MPA Lectures on Galaxy Clustering

Fabian Schmidt, MPA

November 26, 2013

Contents

1 Motivation

Large-scale structure (LSS) surveys which map out the positions and redshifts of millions of galaxies deliver an enormous amount of information. The clustering statistics of these galaxies are measured with an extremely high signal to noise and correspondingly small statistical uncertainty. The number of modes available to a galaxy survey of volume V roughly scales as

$$N_{\rm modes} \sim V\, k_{\rm max}^3 \sim 10^7 \left(\frac{V}{({\rm Gpc}/h)^3}\right) \left(\frac{k_{\rm max}}{0.2\, h\, {\rm Mpc}^{-1}}\right)^3 , \tag{1}$$

where $k_{\rm max}$ is roughly the inverse of the mean galaxy separation. Compare this to the number of modes accessible by the CMB,

$$N_{\rm modes,CMB} \sim \ell_{\rm max}^2 \sim 10^6 . \tag{2}$$

Thus, given that there is only one CMB sky but many Gpc^3 of volume to be mapped out by LSS, the statistical power of galaxy surveys is much greater than that of the CMB.

In order to extract cosmological information from galaxy surveys, we thus need an accurate model for how to connect the observed statistics to theoretical predictions, which are based on the statistics of the initial density fluctuations as well as their evolution under gravity. Unfortunately, the process of galaxy formation is not understood in sufficient detail to be able to connect the initial conditions with observations from first principles (in contrast to the CMB, where this can be done); even if it was possible to simulate the detailed formation of an individual galaxy, simulating a cosmological volume with millions of galaxies would very likely still be computationally infeasible.

The topic of these lectures is to describe a general framework for a perturbative description of galaxy clustering which does not make any specific assumptions about the process of galaxy formation, and is in fact applicable to *any* tracer of LSS. Furthermore, it is applicable to any form of the initial conditions, including those with primordial non-Gaussianity (a topic we will discuss in detail). The generality does come at the expense of many unknown (bias) parameters, which however have physical meaning and can be constrained for example by simulations or models of galaxy formation. Apart from the generality, this general perturbative approach has the key advantage of providing a rigorous error control, unlike heuristic approaches such as halo occupation distribution models, excursion set ansatz, and so on (in the context of this approach, such models provide predictions for bias parameters and relations between them, and are thus contained in this framework at the perturbative level). This framework is not entirely complete yet, and I will point out missing ingredients and open questions along the way.

Two further interesting topics will be discussed: how to connect the actual "intrinsic" clustering (in a sense to be defined below) with observations, which in principle involves various relativistic corrections; and how to connect the initial conditions used for predicting LSS statistics (and for N-body simulations) to the predictions of inflationary models.

2 Galaxy bias

2.1 Local biasing as an effective description

The simplest and most well-known bias expansion is the local expansion in terms of matter density [1, 2, 3]. Typically, the abundance of tracers at position $\mathbf{x}$ is written as a Taylor series in the matter density perturbation evaluated at $\mathbf{x}$. This expansion can either be performed at the time at which tracers are observed (*Eulerian local biasing*), or by following the positions of tracers back to the initial conditions (*Lagrangian biasing*). In this section, we will keep the treatment general, with in-depth discussions of the specifics of Eulerian and Lagrangian biasing deferred to later sections.

Given the complex formation processes of actual tracers such as galaxies and clusters, we clearly do not expect the abundance of tracers to be a truly local function of the total matter density. The reason why the local bias expansion is useful and relevant is because it provides the *correct effective description* on large scales.[1] Suppose that the abundance of tracers at position $\mathbf{x}$ only depends on the

[1] This statement can be violated for certain non-Gaussian density fields, see Sec. 3.

distribution (density) of matter[2] in a finite region around $\mathbf{x}$, of characteristic dimension R_*. We will call R_* the "non-locality scale". Then the statistics of the tracer on scales $r \gg R_*$ can be accurately described by local biasing, with corrections roughly suppressed by $(R_*/r)^2$. In an analogy with effective field theory (EFT), local biasing provides the low-energy effective description of the full, complicated dynamics of the formation of tracers. This description involves a finite number of coefficients, the bias parameters, which can in principle be calculated from the full theory (e.g., a detailed simulation of galaxy formation) or alternatively be determined observationally. Aside from providing a rigorous basis for the local biasing framework, the EFT-inspired approach we describe below also provides a clear physical interpretation for the renormalized bias parameters. That is, they describe the response of the tracer number density to a uniform change in the matter density; these bias parameters are historically referred to as "peak-background split" (PBS) bias parameters. More generally, one can find similar interpretations for the other non-local bias parameters we will encounter throughout as well. The number of bias parameters that are needed depends on how close to the scale R_* one wishes to model the statistics of tracers, and the desired accuracy. We will now review how this works in detail.

Consider a filter function $W_L(\mathbf{x})$ of characteristic size R_L, normalized to unity in 3D space. In EFT language, R_L provides our UV cutoff, introduced to render loop integrals finite. 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

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

and the small-scale density field as the difference between δ and δ_L:

$$\delta_s(\mathbf{x}) = \delta(\mathbf{x}) - \delta_L(\mathbf{x})\,. \tag{4}$$

We can think of $\delta_L(\mathbf{x})$ as the average density within a region $\mathcal{U}$ of size R_L centered on $\mathbf{x}$ (circles in Fig. 1). Our expressions for any observable, including statistics of tracers as well as the bias parameters, should be independent of the artificial cutoff or smoothing scale R_L.

2.2 Bare bias expansion

Consider the number density of a given tracer $n_h(\mathbf{x})$. Throughout, we assume that we are working on a fixed proper time slice, so that when varying $\mathbf{x}$ we are always dealing with tracers at a fixed evolutionary stage. In this case, $n_h(\mathbf{x})$ can be written as a general functional of the matter density field:

$$n_h(\mathbf{x}) = F_h[\delta](\mathbf{x})\,. \tag{5}$$

We now perform the coarse-graining of the matter density field, choosing the scale R_L to be much greater than the non-locality scale R_* of the functional F. Then, the functional dependence on δ in Eq. (5) becomes separates into an ordinary dependence on the variable $\delta_L(\mathbf{x})$ and a functional dependence on $\delta_s(\mathbf{y})$ around $\mathbf{x}$:

$$n_h(\mathbf{x}) = F_{h,L}(\delta_L(\mathbf{x}); \delta_s(\mathbf{y}))\,. \tag{6}$$

where $|\mathbf{x}-\mathbf{y}| \lesssim R_*$. To see this, expand Eq. (5) into a series of linear, quadratic, and higher order functionals. Denoting a volume of scale R_* around $\mathbf{x}$ as $V_*(\mathbf{x})$, any term in this expansion can be written as

$$\left(\prod_{i=1}^{n}\int_{V_*(\mathbf{x})} d^3\mathbf{y}_i\right) F^{(n)}(\mathbf{y}_1,\cdots,\mathbf{y}_n)\delta(\mathbf{y}_1)\cdots\delta(\mathbf{y}_n) = \sum_{k=0}^{n}\binom{n}{k}[\delta_L(\mathbf{x})]^k\left(\prod_{i=k+1}^{n}\int_{V_*(\mathbf{x})} d^3\mathbf{y}_i\right) F^{(n)}(\mathbf{x},\cdots,\mathbf{x},\mathbf{y}_{k+1},\cdots\mathbf{y}_n)\delta_s(\mathbf{y}_{k+1})\cdots\delta_s(\mathbf{y}_n)\,, \tag{7}$$

[2]Here we neglect the dependence of the tracer density on the matter *velocities*, which in principle are equally important. We will consider this in Sec. 2.4.2.

where we have approximated $\delta_L(\mathbf{x})$ as constant over the region $V_*(\mathbf{x})$, consistent with $R_L \gg R_*$ (we will generalize this in Sec. 2.4.1). Thus, each term in $F_h[\delta](\mathbf{x})$ becomes a local function of $\delta_L(\mathbf{x})$ multiplied by a functional of δ_s. Note that, as indicated by the notation, the function $F_{h,L}$ itself depends on R_L. The dependence of $F_{h,L}$ on δ_s is the reason for any departure, or "scatter", of the tracer number density from a deterministic relation $n_h(\mathbf{x}) = n_h[\delta_L(\mathbf{x})]$. Again, 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 [4].

