k_2}{(2\pi)^3}\\ &\quad\times\mathcal{M}_s(k_1)\mathcal{M}_s(k_2)\langle\phi_{\mathbf{k}}\phi_{\mathbf{k}_1}\phi_{\mathbf{k}_2}\rangle\\ &= 2f_{\rm NL}\xi_{\phi\delta,L}(r).\end{aligned} \tag{99}$$

This result can also be derived by using the well-known property of local primordial non-Gaussianity that, in the squeezed limit, the local variance of the density field is rescaled by $\phi(\mathbf{x})$,

$$\big\langle\delta_s^2(\mathbf{x})\big\rangle\big|_{\phi(\mathbf{x})} = \sigma_s^2\left[1+4f_{\rm NL}\phi(\mathbf{x})\right]+\mathcal{O}(f_{\rm NL}^2), \tag{100}$$

and hence

$$\langle y_*(\mathbf{x})\rangle|_{\phi(\mathbf{x})} = 2f_{\rm NL}\phi(\mathbf{x})+\mathcal{O}(f_{\rm NL}^2), \tag{101}$$

which immediately leads to Eq. (99). Note that the correlators involving y_* are independent of the scale R_* for local non-Gaussianity, so that the choice of R_* is arbitrary in this case. We will see below how this changes for other types of non-Gaussianity. Note that Eq. (99) implies that y_* is of order ϕ, i.e. linear in potential perturbations — in contrast to the naive expectation that it is of order δ^2. Finally, we note that $\langle y_*(1)y_*(2)\rangle$ is $\mathcal{O}(f_{\rm NL}^2)$ and hence not included in the following.

B. Correlations

We define the estimator for the correlation function through Eq. (14) as before. The expectation value then becomes

$$\begin{aligned}\langle\hat{\xi}_h(r)\rangle &= \frac{1}{\mathcal{N}^2}\\ &\times\sum_{n,m,n',m'=0}^{\infty}\frac{c_{nm}c_{n'm'}}{n!m!n'!m'!}\langle\delta_L^n(1)y_*^m(1)\delta_L^{n'}(2)y_*^{m'}(2)\rangle-1,\end{aligned} \tag{102}$$

where '1' and '2' stand for two arbitrary locations separated by a distance r, and we have redefined

$$\mathcal{N}\equiv\sum_{n,m=0}^{\infty}\frac{c_{nm}}{n!m!}\langle\delta_L^n y_*^m\rangle. \tag{103}$$

Similarly, we obtain the expectation value of the tracer-matter cross-correlation,

$$\langle\hat{\xi}_{hm}(r)\rangle = \frac{1}{\mathcal{N}}\sum_{n,m;n+m>0}^{\infty}\frac{c_{nm}}{n!m!}\langle\delta_L^n(1)y_*^m(1)\delta_L(2)\rangle. \tag{104}$$

Again, these expressions involve the "bare" bias parameters c_{nm}, and the mixed moments of δ_L, y_* which contain disconnected pieces. In the Gaussian case, Eq. (88) implies the absence of any connected correlators involving δ_L and y_*. The powers of y_* then only add zero-lag pieces to the previous result Eq. (17), which are absorbed by corresponding terms in $\mathcal{N}$. Thus, the final result Eq. (37) does not change in the Gaussian case if we include the dependence on y_*.

Defining for convenience

$$f(\mathbf{x}) = \sum_{n,m=0}^{\infty}\frac{c_{nm}}{n!m!}\delta_L(\mathbf{x})y_*(\mathbf{x})-1, \tag{105}$$

we have

$$\xi_h(r) = \frac{1}{\mathcal{N}^2}\left[\langle f(1)f(2)\rangle-\langle f\rangle^2\right], \tag{106}$$

where $\langle f\rangle=\mathcal{O}(\delta^2)$. In the following, we will expand ξ_h to order δ^4, and simultaneously to linear order in $f_{\rm NL}$ (as long as there are no other sources of non-Gaussianity, going to $\mathcal{O}(\delta^4)$ is also sufficient to retain *all* terms linear in $f_{\rm NL}$). Through the latter restriction, we avoid a large number of quadratic and higher order terms in y_*. We have

$$\mathcal{N} = 1+\langle f\rangle = 1+\frac{c_{20}}{2}\sigma_L^2+c_{11}\langle\delta_L y_*\rangle+\mathcal{O}(\delta^3), \tag{107}$$

and

$$\langle f\rangle^2 = \frac{c_{20}^2}{4}\sigma_L^4+c_{11}c_{20}\langle\delta_L y_*\rangle\sigma_L^2+\mathcal{O}(\delta^5). \tag{108}$$

Hence, ξ_h becomes

$$\begin{aligned}\xi_h = \frac{1}{\mathcal{N}^2}\Big\{\Big\langle &\Big(c_{10}\delta_L + c_{01}y_* + c_{11}\delta_L y_* + \frac{c_{20}}{2}\delta_L^2 + \frac{c_{30}}{6}\delta_L^3\Big)_1 \\ &\Big(c_{10}\delta_L + c_{01}y_* + c_{11}\delta_L y_* + \frac{c_{20}}{2}\delta_L^2 + \frac{c_{30}}{6}\delta_L^3\Big)_2\Big\rangle \\ &- \frac{c_{20}^2}{4}\sigma_L^4 - c_{11}c_{20}\langle\delta_L y_*\rangle\sigma_L^2\Big\} \\ = \frac{1}{\mathcal{N}^2}\Big[&c_{10}^2\langle\delta_L(1)\delta_L(2)\rangle + 2c_{10}c_{01}\langle\delta_L(1)y_*(2)\rangle \\ &+ c_{10}c_{20}\langle\delta_L(1)\delta_L^2(2)\rangle + c_{10}c_{30}\sigma_L^2\langle\delta_L(1)\delta_L(2)\rangle \\ &+ c_{01}c_{30}\sigma_L^2\langle\delta_L(1)y_*(2)\rangle \\ &+ 2c_{11}c_{20}\langle\delta_L(1)y_*(2)\rangle\langle\delta_L(1)\delta_L(2)\rangle \\ &+ 2\frac{c_{20}^2}{4}\langle\delta_L(1)\delta_L(2)\rangle^2\Big] + \mathcal{O}(\delta^5)\end{aligned} \tag{109}$$

where we have used the symmetry under interchange of locations 1 and 2, and

$$\begin{aligned}\langle y_*(1)\delta_L^3(2)\rangle &= 3\langle y_*(1)\delta_L(2)\rangle\sigma_L^2 + \mathcal{O}(f_{\rm NL}^2) \\ \langle\delta_L(1)y_*(1)\delta_L^2(2)\rangle &= 2\langle\delta_L(1)y_*(2)\rangle\langle\delta_L(1)\delta_L(2)\rangle \\ &\quad + \langle\delta_L y_*\rangle\sigma_L^2 + \mathcal{O}(f_{\rm NL}^2).\end{aligned} \tag{110}$$

Perhaps somewhat surprisingly at first, we have to keep these terms whereas terms such as $\langle\delta_L^2(1)y_*(2)\rangle$, $\langle\delta_L(1)y_*(1)\delta_L(2)\rangle$ are higher order in $f_{\rm NL}$ and thus dropped. This is simply because the latter terms do not have disconnected contributions.

Note that all completely disconnected terms, i.e. terms that asymptote to a constant as $r\to\infty$, have canceled as expected. We now use the relations derived in Sec. IV A. Using Eqs. (97)–(99), we obtain

$$\begin{aligned}\xi_h(r) = \frac{1}{\mathcal{N}^2}\Big[&\left(c_{10}^2 + c_{10}c_{30}\sigma_L^2\right)\xi_L(r) + \frac{c_{20}^2}{2}\xi_L(r)^2 \\ &+ \left(2c_{10}c_{01} + c_{01}c_{30}\sigma_L^2 + 2c_{10}c_{20}\sigma_L^2\right)2f_{\rm NL}\xi_{\phi\delta,L}(r) \\ &+ 2c_{11}c_{20}2f_{\rm NL}\xi_{\phi\delta,L}(r)\xi_L(r)\Big].\end{aligned} \tag{111}$$

C. Bivariate PBS bias parameters

In analogy to Sec. II B, we would like to introduce a physically motivated bias parameter which quantifies the response of the tracer number density to a change in the amplitude of small-scale fluctuations, without making reference to any coarse-graining on the scale R_L. The simplest way to parametrize such a dependence is to rescale all perturbations by a factor of $1+\varepsilon$ from their fiducial value, where ε is an infinitesimal parameter. For example, for a given realization of initial conditions of an N-body simulation, one can obtain a realization with a different power spectrum normalization by rescaling the initial density perturbations by $(1+\varepsilon)$.[4] Clearly, the variance of the density field on some scale R, σ_R^2, is then rescaled to $(1+\varepsilon)^2\sigma_R^2$. Note that this means that the scaled cumulants $\langle\delta_*^n\rangle_c/\sigma_*^n$ are invariant, whereas the primordial non-Gaussianity parameter $f_{\rm NL}\sim B_\Phi/P_\Phi^2$, if non-zero, scales as $(1+\varepsilon)^{-1}$ under this transformation. Specifically, under this rescaling δ_L and y_* transform as

$$\begin{aligned}\delta_L(\mathbf{x}) &\to (1+\varepsilon)\delta_L(\mathbf{x}) \\ y_*(\mathbf{x}) &\to y_*(\mathbf{x}) + \left(\varepsilon + \frac{\varepsilon^2}{2}\right)\frac{\delta_s^2(\mathbf{x})}{\sigma_s^2}\,.\end{aligned} \tag{112}$$

Note that the parameter σ_s^2 in the definition of y_* is just a constant normalization, and does not change under the ε-transformation. This is in analogy to keeping $\overline{\rho}$ fixed in the D-transformation in Sec. II B.