The PBS argument we will apply below will allow us to derive the statistics of the tracer without any explicit knowledge of the function $F_{h,L}$. We can formally expand Eq. (6) in a Taylor series,

$$n_h(\mathbf{x}) = \sum_{n=0}^{\infty} \frac{1}{n!} F_{h,L}^{(n)}(0;\delta_s)\,[\delta_L(\mathbf{x})]^n, \tag{8}$$

where $F_{h,L}^{(n)}(0;\delta_s)$ denotes the n-derivative of $F_{h,L}(\delta_L,\delta_s)$ with respect to δ_L evaluated for a fixed realization of the small-scale density δ_s and at $\delta_L = 0$. We now take the expectation value of Eq. (8) over realizations of the density field, in order to obtain an expression for the mean number density of tracers. The assumption of negligible correlation between δ_L and δ_s implies that the two factors in each term of Eq. (8) are independent random variables (cf. the Poisson clustering model [3]):

$$\langle F_{h,L}^{(n)}(0;\delta_s)\,[\delta_L(\mathbf{x})]^n\rangle = \langle F_{h,L}^{(n)}(0;\delta_s)\rangle\,\langle[\delta_L(\mathbf{x})]^n\rangle \tag{9}$$

We then obtain

$$\langle n_h(\mathbf{x})\rangle \;= \langle F_{h,L}(0,\delta_s)\rangle\left(1 + \frac{c_2}{2}\sigma_L^2 + \frac{c_3}{6}\langle\delta_L^3\rangle + \dots\right), \tag{10}$$

where we have defined

$$c_n \equiv \frac{1}{\langle F_{h,L}(0,\delta_s)\rangle}\left\langle F_{h,L}^{(n)}(0,\delta_s)\right\rangle . \tag{11}$$

Note that the c_n are independent of location due to homogeneity, but they do depend on R_L. The bare local bias parameters c_n depend on R_L, but for clarity we will not indicate this dependence explicitly. Further,

$$\sigma_L^2 \equiv \langle\delta_L^2\rangle = \int \frac{d^3k}{(2\pi)^3}|\tilde{W}_L(k)|^2 P(k). \tag{12}$$

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 $\langle n_h\rangle = \langle F_{h,L}(0,\delta_s)\rangle$, the expectation value of the function $F_{h,L}$ at the background density. For finite values of R_L however, $\langle n_h\rangle$ receives contributions from the variance σ_L^2 and higher order moments of the density field coarse-grained with W_L. This just says that for finite regions $\mathcal{U}$, $\langle F_{h,L}(0,\delta_s)\rangle$ does not give the cosmic mean of the tracer abundance, $\langle n_h\rangle$.

We now turn to the correlation function ξ_h of the 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 above, 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. 2.4.1.

The correlation function is defined as

$$\xi_h(r) = \frac{\langle n_h(1)n_h(2)\rangle}{\langle n_h\rangle^2} - 1, \tag{13}$$

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

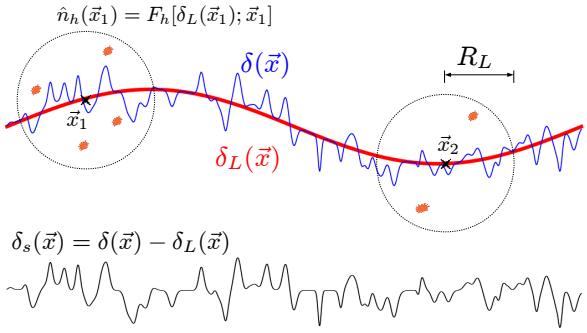

Figure 1: Sketch of the separation of the density field (blue, thin line) into large-scale part δ_L (red, thick line) and a small scale part δ_s (Eq. (46) in Sec. 3; thin black line below), via an arbitrary coarse-graining scale R_L. The tracer density coarse-grained on scale R_L (circles) is described by the function $F_{h,L}(\delta_L; \delta_s)$ [Eq. (6)], where the explicit dependence on δ_s encodes the scatter around the mean relation with δ_L, which is assumed to be uncorrelated with δ_L.

random variables. Using Eq. (8), we can then write Eq. (13) in terms of the statistics of the coarse-grained density field δ_L, and the coefficients c_n:

$$\xi_h(r) = \frac{1}{\mathcal{N}^2} \sum_{n,m=0}^{\infty} \frac{c_n c_m}{n!m!} \langle \delta_L^n(1) \delta_L^m(2) \rangle - 1 \tag{14}$$

$$\mathcal{N} = \sum_{n=0}^{\infty} \frac{c_n}{n!} \langle \delta_L^n \rangle \,. \tag{15}$$

Since the correlation function is an observable, the left hand side has to be independent of the coarse-graining scale R_L. On the other hand, the expression on the r.h.s. involves sums over moments of the density field, which contain disconnected pieces ("loops") such as $\langle \delta_L^2 \rangle$ which clearly depend on R_L, multiplied by the bare bias parameters c_n. Thus, the running of the c_n with R_L has to cancel that of the matter correlators. Further, terms of order $\langle \delta_L^n \rangle$ appear for arbitrarily high n on the r.h.s.. On the other hand, a physically reasonable perturbative bias model for $\xi_h(r)$ should converge for a scale r as long as the *connected* matter correlators, such as the correlation function

$$\xi_L(r) \equiv \langle \delta_L(1) \delta_L(2) \rangle \,, \tag{16}$$

on the scale r 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. (14) 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.

2.3 Renormalizing the bias parameters

Consider a new set of bias parameters b_N defined through

$$b_N = \frac{1}{\mathcal{N}} \sum_{n=N}^{\infty} \frac{c_n}{n!} \frac{n!}{(n-N)!} \langle \delta_L^{n-N} \rangle. \tag{17}$$

As shown in [5], the tracer correlation function [Eq. (14)] expressed in terms of the b_N becomes

$$\xi_h(r) = \sum_{N,M=1}^{\infty} \frac{b_N}{N!} \frac{b_M}{M!} \langle \delta_L^N(1) \delta_L^M(2) \rangle_{\rm nzl} \,. \tag{18}$$

Here, $\langle\cdot\rangle_{\rm nzl}$ 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 App. A in [5] for a mathematical expression of this definition). Eq. (18) also generalizes to the correlation function of two different tracers a and b:

$$\xi_{ab}(r) = \sum_{N,M=1}^{\infty} \frac{b_N^{(a)}}{N!}\frac{b_M^{(b)}}{M!}\langle\delta_L^N(1)\delta_L^M(2)\rangle_{\rm nzl}\,, \tag{19}$$

which includes the tracer-matter cross-correlation function as a special case with $b_N^{(b)} = \delta_{N1}$. Note that there is no b_0; the expressions Eqs. (18)–(19) only involve terms with b_N for $N \geq 1$. In the Gaussian case, Eq. (19) further simplifies to

$$\xi_{ab}(r) = \sum_{N=1}^{\infty} \frac{b_N^{(a)} b_N^{(b)}}{N!}\left[\xi_L(r)\right]^N. \tag{20}$$

Eqs. (18)–(20) 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). Thus, this expansion converges if and only if the finite-lag matter correlators are small on the scale r, which is what we expect from a physical bias expansion.

Finally, it is worth noting that Eqs. (18)–(19) are valid for a general non-Gaussian density field. However, as we will discuss in Sec. 3, a local bias description is in general not sufficient for non-Gaussian density fields where small-scale fluctuations couple to long-wavelength modes.

We now turn to the physical interpretation of the b_N introduced in Eq. (17), which makes use of what has historically been referred to as the peak-background split argument. 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 $\langle n_h\rangle$ in a fictitious Universe with modified background density

$$\overline{\rho}' = \overline{\rho}(1+D)\,, \tag{21}$$

where $\overline{\rho}$ is the actual background density. Thus, we now consider the case where we add a uniform density $D\overline{\rho}$ to the matter density, where D is an infinitesimal parameter. Then, in a region with overdensity δ_L the matter density is perturbed to

$$\rho_L = \overline{\rho}(1+\delta_L) \to \overline{\rho}(1+\delta_L) + \Delta\overline{\rho} = \overline{\rho}(1+\delta_L+D) \tag{22}$$

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 $\langle n_h\rangle$ (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. (10):

$$\langle n_h\rangle|_D = \langle F_{h,L}(0,\delta_s)\rangle \sum_{n=0}^{\infty}\frac{c_n}{n!}\langle(\delta_L+D)^n\rangle, \tag{23}$$

where $F_{h,L}$ and c_n both refer to the Universe with background density $\overline{\rho}$, i.e. $D = 0$. We can now define the *peak-background split bias parameters* b_N ($N \geq 1$) as the derivative of $\langle n_h\rangle$ with respect to D:

$$b_N \equiv \frac{1}{\langle n_h\rangle|_{D=0}}\frac{\partial^N\langle n_h\rangle|_D}{\partial D^N}\Big|_{D=0}. \tag{24}$$

Using Eq. (21), we can also write this as

$$b_N = \frac{\overline{\rho}^N}{\langle n_h\rangle}\frac{\partial^N\langle n_h\rangle}{\partial\overline{\rho}^N}\,, \tag{25}$$

where the derivatives are evaluated at the fiducial value of $\overline{\rho}$. Evaluating the derivative on Eq. (23), we immediately see that the b_N defined through Eq. (24) coincide with the renormalized bias parameters introduced in Eq. (17). Thus, the renormalized bias parameters are exactly the peak-background split bias parameters understood in the sense of Eq. (24). Note also that this is directly connected to the derivation of bias (linear bias in that case) in the relativistic context presented in Sec. 4.3. 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 $\langle n_h \rangle$ for varying $\overline{\rho}'$ it is crucial to keep the age of the Universe fixed (see also Sec. 4.3).

It is worth emphasizing again 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 cut-off, i.e. the filter function W_L and scale R_L.[3]

On the other hand, after renormalization we only need a prediction for $\langle n_h \rangle$ as function of the background density $\overline{\rho}$ to calculate the statistics of tracers; no knowledge of the function $F_{h,L}$ is necessary. Further, the same PBS bias parameters describe both tracer auto-correlation and the cross-correlation with other tracers (including matter), which is a key requirement for a physical bias expansion (the corresponding statement in Fourier space is complicated by small-scale effects such as stochasticity, shot noise and halo exclusion which contribute to the tracer power spectrum at all k).

A careful inspection of Eq. (18) shows that this result does have an apparent flaw: the expression on the right-hand side involves nzl correlators of the coarse-grained matter density δ_L, which depend on R_L (though typically this dependence is very small for $R_L \ll r$). We will turn to this issue in Sec. 2.4.1. First however, we will consider the application of the renormalized bias framework to the well-known case of tracers following a universal mass function.

2.3.1 Renormalized 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

$$\bar{n}_h = \overline{\rho}\, f(\nu_c)\, J_* \tag{26}$$

$$\nu_c \equiv \frac{\delta_c}{\sigma_*}; \quad J_* \equiv \frac{d \ln \sigma_*}{d \ln R_*}, \tag{27}$$

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. (26) is referred to as "universal mass function" and was originally motivated by the excursion set formalism [7]. It is a special case of a more general description of mean tracer abundance we will consider in Sec. 3.5.

In order to derive the bias parameters Eq. (24), we need to know how $\bar{n}_h$ changes under a change in the background density of the Universe [Eq. (21)]. Since we work in the Lagrangian picture, we will ignore the trivial dependence through the $\overline{\rho}$ prefactor in Eq. (26). The threshold δ_c is defined as the fractional overdensity a region must have to collapse[4] 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 $\nu_c = \delta_c/\sigma_* = (\rho_c - \overline{\rho})/\delta\rho_{\rm RMS}$

[3] The b_N are closely related to the resummed bias propagators defined in [6] [see Eqs. (83)–(84) there], while the bare bias parameters c_n correspond to the bare propagators [Eqs. (1)–(2) in that paper].

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

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

$$\rho_c - \overline{\rho}' = (1+\delta_c)\overline{\rho} - (1+D)\overline{\rho} = (\delta_c - D)\overline{\rho}\,. \tag{28}$$

Thus, the significance is modified to

$$\nu_c' = \frac{\rho_c - \overline{\rho}'}{\sigma_*\overline{\rho}} = \frac{\delta_c - D}{\sigma_*}\,. \tag{29}$$

For a mass function of the form Eq. (26), changing the background density is thus equivalent to changing $\delta_c \to \delta_c - D$. Eq. (26) and Eq. (24) thus immediately yield

$$b_N = \frac{(-1)^N}{\langle n_h\rangle}\frac{\partial^N\langle n_h\rangle}{\partial\delta_c^N} = \frac{(-1)^N}{\sigma_*^N}\frac{1}{f(\nu_c)}\frac{d^N f(\nu_c)}{d\nu_c^N}. \tag{30}$$

This is the widely known expression for the peak-background split bias parameters, which in our approach is a special case of Eq. (24).

2.4 Beyond local bias

2.4.1 Non-locality

We now turn to the issue of the residual R_L-dependence on the right-hand side of Eq. (18). As long as $r \gg R_L$, or more precisely, as long as the matter correlators do not have structure on scales of order R_L, the residual dependence on R_L present in the "nzl" statistics of δ_L can be neglected, that is we can set $R_L \to 0$. In general, however, the EFT framework suggests that additional terms need to be included in the bias expansion in order to absorb this R_L dependence. Physically, the reason is that Eq. (18) was obtained from a local bias expansion, whereas we expect corrections to this expansion as r approaches R_*, the non-locality scale of the tracer considered. Instead of just depending on the coarse-grained density, we then expect the abundance of tracers to depend on the local gradient, curvature, and higher derivatives of the coarse-grained density field. Note that isotropy dictates that any dependence on the gradient of the density has to appear at quadratic order. In the following, we will restrict the treatment to linear order in the density for simplicity, in which case the lowest relevant derivative is the curvature (Laplacian) of the density field. We can already guess that the lowest order correction to Eq. (18) will be of the form

$$R_*^2\,\nabla_r^2\xi(r)\,, \tag{31}$$

where $\xi(r)$ is the matter correlation function. This is exactly the type of correction which appears for peaks of the density field [8]. In fact, this term can become observationally relevant not just for $r \sim R_*$ but on the much larger BAO feature scale ($r \sim 120\,{\rm Mpc}/h$) because of the narrow width of the feature and corresponding enhancement of the derivatives of $\xi(r)$. We now show how including the dependence of the tracer number density on derivatives of the coarse-grained density in fact removes the R_L-dependence in Eq. (18) and leads to terms of type Eq. (31).

For a general isotropic filter function (see App. B in [5]) the effect of smoothing on the correlation function $\xi(r)$ can be perturbatively described through

$$\begin{aligned}\xi_L(r) &= \int\frac{d^3k}{(2\pi)^3}|\tilde{W}_L(k)|^2P(k)e^{i\mathbf{k}\cdot\mathbf{r}}\\ &= \int\frac{d^3k}{(2\pi)^3}\left(1-2R_L^2k^2+\mathcal{O}(k^4)\right)P(k)e^{i\mathbf{k}\cdot\mathbf{r}}\\ &= \xi(r) + 2R_L^2\nabla^2\xi(r) + \mathcal{O}(\nabla^4\xi(r)),\end{aligned} \tag{32}$$

by suitable definition of the parameter R_L. The general expansion of $\tilde{W}_L(k)$ and the expansion of $\xi_L(r)$ in terms of derivatives of $\xi(r)$ is given in App. B of [5]. 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:

$$n_h(\mathbf{x}) = F_{h,L}(\delta_L(\mathbf{x}); \nabla^2\delta_L(\mathbf{x}); \delta_s). \tag{33}$$

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. Terms such as $(\boldsymbol{\nabla}\delta_L)^2$ will however appear at second and higher order. 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 tracer auto-correlation becomes

$$\begin{aligned}\xi_h(r) &= c_1^2\langle\delta_L(1)\delta_L(2)\rangle + 2c_1c_{\nabla^2\delta}\langle\delta_L(1)\nabla^2\delta_L(2)\rangle \\ &= c_1^2\left[\xi(r) + 2R_L^2\nabla^2\xi(r)\right] + 2c_1c_{\nabla^2\delta}\nabla^2\xi(r).\end{aligned} \tag{34}$$

Here, we have neglected higher than second derivatives of $\xi(r)$, and defined

$$c_{\nabla^2\delta} = \frac{1}{\langle F_{h,L}(0)\rangle}\left\langle \frac{\partial F_{h,L}}{\partial(\nabla^2\delta_L)}\Big|_{\delta_L=0,\nabla^2\delta_L=0}\right\rangle. \tag{35}$$

Eq. (34) 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. 2.3 for the density itself. Consider the following transformation of the density field:

$$\delta(\mathbf{x}) \to \delta_\alpha(\mathbf{x}) = \delta(\mathbf{x}) + \frac{\alpha}{6\ell^2}\mathbf{x}^2\,, \tag{36}$$

where α is a dimensionless small parameter, and we have added a length scale ℓ. Under this transformation, the Laplacian of the density perturbation shifts by a constant:

$$\nabla^2\delta_\alpha(\mathbf{x}) = \nabla^2\delta(\mathbf{x}) + \frac{\alpha}{\ell^2}\,. \tag{37}$$

Note that Eq. (36) 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 can now define a generalized PBS bias parameter through

$$b_{\nabla^2\delta} = \frac{\ell^2}{\langle n_h\rangle}\frac{\partial\langle n_h(\mathbf{0})\rangle}{\partial\alpha}\Big|_{\alpha=0}\,. \tag{38}$$

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). As in the case of local bias, the parameter defined through Eq. (38) will turn out to be the renormalized bias with respect to curvature of the density field.