We can then define a set of *bivariate PBS bias parameters* b_{NM} by generalizing Eq. (31) to

$$b_{NM} \equiv \frac{1}{\langle\hat{n}_h\rangle_{D=0,\varepsilon=0}}\frac{\partial^{N+M}\langle\hat{n}_h\rangle_{D,\varepsilon}}{\partial D^N\partial\varepsilon^M}\bigg|_{D=0,\varepsilon=0}. \tag{113}$$

These parameters can be understood as follows. Given infinite volume, the average tracer number density is a deterministic function of the mean matter density $\overline{\rho}$ and the amplitude of the fluctuations (parametrized, e.g., through the RMS of the density field on some scale, σ_*). b_{NM} then denotes the $N+M$-th joint derivative of this function with respect to $\ln\overline{\rho}$ and σ_* (more precisely, ε) at some fiducial values of $\overline{\rho}$ and σ_*. Clearly, the parameters b_{NM} are independent of the coarse-graining scale R_L.

As before, our next task is to derive the relation between b_{NM} and c_{nm}. We have from Eq. (92),

$$\begin{aligned}\langle\hat{n}_h\rangle(D,\varepsilon) = \langle F_{h,L}(0)\rangle &\sum_{n,m=0}^{\infty}\frac{c_{nm}}{n!m!} \\ &\times\left\langle\left[(1+\varepsilon)\delta_L + D\right]^n\left[y_* + \left(\varepsilon+\frac{\varepsilon^2}{2}\right)\frac{\delta_s^2}{\sigma_s^2}\right]^m\right\rangle.\end{aligned} \tag{114}$$

We thus have

$$b_{N0} = b_N. \tag{115}$$

In particular,

$$b_{10} = \frac{1}{\mathcal{N}}\left(c_{10} + \frac{c_{30}}{2}\sigma_L^2 + \mathcal{O}(\delta_L^3)\right). \tag{116}$$

Further,

$$\begin{aligned}b_{01} &= \frac{1}{\mathcal{N}}\sum_{n,m}\frac{c_{nm}}{n!m!}\left(n\langle\delta_L^n y_*^m\rangle + m\langle\delta_L^n(1+2y_*)y_*^{m-1}\rangle\right) \\ &= \frac{1}{\mathcal{N}}\left(c_{01} + c_{20}\sigma_L^2 + c_{11}\langle\delta_L y_*\rangle + \frac{c_{30}}{2}\left\langle\delta_L^3\right\rangle + \mathcal{O}(\delta^4)\right).\end{aligned} \tag{117}$$

We can now express the correlation function of tracers at this order, Eq. (109), in terms of the PBS bias parameters. In fact, if we are able to reach the analogous result to the Gaussian case, i.e. that the tracer correlation

[4] Of course, if one initializes using a second-order density field, then the second order part needs to be rescaled by $(1+\varepsilon)^2$.

function is a sum over PBS bias parameters multiplying no-zero-lag correlators, we only need to keep terms up to order δ^2 in b_{NM}, since they always multiply a correlator of at least order δ^2. Note that when extending the treatment to higher order in $f_{\rm NL}$, it is necessary to take into account that y_* transforms nonlinearly with ε [Eq. (112)]. This means that the bias coefficient multiplying correlators containing say $y_*^2(1)$ will not simply be b_{N2}, but involve a linear combination of b_{N1} and b_{N2}.

Let us thus write all mixed "no-zero-lag" terms with the appropriate b_{NM} in front, at order δ^4, $f_{\rm NL}$. We obtain

$$\begin{aligned}\xi_h(r) = b_{10}^2\xi_L(r) + \frac{b_{20}^2}{2}\xi_L^2(r) + 2b_{10}b_{01}\langle\delta_L(1)y_*(2)\rangle \\ + b_{20}b_{11}\langle\delta_L(1)y_*(1)\delta_L^2(2)\rangle_{\rm nzl} \\ + \mathcal{O}(\delta^5, f_{\rm NL}^2).\end{aligned} \tag{118}$$

Here we have used the fact that at this order, $\langle\delta_L(1)\delta_L^2(2)\rangle$, $\langle\delta_L^2(1)y_*(2)\rangle$, $\langle\delta_L(1)\delta_L^3(2)\rangle$, and $\langle y_*(1)\delta_L^3(2)\rangle$ have no *no-zero-lag* pieces. Note also that $\langle\delta_L(1)y_*(2)\rangle = \langle\delta_L(1)y_*(2)\rangle_{\rm nzl}$. Plugging in the expressions for b_{NM} at the relevant order, we obtain

$$\begin{aligned}\xi_h(r) = \frac{1}{\mathcal{N}^2}\Big\{ \left(c_{10}^2 + c_{10}c_{30}\sigma_L^2\right)\xi_L(r) + \frac{c_{20}^2}{2}\xi_L^2(r) \\ + 2\left(c_{10}c_{01} + \frac{c_{30}}{2}c_{01}\sigma_L^2 + c_{10}c_{20}\sigma_L^2\right)2f_{\rm NL}\xi_{\phi\delta,L}(r) \\ + 2c_{20}c_{11}2f_{\rm NL}\xi_{\phi\delta,L}(r)\xi_L(r) \\ \Big\} + \mathcal{O}(\delta^5, f_{\rm NL}^2).\end{aligned} \tag{119}$$

We easily see that this agrees identically with Eq. (111). Thus, the bivariate PBS parameters which we have defined in a coarse graining scale-independent way absorb all coarse graining-scale dependent terms in the "bare" bias parameter expansion Eq. (102), in particular the term $c_1c_2\langle\delta_L(1)\delta_L^2(2)\rangle$. We expect this to hold to any order in the bare bias parameter expansion, although a proof is beyond the scope of this paper.

Thus, the introduction of the bivariate bias parameters Eq. (113) and the resulting expression Eq. (118) achieved exactly what we had wanted. In particular, the leading effect of local primordial non-Gaussianity is quantified by b_{01}, the response of the mean number density of tracers to a rescaling of the amplitude of initial fluctuations. The term $c_1c_2\langle\delta_L(1)\delta_L^2(2)\rangle$ on the other hand is seen as an artifact of the bare bias expansion which is absorbed in the renormalized parameter b_{01}. Apart from the clear physical interpretation, this reordering of the perturbative expansion is also manifestly convergent: higher order terms are guaranteed to be suppressed by powers of $\xi_L(r)$ and $f_{\rm NL}\xi_{\phi\delta,L}(r)$, which only need to be small on the scale of observation r for the perturbative expansion to be valid.

This also remedies a worrying issue with the local bias expansion in the non-Gaussian case: evaluation of Eq. (36) shows that higher order terms ("loop corrections") become comparable to or larger than the leading order expression $b_1^2\xi_L(r)$ on sufficiently large scales, which would indicate a breakdown of the perturbative expansion. The bivariate expansion on the other hand leads to an expansion in which higher order terms are consistently suppressed [Eq. (118)], i.e. all dominating terms are actually lowest order ("tree-level"). For sufficiently large values of $f_{\rm NL}$, one might need to include higher order terms in that parameter. Nevertheless, the expansion will remain convergent.

D. Non-local non-Gaussianity

We now consider the generalization of the results of the last section to arbitrary quadratic non-Gaussianity, i.e. non-Gaussianity that is described to leading order by a 3-point function. The correlators that are relevant for the tracer two-point correlation in the non-Gaussian case, Eqs. (97)–(99), are determined by the behavior of the bispectrum in the squeezed limit, corresponding to triangle configurations where one side is much smaller than the other two. For scale-invariant bispectra, we can write the bispectrum in this limit as

$$B_\phi(\mathbf{k}_l, \mathbf{k}_s, -\mathbf{k}_l - \mathbf{k}_s) \overset{k_l \ll k_s}{=} A\left(\frac{k_l}{k_s}\right)^\alpha P_\phi(k_l)P_\phi(k_s), \tag{120}$$

with A, α being constants (more general shapes can be constructed by linear superposition of bispectra with different A_i, α_i). Local, folded, and equilateral shapes correspond to $\alpha = 0$, 1, and 2, respectively. Eqs. (97)–(99) then generalize to

$$\begin{aligned}\langle\delta_L(1)\delta_L^2(2)\rangle &= \int\frac{d^3k}{(2\pi)^3}e^{i\mathbf{k}\cdot\mathbf{r}}\mathcal{M}_L(k)\int\frac{d^3k_1}{(2\pi)^3} \\ &\qquad\times\mathcal{M}_L^2(k_1)Ak^\alpha P_\phi(k)k_1^{-\alpha}P_\phi(k_1) \\ &= A\sigma^2_{-\alpha,L}\xi_{\phi_\alpha\delta,L}(r)\end{aligned} \tag{121}$$

$$\begin{aligned}\langle\delta_L(1)y_*(2)\rangle &= \frac{1}{2}\Big\langle\delta_L(1)\frac{\delta_s^2(2)}{\sigma_s^2}\Big\rangle \\ &= \frac{1}{2\sigma_s^2}\int\frac{d^3k}{(2\pi)^3}e^{i\mathbf{k}\cdot\mathbf{r}}\mathcal{M}_L(k)\int\frac{d^3k_1}{(2\pi)^3} \\ &\qquad\times\mathcal{M}_s^2(k_1)Ak^\alpha P_\phi(k)k_1^{-\alpha}P_\phi(k_1) \\ &= A\frac{\sigma^2_{-\alpha,s}}{2\sigma_s^2}\xi_{\phi_\alpha\delta,L}(r),\end{aligned} \tag{122}$$

where we have defined the general spectral moment

$$\sigma^2_{n,X} \equiv \int\frac{d^3k}{(2\pi)^3}k^nP(k)|\tilde{W}_X(k)|^2, \tag{123}$$

and the correlation between a non-local function of ϕ and the density field,

$$\xi_{\phi_\alpha\delta,L}(r) \equiv \int\frac{d^3k}{(2\pi)^3}k^\alpha\mathcal{M}(k)P_\phi(k)\tilde{W}_L(k). \tag{124}$$

Again, Eqs. (121)–(122) are valid at leading order in the squeezed limit ($k \ll k_1$, with corrections going as

$(k/k_1)^2$). Inserting these expressions into Eq. (109), and using Eq. (110) we obtain up to $\mathcal{O}(\delta^5, f_{\rm NL}^2)$

$$\xi_h(r) = \frac{1}{\mathcal{N}^2}\Bigg[\left(c_{10}^2 + c_{10}c_{30}\sigma_L^2\right)\xi_L(r) + \frac{c_{20}^2}{2}\xi_L(r)^2 + \left(2c_{10}c_{01}\frac{\sigma^2_{-\alpha,s}}{2\sigma_s^2} + c_{01}c_{30}\frac{\sigma^2_{-\alpha,s}}{2\sigma_s^2}\sigma_L^2 + c_{10}c_{20}\sigma^2_{-\alpha,L}\right) \times A\xi_{\phi_\alpha\delta,L}(r) + 2c_{11}c_{20}\frac{\sigma^2_{-\alpha,s}}{2\sigma_s^2}A\xi_{\phi_\alpha\delta,L}(r)\xi_L(r)\Bigg]. \quad (125)$$

Inspection shows that the bivariate PBS parameters defined in Sec. IV C cannot absorb the R_L-dependent term from $\langle\delta_L(1)\delta_L^2(2)\rangle$. This goes back to the fact that in the presence of a bispectrum of the form Eq. (120), the small-scale perturbations are not rescaled uniformly, but rather in a scale-dependent way: the squeezed-limit result Eq. (100) generalizes to

$$\left\langle\delta_s^2(\mathbf{x})\right\rangle\big|_{\phi_\alpha(\mathbf{x})} = \sigma_s^2 + A\sigma^2_{-\alpha,s}\phi_\alpha(\mathbf{x}), \quad (126)$$

where

$$\phi_\alpha(\mathbf{x}) = \int\frac{d^3k}{(2\pi)^3}k^\alpha\phi(\mathbf{k})e^{i\mathbf{kx}}, \quad (127)$$

and hence

$$\langle y_*(\mathbf{x})\rangle|_{\phi_\alpha(\mathbf{x})} = A\frac{\sigma^2_{-\alpha,s}}{2\sigma_s^2}\phi_\alpha(\mathbf{x}). \quad (128)$$

Thus, the transformation of the density field following Eq. (112) is not the relevant one any more. Instead, we need to rescale the density field through

$$\delta(\mathbf{k}) \to \left(1 + \varepsilon k^{-\alpha}\right)\delta(\mathbf{k}), \quad (129)$$

so that

$$\delta_L(\mathbf{x}) \to \delta_L(\mathbf{x}) + \varepsilon\,\delta_{-\alpha,L}(\mathbf{x})$$
$$y_*(\mathbf{x}) \to y_*(\mathbf{x}) + \frac{\varepsilon}{\sigma_s^2}\delta_s(\mathbf{x})\delta_{-\alpha,s}(\mathbf{x}) + \frac{\varepsilon^2}{2\sigma_s^2}\delta^2_{-\alpha,s}(\mathbf{x})\ , \quad (130)$$

where, in analogy to Eq. (127),

$$\delta_{-\alpha,X}(\mathbf{x}) \equiv \int\frac{d^3k}{(2\pi)^3}k^{-\alpha}\tilde{W}_X(k)\delta(\mathbf{k})\,e^{i\mathbf{kx}}. \quad (131)$$

Note that $\langle\delta_X\delta_{-\alpha,X}\rangle = \sigma^2_{-\alpha,X}$, and $\langle\delta^2_{-\alpha,X}\rangle = \sigma^2_{-2\alpha,X}$. We will continue to assume that the tracer density depends on the small-scale density field only through the variance on some scale R_*, parametrized through y_*. We again define b_{NM} through Eq. (113), but with the transformation Eq. (129), so that these bivariate bias parameters will in general be different from those in Sec. IV C.

As before, our next task is to derive the relation between b_{NM} and c_{nm}. We have from Eq. (92),

$$\langle\hat{n}_h\rangle(D,\varepsilon) = \langle F_{h,L}(0)\rangle\sum_{n,m}\frac{c_{nm}}{n!m!} \times\left\langle\Big[(\delta_L + \varepsilon\,\delta_{-\alpha,L})\delta_L + D\Big]^n \times\left[y_* + \frac{\varepsilon}{\sigma_s^2}\delta_s\delta_{-\alpha,s} + \frac{\varepsilon^2}{\sigma_s^2}\delta^2_{-\alpha,s}\right]^m\right\rangle.$$

We obtain

$$b_{01} = \frac{1}{\mathcal{N}}\sum_{n,m}\frac{c_{nm}}{n!m!} \times\left(n\langle\delta_{-\alpha,L}\delta_L^{n-1}y_*^m\rangle + \frac{m}{\sigma_s^2}\langle\delta_L^n\delta_s\delta_{-\alpha,s}y_*^{m-1}\rangle\right) = \frac{1}{\mathcal{N}}\left(c_{01}\frac{\sigma^2_{-\alpha,s}}{\sigma_s^2} + c_{20}\sigma^2_{-\alpha,L} + \mathcal{O}(\delta^3, f_{\rm NL}^2)\right). \quad (132)$$

As in the case of local non-Gaussianity, we now write all mixed "no-zero-lag" correlators with the appropriate b_{NM} in front, up to $\mathcal{O}(\delta^5, f_{\rm NL}^2)$. Due to the factor of $\sigma^2_{-\alpha,s}/\sigma_s^2$ in the transformation of y_* under the scale-dependent rescaling Eq. (129) (at lowest order), we have to divide by that factor when multiplying correlators involving y_*. We obtain

$$\xi_h(r) = b_{10}^2\xi_L(r) + \frac{b_{20}^2}{2}\xi_L^2(r) + 2b_{10}b_{01}\frac{\sigma_s^2}{\sigma^2_{-\alpha,s}}\langle\delta_L(1)y_*(2)\rangle + b_{20}b_{11}\frac{\sigma_s^2}{\sigma^2_{-\alpha,s}}\langle\delta_L(1)y_*(1)\delta_L^2(2)\rangle_{\rm nzl}$$
$$= b_{10}^2\xi_L(r) + \frac{b_{20}^2}{2}\xi_L^2(r) + b_{10}b_{01}A\xi_{\phi_\alpha\delta,L}(r) + b_{20}b_{11}A\xi_{\phi_\alpha\delta,L}(r)\xi_L(r). \quad (133)$$

Note that the final result is explicitly independent of the scale R_L (as long as r is sufficiently large so that the smoothing effect on $\xi_L(r)$, $\xi_{\phi_\alpha\delta,L}(r)$ is negligible), whereas $\langle\delta_L(1)y_*(2)\rangle$ itself is not since it depends on the spectral moment σ_s^2 [Eq. (122)], which in turn depends on σ_L^2 [Eq. (89)].

Inserting the expressions for b_{NM} at the relevant order,

and using Eqs. (121)–(122), we have

$$\begin{aligned}\xi_h(r) &= \frac{1}{\mathcal{N}^2}\Big\{ \left(c_{10}^2 + c_{10}c_{30}\sigma_L^2\right)\xi_L(r) + \frac{c_{20}^2}{2}\xi_L^2(r) \\ &\quad + \left[2c_{10}c_{01} + 2c_{10}c_{20}\sigma^2_{-\alpha,L}\frac{\sigma_s^2}{\sigma^2_{-\alpha,s}} + c_{30}c_{01}\sigma_L^2\right] \\ &\quad \times \langle\delta_L(1)y_*(2)\rangle \\ &\quad + c_{20}c_{11}\langle\delta_L(1)y_*(1)\delta_L^2(2)\rangle_{\rm nzl}\Big\} \\ &= \frac{1}{\mathcal{N}^2}\Big\{ \left(c_{10}^2 + c_{10}c_{30}\sigma_L^2\right)\xi_L(r) + \frac{c_{20}^2}{2}\xi_L^2(r) \\ &\quad + \left[2c_{10}c_{01} + c_{30}c_{01}\sigma_L^2\right]\frac{\sigma^2_{-\alpha,s}}{2\sigma_s^2}A\xi_{\phi_\alpha\delta,R(r)} \\ &\quad + c_{10}c_{20}\sigma^2_{-\alpha,L}A\xi_{\phi_\alpha\delta,R(r)} \\ &\quad + 2c_{20}c_{11}\xi_L(r)\frac{\sigma^2_{-\alpha,s}}{2\sigma_s^2}A\xi_{\phi_\alpha\delta,L}(r)\Big\}. \end{aligned} \tag{134}$$

This agrees exactly with Eq. (125). The key difference of the expansion of ξ_h in terms of renormalized bias parameters in the case of non-local primordial non-Gaussianity, Eq. (133), from the corresponding result for local non-Gaussianity Eq. (118) is that the bivariate bias parameters are now defined with respect to the scale-dependent rescaling of the density field, Eq. (129), rather than a scale-independent rescaling. We find that it is sufficient in the large-scale limit, even in the case of a non-Gaussianity of general shape, to describe the coarse-grained tracer abundance as a function of $\delta_L(\mathbf{x})$ and $y_*(\mathbf{x})$ in order to absorb the dependence on the coarse-graining scale R_L into the bivariate PBS bias parameters. However, the actual definition of the renormalized bias parameters depends on the shape of primordial non-Gaussianity, in particular the scaling with k_l/k_s in the squeezed limit.