Applying Eq. (38) to Eq. (33) via the chain rule yields

$$\begin{aligned}b_{\nabla^2\delta} &= \frac{\ell^2}{F_{h,L}(0)}\left(\frac{\partial F_{h,L}(\delta_L,0)}{\partial\delta_L}\frac{\partial\delta_L}{\partial\alpha} + \frac{\partial F_{h,L}(0,\nabla^2\delta_L)}{\partial(\nabla^2\delta_L)}\frac{\partial(\nabla^2\delta_L)}{\partial\alpha}\right) \\ &= c_1R_L^2 + c_{\nabla^2\delta}.\end{aligned} \tag{39}$$

Rewriting Eq. (34) in terms of $b_{\nabla^2\delta}$ then yields

$$\xi_h(r) = b_1^2\xi(r) + 2b_1b_{\nabla^2\delta}\nabla^2\xi(r)\,. \tag{40}$$

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, one can show that this continues to arbitrary powers of derivatives,

and that the PBS bias parameters can *entirely* absorb the smoothing effect on ξ_L [5]. Specifically, up to order $\nabla^4\xi(r)$, we obtain

$$\xi_h(r) = b_1^2\xi(r) + 2b_1 b_{\nabla^2\delta}\nabla^2\xi(r) + \left[(b_{\nabla^2\delta})^2 + b_1 b_{\nabla^4\delta}\right]\nabla^4\xi(r)\,. \tag{41}$$

Thus, going beyond local biasing perturbatively is equivalent to adding terms of the type (at linear order in $\xi(r)$)

$$\left[b_{\nabla^{2n}\delta} + \cdots + (b_{\nabla^2\delta})^n\right]\nabla^{2n}\xi(r) \tag{42}$$

to the expression for the tracer correlation function. Note that for a tracer that depends on the matter distribution within a region of size R_*, the coefficients in Eq. (42) will all be of order R_*^{2n}. While in the case of local biasing there is (after renormalization) only one small quantity in which we expand, $\xi(r)$, taking into account of the non-locality introduces a second small parameter $R_*^2\nabla^2\xi(r)$. The ratio of these parameters, which depends on the nature of the tracer, then determines to what order one has to go in the expansion in powers of $\xi(r)$ and derivatives of $\xi(r)$.

In EFT language, Eq. (42) corresponds to an infinite set of higher momentum (irrelevant) operators. If we had a complete description of the formation of tracers, we could predict each bias parameter (coupling constant) in this expansion, whereas in most practical cases these parameters need to be measured observationally. R_* then corresponds to the physical cutoff of the effective theory, that is, our perturbative description of tracer statistics has to break down as $r \to R_*$. Thus, if we measure $b_{\nabla^2\delta} \sim R_*^2$ for a given tracer, this measurement tells us, roughly but in an entirely model-independent way, on what scales the statistics of this tracer can be modeled perturbatively.

Furthermore, the definition of the PBS bias parameter $b_{\nabla^2\delta}$ [Eq. (38)] 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 [5]).

2.4.2 Tidal fields and velocities

So far, we have assumed that the tracer abundance n_h is exclusively determined by the matter density field. In reality, we expect a dependence on the matter velocity as well (density and velocity together describe the matter field completely in the fluid approximation). An immediate consequence of Galilei invariance is that n_h cannot depend on the matter velocity $\boldsymbol{v}$ itself, however it can depend on derivatives $\partial_i v_j$ of the velocity. Note that this implies that there is no velocity bias on large scales. Since the divergence $\partial_i v^i$ is proportional to the density itself, the key new quantity to include is the tidal field t_{ij}, the symmetric trace-free component of $\partial_i v_j$ (the antisymmetric part corresponds to the vorticity, which is negligible on large scales in practice).

Including the dependence of n_h on the tidal field leads of course to another bias hierarchy. Note however that the lowest order contribution is quadratic in t_{ij}, since the lowest order scalar one can create out of t_{ij} is $t_{kl}t^{kl}$. This contribution is of equal perturbative order as the local quadratic bias b_2, and the importance of including the quadratic tidal bias in describing halo clustering has recently been emphasized in [9, 10].

3 Non-Gaussian initial conditions

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. The motivation to study this type of non-Gaussian initial conditions is that it cannot be produced in single-field inflation (see Sec. 5), whereas it is in general present in multi-field inflation, i.e. inflation with at least two light degrees of freedom during inflation. Thus, detecting local primordial non-Gaussianity would rule out single-field inflation. Further, this type of non-Gaussianity has been shown to lead to a large modification of clustering on large scales. General shapes of non-Gaussianity will be considered in Sec. 3.4.

The derivation in Sec. 2 can be generalized to a general non-Gaussian density field, in which Eq. (18) formally retains its validity [5]. 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

$$\xi_h(r) = b_1^2 \xi_L(r) + b_1 b_2 \langle \delta_L(1) \delta_L^2(2) \rangle + \mathcal{O}(\delta_L^4), \tag{43}$$

the second term being the leading non-Gaussian correction. Now consider primordial non-Gaussianity of the local type, and ignore all non-Gaussianity from non-linear gravitational evolution. The matter bispectrum is given by

$$B_m(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \mathcal{M}(k_1)\mathcal{M}(k_2)\mathcal{M}(k_3) 2 f_{\rm NL} \left[P_\phi(k_1) P_\phi(k_2) + 2 \,\text{perm} \right] , \tag{44}$$

where $\mathcal{M}(k)$ is the relation between potential and density given in Eq. (55). In the limit $r \gg R_L$, the second correlator is then given by (see Eq. (57))

$$\langle \delta_L(1) \delta_L^2(2) \rangle = 4 f_{\rm NL} \sigma_L^2 \xi_{\phi\delta}(r), \tag{45}$$

where $\xi_{\phi\delta}$ is the cross-correlation between the matter and primordial Bardeen potential ϕ. Given the appearance of σ_L^2, Eq. (43) 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_*. For local-type NG, 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. 3.4). 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 [11, 12, 13], 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 [13].

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

$$\begin{aligned} \delta_s(\mathbf{x}) &\equiv \delta_*(\mathbf{x}) - \delta_L(\mathbf{x}) \qquad (46) \\ &= \int d^3\mathbf{y} [W_*(\mathbf{x}-\mathbf{y}) - W_L(\mathbf{x}-\mathbf{y})] \delta(\mathbf{y}) \\ &= \int \frac{d^3\mathbf{k}}{(2\pi)^3} \tilde{W}_s(k) \delta(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}, \\ \tilde{W}_s(k) &= \tilde{W}_*(k) - \tilde{W}_L(k). \qquad (47) \end{aligned}$$

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:

$$\begin{aligned} \langle \delta_s(\mathbf{x}) \delta_R(\mathbf{x}) \rangle &= \int \frac{d^3k}{(2\pi)^3} \tilde{W}_s(k) \tilde{W}_R(k) P(k) \\ &\overset{R \gg R_L}{\to} 0, \qquad (48) \end{aligned}$$

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

$$\sigma_s^2\equiv\langle\delta_s^2\rangle=\int\frac{d^3k}{(2\pi)^3}|\tilde{W}_s(k)|^2P(k). \tag{49}$$

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

$$y_*(\mathbf{x})\equiv\frac{1}{2}\left(\frac{\delta_s^2(\mathbf{x})}{\sigma_s^2}-1\right), \tag{50}$$

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. (48)], 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_*,

$$n_h(\mathbf{x})=F_{h,L}(\delta_L(\mathbf{x}),y_*(\mathbf{x});\delta_s). \tag{51}$$

Although our approach here is formally similar to the bivariate local expansion in δ_L and ϕ_L adopted in [14, 11, 15], there is somewhat of a conceptual difference in that we expand 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. 2, including this additional dependence. All arguments about the residual scatter from the deterministic relation $n_h(\mathbf{x})=n_h[\delta_L(\mathbf{x}),y_*(\mathbf{x})]$ and its negligible correlation with long-wavelength perturbations made in Sec. 2 also apply here. In fact, the dependence of $n_h(\mathbf{x})$ 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. (51), we obtain

$$\langle n_h\rangle\;=\langle F_{h,L}(0)\rangle\sum_{n,m}\frac{c_{nm}}{n!m!}\langle\delta_L^n y_*^m\rangle, \tag{52}$$

where we have defined bivariate "bare" bias parameters through

$$c_{nm}\equiv\frac{1}{\langle F_{h,L}(0)\rangle}\left\langle\frac{\partial^{n+m}F_{h,L}}{\partial\delta_L^n\partial y_*^m}\Big|_{\delta_L=0,y_*=0}\right\rangle. \tag{53}$$

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.