We can thus summarize our findings regarding the effect of a primordial bispectrum on the two-point correlations of tracers (non-Gaussian scale-dependent bias) as follows:

- For *local* primordial non-Gaussianity, it is sufficient to include the dependence of the tracer density $\hat{n}_h$ on the local amplitude of small-scale fluctuations δ_s through the variance on some scale R_*. Furthermore, the scale R_* (and whether the dependence on δ_s is actually through the variance on several scales) is irrelevant, as all perturbations δ_s are rescaled uniformly.

- For *non-local separable* bispectra as in Eq. (120), it is still sufficient to parametrize the dependence of $\hat{n}_h$ on the amplitude of small-scale fluctuations through the variance on a single scale R_*. However, the value of the scale R_* now matters as y_* is modulated by an amount that depends on R_* [Eq. (128)]. In particular, if the tracer number density were to depend on the variance of δ_s on several different scales, then the PBS bias parameter b_{01} will be a linear combination of these different dependencies with relative weights controlled by α, i.e. the shape of the bispectrum.

- For *non-separable* bispectra, the renormalization approach we describe here is not able to remove the R_L-dependence in the tracer correlation function. However, such shapes can typically be well approximated by a linear superposition of separable shapes (see e.g. [44]), which then allows the renormalization to proceed as described here.

Thus, we find that in general, a given tracer will respond differently to different shapes of primordial non-Gaussianity, i.e. b_{01} (and b_{NM} in general with $M > 0$) depends on the tracer as well as the shape of the primordial bispectrum. In the following we will study this in the context of simplified models of tracers.

E. Universal mass functions

We begin with a generalization of the universal mass function discussed in Sec. II C. We write the mean abundance of tracers as

$$\bar{n}_h = \bar{n}_h\left(\bar{\rho}, \sigma_*, J_*\right), \tag{135}$$

where the Jacobian J_* is defined in Eq. (39). That is, $\bar{n}_h$ is given as a function of the mean density of the Universe and the variance of the density field smoothed on a scale R_*, as well as its derivative with respect to scale. Under the generalized rescaling Eq. (129), σ_* transforms to lowest order as

$$\sigma_* \to \sigma_*\left[1 + \varepsilon\frac{\sigma^2_{-\alpha,*}}{\sigma_*^2}\right], \tag{136}$$

while the Jacobian transforms as (see also [31])

$$\begin{aligned} J_* &\to J_* + \varepsilon\frac{\sigma^2_{-\alpha,*}}{\sigma_*^2}\left(\frac{d\ln\sigma^2_{-\alpha,*}}{d\ln R_*} - \frac{d\ln\sigma_*^2}{d\ln R_*}\right) \\ &= J_*\left[1 + 2\varepsilon\frac{\sigma^2_{-\alpha,*}}{\sigma_*^2}\left(\frac{d\ln\sigma^2_{-\alpha,*}}{d\ln\sigma_*^2} - 1\right)\right]. \end{aligned} \tag{137}$$

Here we have used $d/d\ln R_* = 2J_*\, d/d\ln\sigma_*^2$. Note that in the local case where $\alpha = 0$, the local Jacobian is not affected by long-wavelength modes. Using Eq. (135), we can then derive the leading non-Gaussian bias through Eq. (113):

$$\begin{aligned} b_{01} &= \frac{1}{\bar{n}_h}\left(\frac{\partial\bar{n}_h}{\partial\ln\sigma_*}\frac{\partial\ln\sigma_*}{\partial\varepsilon} + \frac{\partial\bar{n}_h}{\partial\ln J_*}\frac{\partial\ln J_*}{\partial\varepsilon}\right) \\ &= \left[\frac{1}{\bar{n}_h}\frac{\partial\bar{n}_h}{\partial\ln\sigma_*} + \frac{1}{\bar{n}_h}\frac{\partial\bar{n}_h}{\partial\ln J_*}2\left(\frac{d\ln\sigma^2_{-\alpha,*}}{d\ln\sigma_*^2} - 1\right)\right]\frac{\sigma^2_{-\alpha,*}}{\sigma_*^2} \\ &= \left[b_{01}(\alpha=0) + 2\left(\frac{d\ln\sigma^2_{-\alpha,*}}{d\ln\sigma_*^2} - 1\right)\right]\frac{\sigma^2_{-\alpha,*}}{\sigma_*^2}. \end{aligned} \tag{138}$$

Here, $b_{01}(\alpha = 0)$ is the PBS bias parameter quantifying the effect of local primordial non-Gaussianity for a tracer following Eq. (135), and we have assumed that the tracer density scales linearly with the Jacobian as expected physically. For such tracers, the bias parameters quantifying the response to general non-local non-Gaussianity (in the squeezed limit) are thus directly related to those for local non-Gaussianity. In particular, we recover the results of [31], who first pointed out the contribution by the Jacobian J_*.

We now specialize Eq. (135) to a "truly" universal mass function [Eq. (38)],

$$\bar{n}_h = \overline{\rho}\, f(\nu_c)\, J_*\,, \quad \nu_c \equiv \frac{\delta_c}{\sigma_*}\,, \tag{139}$$

where $f(\nu_c)$ is in general an arbitrary function of ν_c. The results relating $b_{01}(\alpha)$ to $b_{01}(\alpha = 0)$ of course also hold in this case. However, the specific form Eq. (139) further allows us to connect $b_{01}(\alpha = 0)$ to the linear PBS density bias:

$$\begin{aligned} b_{10} &= \frac{1}{\bar{n}_h}\frac{\partial \bar{n}_h}{\partial \ln \overline{\rho}} = -\frac{1}{\sigma_*}\frac{df}{d\nu_c} \\ b_{01}(\alpha=0) &= \frac{1}{\bar{n}_h}\frac{\partial \bar{n}_h}{\partial \ln \sigma_*} = -\frac{\delta_c}{\sigma_*}\frac{df}{d\nu_c} = \delta_c b_{10}. \end{aligned} \tag{140}$$

Note that here b_{10} is the Lagrangian bias, which is why we have not included the derivative with respect to $\ln \overline{\rho}$ of the $\overline{\rho}$ prefactor in Eq. (139) (see also Sec. II C); again, the effect on J_* vanishes for $\alpha = 0$. This is the original relation between the density bias parameter and the response to primordial non-Gaussianity derived in [13, 14, 43]. We point out that these results differ from those of [45], who considered the effect of primordial non-Gaussianity on tracers with local Lagrangian biasing. There, the entire leading order effect of primordial non-Gaussianity is encoded in a scale-dependent $c_2(\mathbf{k}_1, \mathbf{k}_2)$. Thus, a parametrization of $c_2(\mathbf{k}_1, \mathbf{k}_2)$ down to very small scales is necessary in order to predict the amplitude of the scale-dependent bias. This is in contrast to the approach presented here, where one introduces a local dependence on the small-scale fluctuations which absorbs the term proportional to c_2 into a renormalized b_{01}, which is a single number. As a result, the prediction of [45] yields a departure from Eq. (140) for universal mass functions which depends on the precise form of $f(\nu_c)$. While Eq. (140) has been both supported [13, 43, 46] and disputed [47] by simulation results, these different predictions are clearly resolvable with sufficiently large simulations. In particular, our prediction for the scale-dependent bias for a general tracer,

$$b_{01} = \frac{1}{\langle \hat{n}_h \rangle}\frac{\partial \langle \hat{n}_h \rangle}{\partial \varepsilon}\,, \tag{141}$$

which is independent of any assumptions on the mass function of the tracer, provides a rigorous test of our approach which can be applied to simulations.

V. SUMMARY & DISCUSSION

We have shown that the expression of tracer correlations in terms of R_L-independent renormalized bias parameters b_N absorbs all zero-lag correlators present in the expansion of the tracer correlation function in terms of the bare "scatter-plot" bias parameters c_n. We have shown this to all orders for an arbitrary density field. While the proof only applies directly for the auto- and cross-correlation functions and pure density biasing, we expect the result to hold in the case of higher N-point functions and multivariate biasing as well (analogously to the resummed multipoint propagators of [10, 11]). Our key result is a rigorous definition of the renormalized bias parameters in terms of derivatives of the mean number density of tracers with respect to the background density (we call these "peak-background split" bias parameters since their definition is closely related to the commonly adopted definition of PBS biases [33, 48, 49]). It is important to stress that this exact definition is entirely independent of the nature of the tracer considered. Therefore, it provides a rigorous framework in which further assumptions for or modeling of the bias parameters, for example from the excursion set, peak model, or halo occupation distribution, can be embedded.

Our results go beyond previous work on renormalized bias parameters [8] in two ways: first, we show that our result is valid to all orders; second, we rigorously connect the renormalized bias parameters with the peak-background split. We further show that the renormalized bias parameters in the tracer auto-correlation and the tracer-matter cross-correlation agree to all orders. We also expect this to be the case for higher N-point functions, although this remains to be shown. We can summarize this reasoning as in line (a) of Tab. I: the expression of tracer correlations in terms of the bare biases c_n is R_L-dependent at each order due to disconnected correlators (for example $c_2^2\sigma_L^2$). This R_L-dependence is then resummed into R_L-independent bias parameters b_N which are defined with respect to a uniform increase in the matter density.

The underlying assumption in this result is that the clustering of tracers is entirely determined by their dependence on the local matter density. This is not expected to be a good assumption in general. However, our result provides another invaluable tool: whenever the renormalized expression in terms of no-zero-lag correlators exhibits a residual dependence on R_L, we conclude that a biasing purely in terms of matter density is not sufficient.

We first encounter this in the case of the smoothed matter correlation function $\xi_L(r)$, which depends on R_L if $\xi(r)$ has structure on scales smaller than R_L (line (b) in Tab. I). In this case, we are led to introduce bias parameters with respect to the curvature (Laplacian) of the matter density field. In Fourier space, this corresponds to a scale-dependent bias $\propto k^2$. If we further include bias parameters with respect to higher derivatives of the

density field, we can in fact *entirely* absorb the effect of smoothing on $\xi(r)$ [App. B]. Again, this is regardless of the nature of the tracer and the shape of the matter correlation function. Of course, for a smooth correlation function, it is usually sufficient to keep only terms involving the lowest few derivatives of the density field. The renormalized biases with respect to the curvature correspond to derivatives of the mean tracer abundance with respect to a constant shift in the curvature of the density field (Tab. I). As an example, these bias parameters are easily derived for peaks of a Gaussian density field from the results of [40]. We show that the bias parameters obtained in this way indeed match the scale-dependent biases derived in the full, direct calculation of peak correlations [37]. In this context, it is important to point out that the curing of R_L-dependencies, such as that from a smoothing of the correlation function, is a sufficient condition for having to introduce an additional dependence of the tracer density on properties of the matter density field. However, it is not a necessary condition—specific tracers might also exhibit additional dependencies not required by renormalization. One example is peaks of the matter density field, which also exhibit a dependence on quantities such as $(\vec{\nabla}\delta)^2$ [41]. Of course, it is straightforward to include these additional dependencies in the formalism described here, by defining renormalized PBS bias parameters through suitable transformations of the density field.

In the case of a non-Gaussian density field, we find that the tracer correlation function for pure density biasing acquires a strong dependence on R_L if long-wavelength modes are coupled to short wavelength modes (line (c) in Tab. I). The most well known example of this kind is primordial non-Gaussianity of the local type [18]. In this case, we have to add a bias parameter with respect to the amplitude (variance) of small-scale fluctuations. The renormalization procedure then absorbs the R_L-dependent terms such as $c_1c_2\langle\delta_L(1)\delta_L^2(2)\rangle$, and the resulting bivariate bias parameters are given by the derivatives of the mean tracer density with respect to the background density and (essentially) the amplitude of the initial power spectrum—both clearly R_L-independent quantities. Effectively, we obtain an expansion closely related to that of [43], although we did not need to drop any terms or make approximations beyond the large-scale limit (which allows us to evaluate the bispectrum in the squeezed limit).

	R_L-dependence	local quantity	Transformation defining PBS bias parameter
(a)	$c_2^2\sigma_L^2\xi_L(r)$	δ_L	$\rho \to \rho + D\overline{\rho}$
(b)	$\xi_L(r)$	$\nabla^2\delta_L$	$\nabla^2\delta \to \nabla^2\delta + \alpha/\ell^2$
(c)	$c_1c_2\langle\delta_L(1)\delta_L^2(2)\rangle$	y_*	$\delta \to (1+\varepsilon)\delta$

TABLE I: Summary of renormalization procedures introduced to remove various dependencies of tracer correlations on the coarse-graining scale R_L. The variable y_* is defined in Eq. (90).

We also generalize the results to any form of primordial non-Gaussianity given through a bispectrum of potential perturbations. In fact, this provides a good example for how this renormalization approach pays off: we obtain a fully general and exact result (in the large-scale limit), in which the renormalized scale-dependent bias parameter depends on the precise shape of the non-Gaussianity as well as the nature of the tracer. Assuming that the tracer abundance only depends on the variance of the small-scale density field on a single scale, we can however relate the scale-dependent bias parameter for an arbitrary general shape to that for local non-Gaussianity. Further, we can be more restrictive and assume that the tracer follows a universal mass function. In this case, we can relate the scale-dependent bias parameter to the bias parameter with respect to density (as in [13, 14, 43]).

The general procedure also carries over to primordial non-Gaussianity described by higher N-point functions. For example, a non-zero trispectrum which couples long- to short-wavelength modes will introduce a significant R_L-dependence in the tracer correlation function through the term $c_1c_3\langle\delta_L(1)\delta_L^3(2)\rangle$. In order to remedy this, we need to explicitly take into account the dependence of the tracer density on the local skewness $\langle\delta_s^3\rangle$ of the density field, which then yields a corresponding scale-dependent bias contribution (as shown in [31, 50]) which absorbs the R_L-dependent terms.

The main caveat to our results is that we have worked in Lagrangian space throughout. While we expect that the general approach will also be applicable in Eulerian space, the effect of gravitational evolution will in general introduce several further dependencies of the tracer density on the environment, for example velocity and tidal field biases [19–21]. We leave this for future work. Further, we have neglected the effects of supersonic relative motion between baryons and dark matter [51], which are potentially important for low-mass tracers at high redshifts. If relevant, this effect can be included through an additional bias with respect to the relative velocity squared [52, 53]. Note that the statistical properties of this relative velocity are very well understood.

We have also only considered observables in real space. The main obstacle in transforming to Fourier space is the issue of stochasticity in the tracer density field and its scale dependence, which contributes to correlations at all k in Fourier space (although the contributions will asymptote to a constant in the low-k limit). Thus, a well-defined model for correlations on small scales is a necessary prerequisite for a rigorous understanding of Fourier-space correlations.

Further, we have restricted the treatment here to two-point correlations of tracers. The main reason for this is simplicity; we expect no major obstacles in generalizing the results to higher N-point functions in Lagrangian space, such as the tracer bispectrum with non-Gaussian initial conditions. In order for this to be useful however, non-Gaussianities from gravitational evolution will also

have to be included [54–59].

Finally, in the case of primordial non-Gaussianity, we have only considered linear terms in $f_{\rm NL}$, and restricted to the large-scale limit where the bispectrum is evaluated at lowest order in the "squeezed-limit" expansion. The extension to higher powers of $f_{\rm NL}$ is straightforward. The second approximation captures the main effects on large scales, since the subleading term is suppressed by k_l^2/k_s^2, where $k_l \sim 1/r$ is the scale on which we measure correlations, and $k_s \sim 1/R_*$ corresponds to the small-scale fluctuations. For example, in case of local non-Gaussianity, the subleading term is expected to lead to a small approximately scale-independent bias.

These caveats notwithstanding, we hope these results provide the starting point for a rigorous treatment of biasing of general tracers in the context of cosmological perturbation theory.

Acknowledgments

VD and FS thank Marc Kamionkowski and Physics and Astronomy at Johns Hopkins University for hospitality. FS thanks Tobias Baldauf, Eiichiro Komatsu, Daan Meerburg, Roman Scoccimarro, and Svetlin Tassev for discussions. FS was supported by the National Aeronautics and Space Administration through Einstein Postdoctoral Fellowship Award Number PF2-130100 issued by the Chandra X-ray Observatory Center, which is operated by the Smithsonian Astrophysical Observatory for and on behalf of the National Aeronautics Space Administration under contract NAS8-03060. DJ is supported by DoE SC-0008108 and NASA NNX12AE86G. VD acknowledges support by the Swiss National Science Foundation.

Appendix A: Proof of Eq. (36) for a general density field

Let us denote as Π_n the set of all partitions of the set

$$\{\underbrace{1,\,1,\,\ldots,\,1}_{n\text{ times}}\}, \tag{A1}$$

where the elements of the set are considered distinguishable. We call the elements B of any given $\pi \in \Pi_n$ "blocks", with π having $|\pi|$ blocks where $|\pi|$ is the cardinality, or number of elements, of π. Clearly, $|\pi| \leq n$, and the blocks of any partition in Π_n satisfy

$$\sum_{B\in\pi} |B| = n. \tag{A2}$$

For example for $n = 4$ there are four distinct partitions with one block B_1 with $|B_1| = 1$ and one block B_2 with $|B_2| = 3$. Then, the moment for an arbitrary density field δ is given in terms of the cumulants (connected correlators) by

$$\langle \delta^n \rangle = \sum_{\pi\in\Pi_n} \prod_{B\in\pi} \langle \delta^{|B|} \rangle_c \,. \tag{A3}$$

For example, in this sum the trivial partition $\pi = \big\{\{1,\,1,\,\ldots,\,1\}\big\}$ (with a single block B with $|B| = n$) corresponds to $\langle\delta^n\rangle_c$. Note that since $\langle\delta\rangle = 0$, any partition where $|B| = 1$ for any $B \in \pi$ yields a vanishing contribution.

Similarly, let us denote as $\Pi_{n,m}$ the set of all partitions of

$$\{\underbrace{1,\,1,\,\ldots,\,1}_{n\text{ times}},\,\underbrace{2,\,2,\,\ldots,\,2}_{m\text{ times}}\}. \tag{A4}$$

We can then write

$$\begin{aligned}\langle \delta_1^n \delta_2^m \rangle &= \sum_{\rho\in\Pi_{n,m}} \prod_{B\in\rho} \Big\langle \prod_{a\in B} \delta_a \Big\rangle_c \\ &= \sum_{\rho\in\Pi_{n,m}} \prod_{B\in\rho} \langle \delta_1^{n_1(B)} \delta_2^{n_2(B)} \rangle_c \,,\end{aligned} \tag{A5}$$

where in the first line a runs over the elements of the block B, and in the second line we have defined as $n_1(B)$ the number of elements '1' in block B, and correspondingly for $n_2(B)$ (so that $n_1(B) + n_2(B) = |B|$). This simplifies the result since the cumulants are independent of the order of products of δ_1 and δ_2, and only depend on the overall power of each.