3.1 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,

$$\begin{aligned}B(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)&=\mathcal{M}(k_1)\mathcal{M}(k_2)\mathcal{M}(k_3)B_\phi(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\\ B_\phi(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)&=2f_{\rm NL}[P_\phi(k_1)P_\phi(k_2)+(2\text{ cyclic})].\end{aligned} \tag{54}$$

Here,

$$\mathcal{M}(k)=\frac{2}{3}\frac{k^2T(k)g(z)}{\Omega_{m0}H_0^2(1+z)} \tag{55}$$

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

$$\delta(\mathbf{k}, z) = \mathcal{M}(k)\phi(\mathbf{k}), \tag{56}$$

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 of [5],

$$\begin{aligned}\langle\delta_L(1)\delta_L^2(2)\rangle = &\int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{r}}\mathcal{M}_L(k)\int\frac{d^3k_1}{(2\pi)^3}\int\frac{d^3k_2}{(2\pi)^3}\\ &\times \mathcal{M}_L(k_1)\mathcal{M}_L(k_2)\langle\phi_{\mathbf{k}}\phi_{\mathbf{k}_1}\phi_{\mathbf{k}_2}\rangle\\ &= 4f_{\rm NL}\sigma_L^2\xi_{\phi\delta,L}(r),\end{aligned} \tag{57}$$

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.

$$\xi_{\phi\delta,L}(r) = \int\frac{d^3k}{(2\pi)^3}e^{i\mathbf{k}\cdot\mathbf{r}}\tilde{W}_L(k)\int\frac{d^3k_1}{(2\pi)^3}\langle\delta(\mathbf{k})\phi(\mathbf{k}_1)\rangle. \tag{58}$$

In deriving Eq. (57), 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. See [16] for a discussion of this issue.

Similarly, at leading order in $f_{\rm NL}$,

$$\begin{aligned}\langle\delta_L(1)y_*(2)\rangle &= \frac{1}{2}\Big\langle\delta_L(1)\frac{\delta_s^2(2)}{\sigma_s^2}\Big\rangle\\ &= \frac{1}{2\sigma_s^2}\int\frac{d^3k}{(2\pi)^3}e^{i\mathbf{k}\cdot\mathbf{r}}\mathcal{M}_L(k)\int\frac{d^3k_1}{(2\pi)^3}\int\frac{d^3k_2}{(2\pi)^3}\\ &\qquad\times\mathcal{M}_s(k_1)\mathcal{M}_s(k_2)\langle\phi_{\mathbf{k}}\phi_{\mathbf{k}_1}\phi_{\mathbf{k}_2}\rangle\\ &= 2f_{\rm NL}\xi_{\phi\delta,L}(r).\end{aligned} \tag{59}$$

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 $\phi(\mathbf{x})$,

$$\big\langle\delta_s^2(\mathbf{x})\big\rangle\big|_{\phi(\mathbf{x})} = \sigma_s^2\left[1 + 4f_{\rm NL}\phi(\mathbf{x})\right] + \mathcal{O}(f_{\rm NL}^2), \tag{60}$$

and hence

$$\langle y_*(\mathbf{x})\rangle|_{\phi(\mathbf{x})} = 2f_{\rm NL}\phi(\mathbf{x}) + \mathcal{O}(f_{\rm NL}^2), \tag{61}$$

which immediately leads to Eq. (59). 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. (59) 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.

3.2 Correlations

The expectation value of the correlation function becomes

$$\begin{aligned}\langle\hat{\xi}_h(r)\rangle &= \frac{1}{\mathcal{N}^2}\\ &\times\sum_{n,m,n',m'=0}^{\infty}\frac{c_{nm}c_{n'm'}}{n!m!n'!m'!}\langle\delta_L^n(1)y_*^m(1)\delta_L^{n'}(2)y_*^{m'}(2)\rangle - 1,\end{aligned} \tag{62}$$

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

$$\mathcal{N} \equiv \sum_{n,m=0}^{\infty} \frac{c_{nm}}{n!m!} \langle \delta_L^n y_*^m \rangle. \tag{63}$$

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

$$\langle \hat{\xi}_{hm}(r) \rangle = \frac{1}{\mathcal{N}} \sum_{n,m;n+m>0}^{\infty} \frac{c_{nm}}{n!m!} \langle \delta_L^n(1) y_*^m(1) \delta_L(2) \rangle. \tag{64}$$

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. (48) 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. (14), which are absorbed by corresponding terms in $\mathcal{N}$. Thus, the final result Eq. (20) does not change in the Gaussian case if we include the dependence on y_*.

Defining for convenience

$$f(\mathbf{x}) = \sum_{n,m=0}^{\infty} \frac{c_{nm}}{n!m!} \delta_L(\mathbf{x}) y_*(\mathbf{x}) - 1, \tag{65}$$

we have

$$\xi_h(r) = \frac{1}{\mathcal{N}^2} \left[\langle f(1) f(2) \rangle - \langle f \rangle^2 \right], \tag{66}$$

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

$$\mathcal{N} = 1 + \langle f \rangle = 1 + \frac{c_{20}}{2} \sigma_L^2 + c_{11} \langle \delta_L y_* \rangle + \mathcal{O}(\delta^3), \tag{67}$$

and

$$\langle f \rangle^2 = \frac{c_{20}^2}{4} \sigma_L^4 + c_{11} c_{20} \langle \delta_L y_* \rangle \sigma_L^2 + \mathcal{O}(\delta^5). \tag{68}$$

Hence, ξ_h becomes

$$\begin{aligned}
\xi_h = \frac{1}{\mathcal{N}^2} \Big\{ \Big\langle & \left(c_{10}\delta_L + c_{01} y_* + c_{11} \delta_L y_* + \frac{c_{20}}{2} \delta_L^2 + \frac{c_{30}}{6} \delta_L^3 \right)_1 \\
& \left(c_{10}\delta_L + c_{01} y_* + c_{11} \delta_L y_* + \frac{c_{20}}{2} \delta_L^2 + \frac{c_{30}}{6} \delta_L^3 \right)_2 \Big\rangle \\
& - \frac{c_{20}^2}{4} \sigma_L^4 - c_{11} c_{20} \langle \delta_L y_* \rangle \sigma_L^2 \Big\} \\
= \frac{1}{\mathcal{N}^2} \Big[& c_{10}^2 \langle \delta_L(1) \delta_L(2) \rangle + 2 c_{10} c_{01} \langle \delta_L(1) y_*(2) \rangle \\
& + c_{10} c_{20} \langle \delta_L(1) \delta_L^2(2) \rangle + c_{10} c_{30} \sigma_L^2 \langle \delta_L(1) \delta_L(2) \rangle \\
& + c_{01} c_{30} \sigma_L^2 \langle \delta_L(1) y_*(2) \rangle \\
& + 2 c_{11} c_{20} \langle \delta_L(1) y_*(2) \rangle \langle \delta_L(1) \delta_L(2) \rangle \\
& + 2 \frac{c_{20}^2}{4} \langle \delta_L(1) \delta_L(2) \rangle^2 \Big] + \mathcal{O}(\delta^5)
\end{aligned} \tag{69}$$

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

$$\begin{aligned}
\langle y_*(1) \delta_L^3(2) \rangle &= 3 \langle y_*(1) \delta_L(2) \rangle \sigma_L^2 + \mathcal{O}(f_{\rm NL}^2) \\
\langle \delta_L(1) y_*(1) \delta_L^2(2) \rangle &= 2 \langle \delta_L(1) y_*(2) \rangle \langle \delta_L(1) \delta_L(2) \rangle \\
&\quad + \langle \delta_L y_* \rangle \sigma_L^2 + \mathcal{O}(f_{\rm NL}^2).
\end{aligned} \tag{70}$$

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. 3.1. Using Eqs. (57)–(59), we obtain

$$\begin{aligned}\xi_h(r) = \frac{1}{\mathcal{N}^2}\Big[&\left(c_{10}^2 + c_{10}c_{30}\sigma_L^2\right)\xi_L(r) + \frac{c_{20}^2}{2}\xi_L(r)^2 \\ &+ \left(2c_{10}c_{01} + c_{01}c_{30}\sigma_L^2 + 2c_{10}c_{20}\sigma_L^2\right) 2f_{\rm NL}\xi_{\phi\delta,L}(r) \\ &+ 2c_{11}c_{20}2f_{\rm NL}\xi_{\phi\delta,L}(r)\xi_L(r)\Big]. \end{aligned} \tag{71}$$

3.3 Bivariate PBS bias parameters

In analogy to Sec. 2.3, 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 $(1+\varepsilon)$.[5] 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

$$\begin{aligned}&\delta_L(\mathbf{x}) \to (1+\varepsilon)\delta_L(\mathbf{x}) \\ &y_*(\mathbf{x}) \to y_*(\mathbf{x}) + \left(\varepsilon + \frac{\varepsilon^2}{2}\right)\frac{\delta_s^2(\mathbf{x})}{\sigma_s^2}\,. \end{aligned} \tag{72}$$

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. 2.3.