Our goal is to reorder the sum in Eq. (A5). We assign two further numbers (non-negative integers to be precise) to each $\rho \in \Pi_{n,m}$:

$$\bar{N}_1(\rho) = \sum_{B\in\rho}^{n_2(B)=0} n_1(B); \quad \bar{N}_2(\rho) = \sum_{B\in\rho}^{n_1(B)=0} n_2(B) \,. \tag{A6}$$

In other words, for a given partition ρ, $\bar{N}_1(\rho)$ counts the number of elements '1' that are in blocks that *only* contain '1', while $\bar{N}_2(\rho)$ counts the number of '2's that are in blocks only containing '2'. These numbers are of course uniquely defined for each ρ. Moreover, $\bar{N}_1$, $\bar{N}_2$ define a partition of $\Pi_{n,m}$, i.e. each $\rho \in \Pi_{n,m}$ is member of one and only one subset of $\Pi_{n,m}$ defined as containing all ρ with a specific value of $\bar{N}_1$ and $\bar{N}_2$. Equivalently, the relation $\rho \sim \sigma$ defined for any $\rho, \sigma \in \Pi_{n,m}$ through

$$\rho \sim \sigma \;\Leftrightarrow\; \bar{N}_1(\rho) = \bar{N}_1(\sigma) \wedge \bar{N}_2(\rho) = \bar{N}_2(\sigma) \tag{A7}$$

is an equivalence relation on $\Pi_{n,m}$. We can then split the sum in Eq. (A5) into sums over these disjoint subsets of $\Pi_{n,m}$:

$$\langle \delta_1^n \delta_2^m \rangle = \sum_{k=0}^{n} \sum_{l=0}^{m} \sum_{\rho\in\Pi_{n,m}}^{\bar{N}_1(\rho)=k;\,\bar{N}_2(\rho)=l} \prod_{B\in\rho} \langle \delta_1^{n_1(B)} \delta_2^{n_2(B)} \rangle_c \,. \tag{A8}$$

Consider the sum over all partitions in one of these subsets:

$$\sum_{\rho\in\Pi_{n,m}}^{\bar{N}_1(\rho)=k;\,\bar{N}_2(\rho)=l}\prod_{B\in\rho}\langle\delta_1^{n_1(B)}\delta_2^{n_2(B)}\rangle_c\,. \tag{A9}$$

A partition $\rho\in\Pi_{n,m}$ can be thought of as one specific way of distributing n black balls and m red balls into arbitrarily many (initially empty) boxes. These boxes correspond to the cumulants in Eq. (A5) (of course empty boxes are trivial, because they yield 1 in the product in Eq. (A5); boxes with only one ball lead to a zero contribution). The sum in Eq. (A9) runs over all possible ways of distributing these balls that have exactly k black balls which are in boxes with only black balls, and l red balls which are in boxes with only red balls. Correspondingly, the remaining $n-k$ black and $m-l$ red balls are in boxes with *both* black and red balls. There are $\binom{n}{k}$ ways of selecting k black balls out of n, and $\binom{m}{l}$ ways for the red balls. Given such a selection of k out of n and l out of m, the sum in Eq. (A9) thus runs over all ways of partitioning k black balls into boxes, m red balls into a different set of boxes, and finally $n-k$ black and $m-l$ red balls into a third set of boxes such that each of these boxes contains at least one black and one red. Mathematically, we can write Eq. (A9) as

$$\begin{aligned}\sum_{\rho\in\Pi_{n,m}}^{\bar{N}_1(\rho)=k;\,\bar{N}_2(\rho)=l}\prod_{B\in\rho}\langle\delta_1^{n_1(B)}\delta_2^{n_2(B)}\rangle_c &= \binom{n}{k}\left(\sum_{\pi_1\in\Pi_k}\prod_{B_1\in\pi_1}\langle\delta_1^{|B_1|}\rangle_c\right)\binom{m}{l}\left(\sum_{\pi_2\in\Pi_l}\prod_{B_2\in\pi_2}\langle\delta_2^{|B_2|}\rangle_c\right)\\ &\quad\times\sum_{\rho\in\Pi^{\rm nzl}_{n-k,m-l}}\prod_{B\in\rho}\langle\delta_1^{n_1(B)}\delta_2^{n_2(B)}\rangle_c\\ &= \binom{n}{k}\langle\delta_1^k\rangle\binom{m}{l}\langle\delta_2^l\rangle\sum_{\rho\in\Pi^{\rm nzl}_{n-k,m-l}}\prod_{B\in\rho}\langle\delta_1^{n_1(B)}\delta_2^{n_2(B)}\rangle_c\,.\end{aligned} \tag{A10}$$

Here, we have used Eq. (A3), and defined the subset $\Pi^{\rm nzl}_{n,m}$ of the set of all partitions $\Pi_{n,m}$ through

$$\begin{aligned}\Pi^{\rm nzl}_{n,m} &= \Big\{\rho\in\Pi_{n,m}:\ \bar{N}_1(\rho)=0=\bar{N}_2(\rho)\Big\}\\ &= \Big\{\rho\in\Pi_{n,m}:\ \forall B\in\rho\ n_1(B)>0\wedge n_2(B)>0\Big\}.\end{aligned}$$

In other words, $\Pi^{\rm nzl}_{n,m}$ contains all partitions in which each block has at least one element '1' and at least one element '2'. That is, for any $\rho\in\Pi^{\rm nzl}_{n,m}$, the product of correlators

$$\prod_{B\in\rho}\langle\delta_1^{n_1(B)}\delta_2^{n_2(B)}\rangle_c \tag{A11}$$

does not contain any zero-lag pieces. It is then natural to define the no-zero-lag correlator for a general density field as

$$\langle\delta_1^n\delta_2^m\rangle_{\rm nzl}\equiv\sum_{\rho\in\Pi^{\rm nzl}_{n,m}}\prod_{B\in\rho}\langle\delta_1^{n_1(B)}\delta_2^{n_2(B)}\rangle_c\,, \tag{A12}$$

which reduces to Eq. (25) for a Gaussian density field.

Finally, we can sum Eq. (A10) over all values of $\bar{N}_1(\rho)$, $\bar{N}_2(\rho)$ to obtain

$$\langle\delta_1^n\delta_2^m\rangle=\sum_{k=0}^{n}\sum_{l=0}^{m}\binom{n}{k}\langle\delta_1^k\rangle\binom{m}{l}\langle\delta_2^l\rangle\langle\delta_1^{n-k}\delta_2^{m-l}\rangle_{\rm nzl}\,. \tag{A13}$$

We are now ready to prove Eq. (36) for general non-Gaussian matter density fields. Plugging the relation between b_N and c_n, Eq. (33) into Eq. (36), and relabeling $N\to n-N$, $M\to m-M$, yields

$$\begin{aligned}\langle\hat{\xi}_h(r)\rangle &= \frac{1}{\mathcal{N}^2}\sum_{n,m=0}^{\infty}\frac{c_nc_m}{n!m!}\sum_{N=0}^{n-1}\sum_{M=0}^{m-1}\binom{n}{N}\binom{m}{M}\\ &\qquad\times\langle\delta_L^N\rangle\langle\delta_L^M\rangle\langle\delta_L^{n-N}(1)\delta_L^{m-M}(2)\rangle_{\rm nzl}\\ &= \frac{1}{\mathcal{N}^2}\sum_{n,m=0}^{\infty}\frac{c_nc_m}{n!m!}\left[\langle\delta_l^n(1)\delta_l^m(2)\rangle-\langle\delta_L^n\rangle\langle\delta_L^m\rangle\right]\\ &= \frac{1}{\mathcal{N}^2}\sum_{n,m=0}^{\infty}\frac{c_nc_m}{n!m!}\langle\delta_l^n(1)\delta_l^m(2)\rangle-1.\end{aligned} \tag{A14}$$

In the second line, we have used Eq. (A13) and subtracted out the completely disconnected contribution $\langle\delta_L^n\rangle\langle\delta_L^m\rangle$ (which is not included in Eq. (36) as we start the sum from $N=1$, $M=1$). In the third line, we have used Eq. (22). This is identical to Eq. (17), and thus proves the relation Eq. (36) between PBS bias parameters and two-point correlations for a general non-Gaussian density field at all orders.

Appendix B: Curvature bias to higher order

Let us consider a general spherically symmetric filter function,

$$W_L(\mathbf{x}) = \frac{1}{4\pi R_L^3}\, f\left(\frac{|\mathbf{x}|}{R_L}\right), \tag{B1}$$

where $f(y)$ is a dimensionless function defined on $[0,\infty)$. We have pulled out a factor of $(4\pi R_L^3)^{-1}$ for convenience. The Fourier transform of $W_L(\mathbf{x})$ can then be written as

$$\begin{aligned}\tilde{W}_L(k) &= \sum_{n=0}^{\infty}(kR_L)^{2n}\frac{(-1)^n}{(2n+1)!}f_n \\ f_n &\equiv \int_0^\infty dy\, y^{2n+2} f(y).\end{aligned} \tag{B2}$$

The normalization constraint $\int d^3\mathbf{x}\, W_L(\mathbf{x}) = 1$ is equivalent to $f_0 = 1$. Since R_L is an arbitrary parameter, we can further choose one of the f_n with $n \geq 1$ to assume some desired value. Specifically, in Sec. III we have chosen R_L so that $f_1 = 6$ and hence $W(k) = 1 - k^2R_L^2 + \cdots$.

The smoothed correlation function can then be written *exactly* as

$$\xi_L(r) = \sum_{n,m=0}^{\infty}\frac{f_n f_m R_L^{2(n+m)}}{(2n+1)!(2m+1)!}\nabla^{2(n+m)}\xi(r). \tag{B3}$$

Note that the factor of $(-1)^{n+m}$ from the expansion of $\tilde{W}_L$ cancels with the $i^{2(n+m)}$ from converting powers of k into derivatives.