We can then define a set of *bivariate PBS bias parameters* b_{NM} by generalizing Eq. (24) to

$$b_{NM} \equiv \frac{1}{\langle n_h\rangle_{D=0,\varepsilon=0}}\frac{\partial^{N+M}\langle n_h\rangle_{D,\varepsilon}}{\partial D^N\partial\varepsilon^M}\bigg|_{D=0,\varepsilon=0}. \tag{73}$$

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. (52),

$$\begin{aligned}\langle n_h\rangle(D,\varepsilon) = \langle F_{h,L}(0)\rangle &\sum_{n,m=0}^{\infty}\frac{c_{nm}}{n!m!} \\ &\times\left\langle\,[(1+\varepsilon)\delta_L + D]^n\left[y_* + \left(\varepsilon+\frac{\varepsilon^2}{2}\right)\frac{\delta_s^2}{\sigma_s^2}\right]^m\right\rangle. \end{aligned} \tag{74}$$

We thus have

$$b_{N0} = b_N. \tag{75}$$

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

In particular,

$$b_{10} = \frac{1}{\mathcal{N}}\left(c_{10} + \frac{c_{30}}{2}\sigma_L^2 + \mathcal{O}(\delta_L^3)\right). \tag{76}$$

Further,

$$\begin{aligned} b_{01} &= \frac{1}{\mathcal{N}}\sum_{n,m}\frac{c_{nm}}{n!m!}\left(n\langle\delta_L^n y_*^m\rangle + m\langle\delta_L^n(1+2y_*)y_*^{m-1}\rangle\right) \\ &= \frac{1}{\mathcal{N}}\left(c_{01} + c_{20}\sigma_L^2 + c_{11}\langle\delta_L y_*\rangle + \frac{c_{30}}{2}\left\langle\delta_L^3\right\rangle + \mathcal{O}(\delta^4)\right). \end{aligned} \tag{77}$$

We can now express the correlation function of tracers at this order, Eq. (69), 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 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. (72)]. 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 $\delta^4, f_{\rm NL}$. We obtain

$$\begin{aligned} \xi_h(r) = b_{10}^2\xi_L(r) &+ \frac{b_{20}^2}{2}\xi_L^2(r) + 2b_{10}b_{01}\langle\delta_L(1)y_*(2)\rangle \\ &+ b_{20}b_{11}\langle\delta_L(1)y_*(1)\delta_L^2(2)\rangle_{\rm nzl} \\ &+ \mathcal{O}(\delta^5, f_{\rm NL}^2). \end{aligned} \tag{78}$$

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

$$\begin{aligned} \xi_h(r) = \frac{1}{\mathcal{N}^2}\Big\{ &\left(c_{10}^2 + c_{10}c_{30}\sigma_L^2\right)\xi_L(r) + \frac{c_{20}^2}{2}\xi_L^2(r) \\ &+ 2\left(c_{10}c_{01} + \frac{c_{30}}{2}c_{01}\sigma_L^2 + c_{10}c_{20}\sigma_L^2\right)2f_{\rm NL}\xi_{\phi\delta,L}(r) \\ &+ 2c_{20}c_{11}2f_{\rm NL}\xi_{\phi\delta,L}(r)\xi_L(r) \\ &\Big\} + \mathcal{O}(\delta^5, f_{\rm NL}^2). \end{aligned} \tag{79}$$

We easily see that this agrees identically with Eq. (71). 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. (62), 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. (73) and the resulting expression Eq. (78) 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. (18) 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 highly undesirable breakdown of the perturbative expansion on *large scales*. The bivariate expansion on the

other hand leads to an expansion in which higher order terms are consistently suppressed [Eq. (78)], 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.

3.4 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. (57)–(59), 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

$$B_\phi(\mathbf{k}_l, \mathbf{k}_s, -\mathbf{k}_l - \mathbf{k}_s) \overset{k_l \ll k_s}{=} A \left(\frac{k_l}{k_s}\right)^\alpha P_\phi(k_l) P_\phi(k_s), \tag{80}$$

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. (57)–(59) then generalize to

$$\begin{aligned}
\langle \delta_L(1) \delta_L^2(2) \rangle &= \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{r}} \mathcal{M}_L(k) \int \frac{d^3k_1}{(2\pi)^3} \\
&\qquad \times \mathcal{M}_L^2(k_1) A k^\alpha P_\phi(k) k_1^{-\alpha} P_\phi(k_1) \\
&= A \sigma^2_{-\alpha,L} \xi_{\phi_\alpha \delta, L}(r) \qquad (81) \\
\langle \delta_L(1) y_*(2) \rangle &= \frac{1}{2} \Big\langle \delta_L(1) \frac{\delta_s^2(2)}{\sigma_s^2} \Big\rangle \\
&= \frac{1}{2\sigma_s^2} \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{r}} \mathcal{M}_L(k) \int \frac{d^3k_1}{(2\pi)^3} \\
&\qquad \times \mathcal{M}_s^2(k_1) A k^\alpha P_\phi(k) k_1^{-\alpha} P_\phi(k_1) \\
&= A \frac{\sigma^2_{-\alpha,s}}{2\sigma_s^2} \xi_{\phi_\alpha \delta, L}(r), \qquad (82)
\end{aligned}$$

where we have defined the general spectral moment

$$\sigma^2_{n,X} \equiv \int \frac{d^3k}{(2\pi)^3} k^n P(k) |\tilde{W}_X(k)|^2, \tag{83}$$

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

$$\xi_{\phi_\alpha \delta, L}(r) \equiv \int \frac{d^3k}{(2\pi)^3} k^\alpha \mathcal{M}(k) P_\phi(k) \tilde{W}_L(k). \tag{84}$$

Again, Eqs. (81)–(82) 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. (69), and using Eq. (70) we obtain up to $\mathcal{O}(\delta^5, f_{\rm NL}^2)$

$$\begin{aligned}
\xi_h(r) = \frac{1}{\mathcal{N}^2} \Bigg[& \left(c_{10}^2 + c_{10} c_{30} \sigma_L^2\right) \xi_L(r) + \frac{c_{20}^2}{2} \xi_L(r)^2 \\
& + \left(2 c_{10} c_{01} \frac{\sigma^2_{-\alpha,s}}{2\sigma_s^2} + c_{01} c_{30} \frac{\sigma^2_{-\alpha,s}}{2\sigma_s^2} \sigma_L^2 + c_{10} c_{20} \sigma^2_{-\alpha,L} \right) \\
& \quad \times A \xi_{\phi_\alpha \delta, L}(r) \\
& + 2 c_{11} c_{20} \frac{\sigma^2_{-\alpha,s}}{2\sigma_s^2} A \xi_{\phi_\alpha \delta, L}(r) \xi_L(r) \Bigg].
\end{aligned} \tag{85}$$

Inspection shows that the bivariate PBS parameters defined in Sec. 3.3 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. (80), the small-scale perturbations are not rescaled uniformly, but rather in a scale-dependent way: the squeezed-limit result Eq. (60) generalizes to

$$\left\langle\delta_s^2(\mathbf{x})\right\rangle\big|_{\phi_\alpha(\mathbf{x})} = \sigma_s^2 + A\sigma^2_{-\alpha,s}\phi_\alpha(\mathbf{x}), \tag{86}$$

where

$$\phi_\alpha(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} k^\alpha \phi(\mathbf{k}) e^{i\mathbf{kx}}, \tag{87}$$

and hence

$$\langle y_*(\mathbf{x})\rangle|_{\phi_\alpha(\mathbf{x})} = A\frac{\sigma^2_{-\alpha,s}}{2\sigma_s^2}\phi_\alpha(\mathbf{x}). \tag{88}$$

Thus, the transformation of the density field following Eq. (72) is not the relevant one any more. Instead, we need to rescale the density field through

$$\delta(\mathbf{k}) \to \left(1 + \varepsilon k^{-\alpha}\right)\delta(\mathbf{k}), \tag{89}$$

so that

$$\begin{aligned}
\delta_L(\mathbf{x}) &\to \delta_L(\mathbf{x}) + \varepsilon\,\delta_{-\alpha,L}(\mathbf{x}) \\
y_*(\mathbf{x}) &\to y_*(\mathbf{x}) + \frac{\varepsilon}{\sigma_s^2}\delta_s(\mathbf{x})\delta_{-\alpha,s}(\mathbf{x}) + \frac{\varepsilon^2}{\sigma_s^2}\delta^2_{-\alpha,s}(\mathbf{x})\ ,
\end{aligned} \tag{90}$$

where, in analogy to Eq. (87),

$$\delta_{-\alpha,X}(\mathbf{x}) \equiv \int \frac{d^3k}{(2\pi)^3} k^{-\alpha}\tilde{W}_X(k) e^{i\mathbf{kx}}. \tag{91}$$

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. (73), but with the transformation Eq. (89), so that these bivariate bias parameters will in general be different from those in Sec. 3.3. As before, our next task is to derive the relation between b_{NM} and c_{nm}. We have from Eq. (52),

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

We obtain

$$\begin{aligned}
b_{01} &= \frac{1}{\mathcal{N}}\sum_{n,m}\frac{c_{nm}}{n!m!} \\
&\quad\times\left(n\langle\delta_{-\alpha,L}\delta_L^{n-1}y_*^m\rangle + \frac{m}{\sigma_s^2}\langle\delta_L^n\delta_s\delta_{-\alpha,s}y_*^{m-1}\right\rangle\Big) \\
&= \frac{1}{\mathcal{N}}\left(c_{01}\frac{\sigma^2_{-\alpha,s}}{\sigma_s^2} + c_{20}\sigma^2_{-\alpha,L} + \mathcal{O}(\delta^3, f_{\rm NL}^2)\right).
\end{aligned} \tag{92}$$

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. (89) (at lowest order), we have to divide by that factor when multiplying correlators involving y_*. We obtain

$$\begin{aligned}\xi_h(r) &= b_{10}^2\xi_L(r) + \frac{b_{20}^2}{2}\xi_L^2(r) + 2b_{10}b_{01}\frac{\sigma_s^2}{\sigma_{-\alpha,s}^2}\langle\delta_L(1)y_*(2)\rangle \\ &\quad + b_{20}b_{11}\frac{\sigma_s^2}{\sigma_{-\alpha,s}^2}\langle\delta_L(1)y_*(1)\delta_L^2(2)\rangle_{\rm nzl} \\ &= b_{10}^2\xi_L(r) + \frac{b_{20}^2}{2}\xi_L^2(r) + b_{10}b_{01}A\xi_{\phi_\alpha\delta,L}(r) \\ &\quad + b_{20}b_{11}A\xi_{\phi_\alpha\delta,L}(r)\xi_L(r). \end{aligned} \tag{93}$$

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. (82)], which in turn depends on σ_L^2 [Eq. (49)].

Inserting the expressions for b_{NM} at the relevant order, and using Eqs. (81)–(82), we have

$$\begin{aligned}\xi_h(r) &= \frac{1}{\mathcal{N}^2}\Bigg\{ \left(c_{10}^2 + c_{10}c_{30}\sigma_L^2\right)\xi_L(r) + \frac{c_{20}^2}{2}\xi_L^2(r) \\ &\quad + \left[2c_{10}c_{01} + 2c_{10}c_{20}\sigma_{-\alpha,L}^2\frac{\sigma_s^2}{\sigma_{-\alpha,s}^2} + c_{30}c_{01}\sigma_L^2\right] \\ &\quad\quad \times\langle\delta_L(1)y_*(2)\rangle \\ &\quad + c_{20}c_{11}\langle\delta_L(1)y_*(1)\delta_L^2(2)\rangle_{\rm nzl}\Bigg\} \\ &= \frac{1}{\mathcal{N}^2}\Bigg\{ \left(c_{10}^2 + c_{10}c_{30}\sigma_L^2\right)\xi_L(r) + \frac{c_{20}^2}{2}\xi_L^2(r) \\ &\quad + \left[2c_{10}c_{01} + c_{30}c_{01}\sigma_L^2\right]\frac{\sigma_{-\alpha,s}^2}{2\sigma_s^2}A\xi_{\phi_\alpha\delta,R(r)} \\ &\quad + c_{10}c_{20}\sigma_{-\alpha,L}^2A\xi_{\phi_\alpha\delta,R(r)} \\ &\quad + 2c_{20}c_{11}\xi_L(r)\frac{\sigma_{-\alpha,s}^2}{2\sigma_s^2}A\xi_{\phi_\alpha\delta,L}(r)\Bigg\}. \end{aligned} \tag{94}$$

This agrees exactly with Eq. (85). 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. (93), from the corresponding result for local non-Gaussianity Eq. (78) is that the bivariate bias parameters are now defined with respect to the scale-dependent rescaling of the density field, Eq. (89), 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 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. (80), it is still sufficient to parametrize the dependence of 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. (88)]. 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. [17]), 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.

3.5 Universal mass functions

We begin with a generalization of the universal mass function discussed in Sec. 2.3.1. We write the mean abundance of tracers as

$$\bar{n}_h = \bar{n}_h\left(\bar{\rho}, \sigma_*, J_*\right) , \tag{95}$$

where the Jacobian J_* is defined in Eq. (27). 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. (89), σ_* transforms to lowest order as

$$\sigma_* \to \sigma_* \left[1 + \varepsilon \frac{\sigma^2_{-\alpha,*}}{\sigma^2_*}\right] , \tag{96}$$

while the Jacobian transforms as (see also [18])

$$\begin{aligned} J_* &\to J_* + \varepsilon \frac{\sigma^2_{-\alpha,*}}{\sigma^2_*}\left(\frac{d\ln\sigma^2_{-\alpha,*}}{d\ln R_*} - \frac{d\ln\sigma^2_*}{d\ln R_*}\right) \\ &= J_*\left[1 + 2\varepsilon \frac{\sigma^2_{-\alpha,*}}{\sigma^2_*}\left(\frac{d\ln\sigma^2_{-\alpha,*}}{d\ln\sigma^2_*} - 1\right)\right] . \end{aligned} \tag{97}$$

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. (95), we can then derive the leading non-Gaussian bias through Eq. (73):

$$\begin{aligned} b_{01} &= \frac{1}{\bar{n}_h}\left(\frac{\partial\bar{n}_h}{\partial\ln\sigma_*}\frac{\partial\ln\sigma_*}{\partial\varepsilon} + \frac{\partial\bar{n}_h}{\partial\ln J_*}\frac{\partial\ln J_*}{\partial\varepsilon}\right) \\ &= \left[\frac{1}{\bar{n}_h}\frac{\partial\bar{n}_h}{\partial\ln\sigma_*} + \frac{1}{\bar{n}_h}\frac{\partial\bar{n}_h}{\partial\ln J_*} 2\left(\frac{d\ln\sigma^2_{-\alpha,*}}{d\ln\sigma^2_*} - 1\right)\right]\frac{\sigma^2_{-\alpha,*}}{\sigma^2_*} \\ &= \left[b_{01}(\alpha=0) + 2\left(\frac{d\ln\sigma^2_{-\alpha,*}}{d\ln\sigma^2_*} - 1\right)\right]\frac{\sigma^2_{-\alpha,*}}{\sigma^2_*} . \end{aligned} \tag{98}$$

Here, $b_{01}(\alpha = 0)$ is the PBS bias parameter quantifying the effect of local primordial non-Gaussianity for a tracer following Eq. (95), 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 [18], who first pointed out the contribution by the Jacobian J_*.

We now specialize Eq. (95) to a "truly" universal mass function [Eq. (26)],

$$\bar{n}_h = \bar{\rho}\, f(\nu_c)\, J_* , \quad \nu_c \equiv \frac{\delta_c}{\sigma_*} , \tag{99}$$

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. (99) further allows us to connect $b_{01}(\alpha = 0)$ to the linear PBS density bias:

$$\begin{aligned} b_{10} &= \frac{1}{\bar{n}_h}\frac{\partial \bar{n}_h}{\partial \ln \overline{\rho}} = -\frac{1}{\sigma_*}\frac{df}{d\nu_c} \\ b_{01}(\alpha = 0) &= \frac{1}{\bar{n}_h}\frac{\partial \bar{n}_h}{\partial \ln \sigma_*} = -\frac{\delta_c}{\sigma_*}\frac{df}{d\nu_c} = \delta_c b_{10}. \end{aligned} \tag{100}$$

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. (99) (see also Sec. 2.3.1); 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 [19, 20, 11].

4 From correlations to observations: relativistic effects

Forecasts indicate that future LSS surveys including BOSS, HETDEX and Euclid will probe values of local $f_{\rm NL}$ down to a few. This is of the same order of magnitude as relativistic terms we have neglected in the quasi-Newtonian treatment of galaxy clustering described above, i.e. terms that are suppressed by powers of aH/k on small scales. Galaxy formation happens on scales that are much smaller than the horizon, and the quasi-Newtonian treatment is completely sufficient for that process. However, the observed clustering of tracers over large scales receives corrections from the Newtonian result. There are two main corrections:

- The observationally inferred locations of tracers (via positions on the sky and measured redshift) are not the actual positions due to light propagation effects in the perturbed FRW spacetime.
- We measure the overdensities of tracers relative to a mean on a constant-observed-redshift surface, which does not coincide with a constant-proper-time surface on which biasing is naturally defined (as seen in the previous section).

Fortunately, these effects are only important on very large scales, and so linear perturbation theory is sufficient to describe them for all practical purposes.