We can generalize the transformation Eq. (61) as follows:

$$\delta_\alpha(\mathbf{y}) = \delta(\mathbf{y}) + \alpha\sum_{N=0}^{\infty}\frac{g_N}{\ell^{2N}}|\mathbf{y}|^{2N}, \tag{B4}$$

where α and g_N are dimensionless parameters, and the case $N = 0$ corresponds to the transformation used to derive the PBS density bias (with $g_0 = 1$ and $\alpha = D$, Sec. II B). Using the expansion of the filter function, we obtain

$$\delta_{L,\alpha}(\mathbf{0}) = \delta_L(\mathbf{0}) + \alpha\sum_{N=0}^{\infty}\left(\frac{R_L}{\ell}\right)^{2N} g_N f_N. \tag{B5}$$

Further, using that

$$\begin{aligned}\nabla^2|\mathbf{y}|^{2N} &= (2N+1)2N|\mathbf{y}|^{2N-2} \\ \Rightarrow \nabla^{2n}|\mathbf{y}|^{2N} &= \frac{(2N+1)!}{(2N-2n+1)!}|\mathbf{y}|^{2(N-n)},\end{aligned} \tag{B6}$$

for $n \leq N$, we obtain in analogy to Eq. (B3)

$$\begin{aligned}&\nabla^{2n}\delta_{L,\alpha}(\mathbf{0}) = \nabla^{2n}\delta_L(\mathbf{0}) \\ &+ \frac{\alpha}{\ell^{2n}}\sum_{N=n}^{\infty}\left(\frac{R_L}{\ell}\right)^{2(N-n)}\frac{(2N+1)!}{(2N-2n+1)!}g_N f_{N-n}.\end{aligned} \tag{B7}$$

The transformation Eq. (63) corresponds to $g_1 = 1/6$ with all other g_n equal to zero. Hence,

$$\begin{aligned}\nabla^2\delta_{L,\alpha}(\mathbf{0}) &= \nabla^2\delta_L(\mathbf{0}) + \frac{\alpha}{\ell^2}3!\frac{1}{6} \\ &= \nabla^2\delta_L(\mathbf{0}) + \frac{\alpha}{\ell^2}\end{aligned} \tag{B8}$$

as intended (with all higher derivatives being unaffected). More generally, we can define $g_N = 1/(2N+1)!$ for a fixed N, with all other $g_n = 0$, so that

$$\begin{aligned}&\nabla^{2n}\delta_{L,\alpha^{(N)}}(\mathbf{0}) = \nabla^{2n}\delta_L(\mathbf{0}) \\ &+ \frac{\alpha^{(N)}}{\ell^{2n}}\left(\frac{R_L}{\ell}\right)^{2(N-n)}\frac{1}{(2N-2n+1)!}f_{N-n},\end{aligned} \tag{B9}$$

where $n \leq N$. In particular,

$$\begin{aligned}\delta_{L,\alpha^{(N)}}(\mathbf{0}) &= \delta_L(\mathbf{0}) + \alpha^{(N)}\left(\frac{R_L}{\ell}\right)^{2N}\frac{1}{(2N+1)!}f_N \\ \nabla^{2N}\delta_{L,\alpha^{(N)}}(\mathbf{0}) &= \nabla^{2N}\delta_L(\mathbf{0}) + \frac{\alpha^{(N)}}{\ell^{2N}}.\end{aligned} \tag{B10}$$

Now, if we write

$$\hat{n}_h(\mathbf{x}) = F_{h,L}\left(\left\{\nabla^{2n}\delta_L(\mathbf{x})\right\}_{n=0}^{\infty};\mathbf{x}\right), \tag{B11}$$

we can define the generalized, bare linear curvature bias parameters (at linear order in $\nabla^{2n}\delta$) through

$$c_{\nabla^{2n}\delta} \equiv \frac{1}{F_{h,L}(0)}\frac{\partial F_{h,L}\left(\left\{\nabla^{2m}\delta_L(\mathbf{x})\right\}_{m=0}^{\infty};\mathbf{x}\right)}{\partial(\nabla^{2n}\delta)}\Big|_0. \tag{B12}$$

Further, we can define PBS bias parameters as

$$b_{\nabla^{2N}\delta} = \frac{\ell^{2N}}{\langle\hat{n}_h\rangle}\frac{\partial\langle\hat{n}_h\rangle_{\alpha^{(N)}}}{\partial\alpha^{(N)}}\Big|_{\alpha^{(N)}=0} \tag{B13}$$

$$\begin{aligned}&= \ell^{2N}\sum_{n=0}^{N}c_{\nabla^{2n}\delta}\frac{\partial\nabla^{2n}\delta_{L,\alpha^{(N)}}}{\partial\alpha^{(N)}} \\ &= \sum_{n=0}^{N}c_{\nabla^{2n}\delta}\frac{R_L^{2(N-n)}}{(2N-2n+1)!}f_{N-n},\end{aligned} \tag{B14}$$

where we have taken out powers of ℓ to make the expression for the correlation function below simpler. Note that the usual PBS density bias is obtained as special case for $N = 0$.

In terms of the bare bias parameters, and to linear order in matter correlations, we can write the tracer cor-

relation function as

$$\begin{aligned}\xi_h(r) &= \sum_{n,m=0}^{\infty} c_{\nabla^{2n}\delta} c_{\nabla^{2m}\delta} \nabla^{2(n+m)} \xi_L(r) \\ &= \sum_{n,m=0}^{\infty} c_{\nabla^{2n}\delta} c_{\nabla^{2m}\delta} \\ &\quad \times \sum_{\tilde{N},\tilde{M}=0}^{\infty} \frac{f_{\tilde{N}} f_{\tilde{M}} R_L^{2(\tilde{N}+\tilde{M})}}{(2\tilde{N}+1)!(2\tilde{M}+1)!} \nabla^{2(\tilde{N}+\tilde{M}+n+m)} \xi(r) \\ &= \sum_{N,M=0}^{\infty} \sum_{n,m=0}^{N,M} c_{\nabla^{2n}\delta} c_{\nabla^{2m}\delta} \qquad \text{(B15)} \\ &\quad \times \frac{f_{N-n} f_{M-m} R_L^{2(N-n+M-m)}}{(2N-2n+1)!(2M-2m+1)!} \nabla^{2(N+M)} \xi(r).\end{aligned}$$

In the second line, we have shifted the sum by defining $N \equiv \tilde{N} + n$, $M \equiv \tilde{M} + m$, while in the third line we have reordered the sum over n, m and N, M.

Now let us write down the expected result in terms of PBS bias parameters, i.e. assuming that the parameters $b_{\nabla^{2N}\delta}$ have absorbed all dependencies on R_L:

$$\xi_h(r) = \sum_{N,M=0}^{\infty} b_{\nabla^{2N}\delta} b_{\nabla^{2M}\delta} \nabla^{2(N+M)} \xi(r), \qquad \text{(B16)}$$

where we have again restricted to linear order in matter correlations. Using Eq. (B14), this yields

$$\begin{aligned}\xi_h(r) = &\sum_{N,M=0}^{\infty} \sum_{n,m=0}^{N,M} c_{\nabla^{2n}\delta} c_{\nabla^{2m}\delta} \\ &\times \frac{R_L^{2(N+M-n-m)} f_{N-n} f_{M-m}}{(2N-2n+1)!(2M-2m+1)!} \nabla^{2(N+M)} \xi(r),\end{aligned}$$

which agrees with the exact result Eq. (B15). Thus, the PBS bias parameters defined in Eq. (B13) are indeed able to completely absorb the effects of smoothing on the correlation function (at linear order). Moreover, if we assume that there is a characteristic scale ℓ describing the dependence of the tracer density on the derivatives of the density field, i.e.

$$b_{\nabla^{2N}\delta} \sim \ell^{2N}, \qquad \text{(B17)}$$

then the expression Eq. (B16) is an expansion in terms of

$$(\ell^2 \nabla^2)^N \xi(r). \qquad \text{(B18)}$$

If the matter correlation function does not have significant structure on scales below ℓ, then this quantity is progressively suppressed at higher N and Eq. (B16) is rapidly convergent.

Appendix C: Derivation of Eqs. (97)–(99)

In this appendix we derive the squeezed-limit expressions Eqs. (97)–(99). We begin with the correlator Eq. (97):

$$\begin{aligned}\langle \delta_L(1) \delta_L^2(2) \rangle = &\int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{r}} \mathcal{M}_L(k) \int \frac{d^3k_1}{(2\pi)^3} \int \frac{d^3k_2}{(2\pi)^3} \\ &\times \mathcal{M}_L(k_1) \mathcal{M}_L(k_2) \langle \phi_{\mathbf{k}} \phi_{\mathbf{k}_1} \phi_{\mathbf{k}_2} \rangle.\end{aligned}$$

Using the definition of the bispectrum,

$$\langle \phi_{\mathbf{k}} \phi_{\mathbf{k}_1} \phi_{\mathbf{k}_2} \rangle = (2\pi)^3 \delta_D(\mathbf{k} + \mathbf{k}_1 + \mathbf{k}_2) B_\phi(\mathbf{k}, \mathbf{k}_1, \mathbf{k}_2), \quad \text{(C1)}$$

we have

$$\begin{aligned}\langle \delta_L(1) \delta_L^2(2) \rangle = &\int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{r}} \mathcal{M}_L(k) \int \frac{d^3k_1}{(2\pi)^3} \mathcal{M}_L(k_1) \\ &\times \mathcal{M}_L(|\mathbf{k}+\mathbf{k}_1|) B_\phi(k, k_1, |\mathbf{k}+\mathbf{k}_1|). \qquad \text{(C2)}\end{aligned}$$

We now expand the integrand in powers of $q = k/k_1$. Further, we define $\mu = \hat{\mathbf{k}} \cdot \hat{\mathbf{k}}_1$. We then obtain for any function $F(k)$

$$\begin{aligned}F(|\mathbf{k}+\mathbf{k}_1|) &= F(k_1) \Big[1 + \left(q\mu + \frac{1}{2} q^2 [1 - 2\mu^2] \right) a(k_1) \\ &\quad + \frac{1}{2} q^2 \mu^2 b(k_1) \Big] + \mathcal{O}(q^3), \\ a &\equiv \frac{\partial \ln F(k_1)}{\partial \ln k_1} \\ b &\equiv \frac{1}{F(k_1)} \frac{\partial^2 F(k_1)}{\partial (\ln k_1)^2} \qquad \text{(C3)}\end{aligned}$$

while

$$\begin{aligned}B_\phi(k, k_1, |\mathbf{k}+\mathbf{k}_1|) &= 2 f_{\rm NL} \Big\{ P_\phi(k) \left[P_\phi(k_1) + P_\phi(|\mathbf{k}+\mathbf{k}_1|) \right] \\ &\quad + P_\phi(k_1) P_\phi(|\mathbf{k}+\mathbf{k}_1|) \Big\} \\ &= 2 f_{\rm NL} P_\phi(k) P_\phi(k_1) \left[2 + (2q\mu + [1-2\mu^2] q^2) n_\phi \right] + \mathcal{O}(q^3),\end{aligned}$$

where $n_\phi = n_s - 4$ and we have assumed a pure power-law $P_\phi(k)$ for simplicity. Inserting these expressions into Eq. (C2), we see that the terms $\propto q\mu$ vanish once the integral over μ is performed. Thus, we obtain

$$\begin{aligned}\langle \delta_L(1) \delta_L^2(2) \rangle = &4 f_{\rm NL} \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{r}} \mathcal{M}_L(k) P_\phi(k) \\ &\times \int \frac{d^3k_1}{(2\pi)^3} \mathcal{M}_L^2(k_1) P_\phi(k_1) \left[1 + \mathcal{O}(q^2) \right].\end{aligned}$$

The k_1 integral yields σ_L^2, leading to Eq. (97) with corrections suppressed in the large-scale limit by $(k/k_1)^2 \sim (kR_L)^2$. We now turn to Eq. (99). Since

$$\langle \delta_L(1) y_*(2) \rangle = \frac{1}{2\sigma_s^2} \langle \delta_L \delta_s^2(2) \rangle, \qquad \text{(C4)}$$

this just differs from Eq. (97) through the prefactor and the different filter function, and the above results immediately lead to the third line of Eq. (99).

[1] F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro, Phys. Rep. **367**, 1 (2002), arXiv:astro-ph/0112551.
[2] J. N. Fry and E. Gaztanaga, Astrophys. J. **413**, 447 (1993), arXiv:astro-ph/9302009.
[3] A. S. Szalay, Astrophys. J. **333**, 21 (1988).
[4] P. Coles, MNRAS **262**, 1065 (1993).
[5] R. J. Scherrer and D. H. Weinberg, Astrophys. J. **504**, 607 (1998), arXiv:astro-ph/9712192.
[6] M. Manera and E. Gaztañaga, MNRAS **415**, 383 (2011), 0912.0446.
[7] K. C. Chan and R. Scoccimarro, ArXiv e-prints (2012), 1204.5770.
[8] P. McDonald, Phys. Rev. D **74**, 103512 (2006), arXiv:astro-ph/0609413.
[9] D. Jeong and E. Komatsu, Astrophys. J. **691**, 569 (2009), 0805.2632.
[10] F. Bernardeau, M. Crocce, and R. Scoccimarro, Phys. Rev. D **78**, 103521 (2008), 0806.2334.
[11] T. Matsubara, Phys. Rev. D **83**, 083518 (2011), 1102.4619.
[12] P. McDonald and A. Roy, JCAP **8**, 20 (2009), 0902.0991.
[13] N. Dalal, O. Doré, D. Huterer, and A. Shirokov, Phys. Rev. D **77**, 123514 (2008), 0710.4560.
[14] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, JCAP **8**, 31 (2008), 0805.3580.
[15] H. J. Mo, Y. P. Jing, and S. D. M. White, MNRAS **284**, 189 (1997), arXiv:astro-ph/9603039.
[16] R. Scoccimarro, L. Hui, M. Manera, and K. C. Chan, Phys. Rev. D **85**, 083002 (2012), 1108.5512.
[17] M. Manera, R. K. Sheth, and R. Scoccimarro, MNRAS **402**, 589 (2010), 0906.1314.
[18] P. McDonald, Phys. Rev. D **78**, 123519 (2008), 0806.1061.
[19] V. Desjacques and R. K. Sheth, Phys. Rev. D **81**, 023526 (2010), 0909.4544.
[20] K. C. Chan, R. Scoccimarro, and R. K. Sheth, Phys. Rev. D **85**, 083509 (2012), 1201.3614.
[21] T. Baldauf, U. Seljak, V. Desjacques, and P. McDonald, ArXiv e-prints (2012), 1201.4827.
[22] D. Jeong, F. Schmidt, and C. M. Hirata, Phys. Rev. D **85**, 023504 (2012), 1107.5427.
[23] J. Yoo, A. L. Fitzpatrick, and M. Zaldarriaga, Phys. Rev. D **80**, 083514 (2009), 0907.0707.
[24] A. Challinor and A. Lewis, ArXiv e-prints (2011), 1105.5292.
[25] C. Bonvin and R. Durrer, ArXiv e-prints (2011), 1105.5280.
[26] T. Baldauf, U. Seljak, L. Senatore, and M. Zaldarriaga, ArXiv e-prints (2011), 1106.5507.
[27] A. Dekel and O. Lahav, Astrophys. J. **520**, 24 (1999), arXiv:astro-ph/9806193.
[28] N. Frusciante and R. K. Sheth, JCAP **11**, 016 (2012), 1208.0229.
[29] F. Bernardeau, M. Crocce, and E. Sefusatti, Phys. Rev. D **82**, 083507 (2010), 1006.4656.
[30] W. H. Press and P. Schechter, Astrophys. J. **187**, 425 (1974).
[31] V. Desjacques, D. Jeong, and F. Schmidt, Phys. Rev. D **84**, 063512 (2011), 1105.3628.
[32] S. Ferraro, K. M. Smith, D. Green, and D. Baumann, ArXiv e-prints (2012), 1209.2175.
[33] N. Kaiser, Astrophys. J. Lett. **284**, L9 (1984).
[34] S. Matarrese, F. Lucchin, and S. A. Bonometto, Astrophys. J. Lett. **310**, L21 (1986).
[35] L. G. Jensen and A. S. Szalay, Astrophys. J. Lett. **305**, L5 (1986).
[36] T. Matsubara, Astrophys. J. Supp. **101**, 1 (1995), arXiv:astro-ph/9501056.
[37] V. Desjacques, Phys. Rev. D **78**, 103503 (2008), 0806.0007.
[38] V. Desjacques, M. Crocce, R. Scoccimarro, and R. K. Sheth, Phys. Rev. D **82**, 103529 (2010), 1009.3449.
[39] T. Matsubara, Astrophys. J. **525**, 543 (1999), arXiv:astro-ph/9906029.
[40] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, Astrophys. J. **304**, 15 (1986).
[41] V. Desjacques, Phys. Rev. D **87**, 043505 (2013), 1211.4128.
[42] F. Schmidt and M. Kamionkowski, Phys. Rev. D **82**, 103002 (2010), 1008.0638.
[43] T. Giannantonio and C. Porciani, Phys. Rev. D **81**, 063530 (2010), 0911.0017.
[44] K. M. Smith and M. Zaldarriaga, MNRAS **417**, 2 (2011), arXiv:astro-ph/0612571.
[45] T. Matsubara, Phys. Rev. D **86**, 063518 (2012), 1206.0562.
[46] V. Desjacques, U. Seljak, and I. T. Iliev, Mon. Not. R. Astron. Soc. **396**, 85 (2009), 0811.2748.
[47] M. Grossi, L. Verde, C. Carbone, K. Dolag, E. Branchini, F. Iannuzzi, S. Matarrese, and L. Moscardini, MNRAS **398**, 321 (2009), 0902.2013.
[48] R. K. Sheth and G. Tormen, MNRAS **308**, 119 (1999), arXiv:astro-ph/9901122.
[49] S. Cole and N. Kaiser, MNRAS **237**, 1127 (1989).
[50] K. M. Smith, S. Ferraro, and M. LoVerde, JCAP **3**, 032 (2012), 1106.0503.
[51] D. Tseliakhovich and C. Hirata, Phys. Rev. D **82**, 083520 (2010), 1005.2416.
[52] N. Dalal, U.-L. Pen, and U. Seljak, JCAP **11**, 007 (2010), 1009.4704.
[53] J. Yoo, N. Dalal, and U. Seljak, JCAP **7**, 018 (2011), 1105.3732.
[54] E. Sefusatti and E. Komatsu, Phys. Rev. D **76**, 083004 (2007), 0705.0343.
[55] D. Jeong and E. Komatsu, Astrophys. J. **703**, 1230 (2009), 0904.0497.
[56] E. Sefusatti, Phys. Rev. D **80**, 123002 (2009), 0905.0717.
[57] M. Liguori, E. Sefusatti, J. R. Fergusson, and E. P. S. Shellard, Advances in Astronomy **2010**, 980523 (2010), 1001.4707.
[58] E. Sefusatti, M. Crocce, and V. Desjacques, MNRAS **425**, 2903 (2012), 1111.6966.
[59] T. Baldauf, U. Seljak, and L. Senatore, ArXiv e-prints (2010), 1011.1513.