An acute reader might notice that primordial non-Gaussianity is a second order effect in perturbation theory, as it arises from interactions of fields in the early Universe (inflation); so can we really get by with a linear treatment if we want to understand both relativistic effects and primordial non-Gaussianity ? The answer is yes, in a certain limit, by making use of the gauge freedom in GR. We will return to this subtle but important point in Sec. 5.

4.1 Perturbed photon geodesics

In the absence of perturbations, photon geodesics are given by straight lines in conformal coordinates,

$$\bar{x}^\mu(\chi) = (\eta_0 - \chi, \hat{\boldsymbol{n}}\chi)\,, \tag{101}$$

where we have chosen the comoving distance χ as affine parameter. Correspondingly, for a photon arriving from a direction $\hat{\boldsymbol{n}}$ with redshift $\tilde{z}$, we assign an "observed" position of emission x^μ given by

$$\begin{aligned} \tilde{x}^0 &= \eta_0 - \tilde{\chi} \\ \tilde{x}^i &= \hat{n}^i\,\tilde{\chi} \\ \tilde{\chi} &\equiv \bar{\chi}(\tilde{z}), \end{aligned} \tag{102}$$

where $\bar{\chi}(\tilde{z})$ denotes the comoving distance-redshift relation in the background Universe. Here we have chosen the observer to reside at the spatial origin without loss of generality. The coordinate time of the observer who is assumed to be comoving is fixed by the condition of a fixed proper time t_0 at

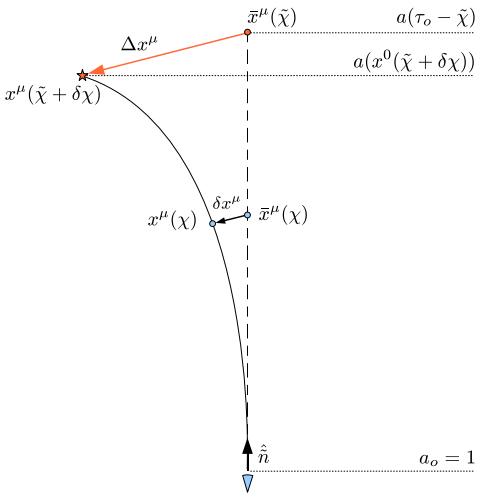

Figure 2: Sketch of perturbed photon geodesics illustrating our notation. The observer is located at the bottom. The solid line indicates the actual photon geodesic tracing back to the source indicated by a star. The dashed line shows the apparent background photon geodesic tracing back to an inferred source position indicated by a circle.

observation. On the other hand, the actual spacetime point of emission, denoted with x^μ, is displaced from the observed positions by Δx^μ (see also Fig. 2),

$$x^\mu = \tilde{x}^\mu + \Delta x^\mu(\hat{\boldsymbol{n}}, \tilde{z}). \tag{103}$$

Further, we will need the scale factor at emission. It is related to the inferred emission scale factor $\tilde{a} \equiv 1/(1+\tilde{z})$ by

$$\frac{a(x^0(\hat{\boldsymbol{n}}, \tilde{z}))}{\tilde{a}} = 1 + \delta z\,. \tag{104}$$

δz is essentially the difference between the observed redshift $\tilde{z}$ and the redshift $\bar{z}$ that would be observed in an unperturbed Universe, specifically

$$\bar{z} - \tilde{z} = -(1+\tilde{z})\delta z\,. \tag{105}$$

Note that the displacements δz, Δx^μ are not observable in general (they depend on which gauge, or frame, the spacetime perturbations are described in). However, δz evaluated in synchronous-comoving gauge is equal to an observable as we will discuss below.

In this section, we work in synchronous-comoving (sc) gauge, for reasons that will become clear below. Specifically, we write

$$ds^2 = a^2(\tau)\left\{-d\tau^2 + \left[\,(1+2D)\,\delta_{ij} + 2E_{ij}\right] dx^i\, dx^j\right\}, \tag{106}$$

where τ is the conformal time, D is a scalar metric perturbation while E_{ij} is transverse and traceless, and related to the scalar perturbation E by

$$E_{ij} = \left(\partial_i\partial_j - \frac{1}{3}\delta_{ij}\nabla^2\right) E. \tag{107}$$

4.2 Observed galaxy number density

The observed number of galaxies contained within a volume $\tilde{V}$ defined in terms of the observed coordinates is given by a (gauge-invariant) integral over a three-form

$$\begin{aligned} N &= \int_{\tilde{V}} \sqrt{-g(x^\alpha)}\, n_g(x^\alpha)\, \varepsilon_{\mu\nu\rho\sigma} u^\mu(x^\alpha) \frac{\partial x^\nu}{\partial \tilde{x}^1}\frac{\partial x^\rho}{\partial \tilde{x}^2}\frac{\partial x^\sigma}{\partial \tilde{x}^3} d^3\tilde{\mathbf{x}} \\ &= \int_{\tilde{V}} \sqrt{-g}\, n_g(x^\alpha)\, \frac{1}{a(x^0)} \varepsilon_{ijk} \frac{\partial x^i}{\partial \tilde{x}^1}\frac{\partial x^j}{\partial \tilde{x}^2}\frac{\partial x^k}{\partial \tilde{x}^3} d^3\tilde{\mathbf{x}} \\ &= \int_{\tilde{V}} \sqrt{-g}\, n_g(x^\alpha)\, \frac{1}{a(x^0)} \left|\frac{\partial x^i}{\partial \tilde{x}^j}\right| d^3\tilde{\mathbf{x}}. \end{aligned} \tag{108}$$

Here, $\mathbf{x}$ is given in terms of $\tilde{\mathbf{x}}$ by Eq. (103), $\varepsilon_{\mu\nu\rho\sigma}$ is the Levi-Civita tensor, and n_g is the physical number density of galaxies as a function of "true" comoving locations (in synchronous-comoving gauge). In the second line, we have adopted the synchronous-comoving gauge, where the oberver velocities reduce to $u^\mu = (1/a, 0, 0, 0)$. In this case, not surprisingly, the volume element reduces to the purely spatial Jacobian $|\partial x^i/\partial \tilde{x}^j|$. Note that perturbations enter Eq. (108) in three places: through the determinant $\sqrt{-g}$; through the position- and redshift-dependence of the galaxy density n_g, and through the Jacobian $|\partial x^i/\partial \tilde{x}^j|$. This Jacobian is

$$\left|\frac{\partial x^i}{\partial \tilde{x}^j}\right| = \left|\delta^i_j + \frac{\partial \Delta x^i}{\partial \tilde{x}^j}\right| = 1 + \frac{\partial \Delta x^i}{\partial \tilde{x}^i}, \tag{109}$$

where we have worked to first order in the displacements $\Delta\mathbf{x}$. Furthermore, noting that $\sqrt{-\bar{g}} = a^4$, where $\bar{g}_{\mu\nu}$ is the background metric, we have

$$\sqrt{-g} = a^4 \left(1 + \frac{1}{2}\delta g^\mu_\mu\right). \tag{110}$$

Finally, the galaxy density perturbations are usually measured with respect to the average density of galaxies at fixed observed redshift, $\bar{n}_g(\tilde{z})$. We assume that when averaged over the whole survey, $\langle \delta z\rangle = 0$ so that $\langle \tilde{z}\rangle = \bar{z}$ [Eq. (104)]. Also, in this paper, we follow common convention and define the galaxy density perturbations $\delta_g^{\rm sc}$ with respect to the *comoving* galaxy density. We thus have for the intrinsic comoving galaxy density

$$a^3(\bar{z}) n_g(\mathbf{x}, \bar{z}) = a^3(\bar{z})\bar{n}_g(\bar{z})\,[1 + \delta_g^{\rm sc}(\mathbf{x}, \bar{z})], \tag{111}$$

where $\bar{z}$ again denotes the redshift that would have been observed in an unperturbed Universe, and $\delta_g^{\rm sc}$ denotes the intrinsic fluctuations in the comoving galaxy density in synchronous-comoving gauge. Using Eq. (104) and expanding to first order, we obtain

$$\begin{aligned} a^3(z) n_g(\mathbf{x}, \bar{z}) = a^3(\tilde{z})\bar{n}_g(\tilde{z})\,[1 + \delta_g^{\rm sc}(\tilde{\mathbf{x}})] \\ - (1+\tilde{z})\frac{d(a^3\bar{n}_g)}{dz}\Big|_{z=\tilde{z}} \delta z. \end{aligned} \tag{112}$$

We define

$$b_e \equiv \frac{d\ln(a^3\bar{n}_g)}{d\ln a}\Big|_{\tilde{z}} = -(1+\tilde{z})\frac{d\ln(a^3\bar{n}_g)}{dz}\Big|_{\tilde{z}}. \tag{113}$$

Note that the distinction between $\delta_g^{\rm sc}(\tilde{\mathbf{x}})$ and $\delta_g^{\rm sc}(\mathbf{x})$ is second order (this effect, analogous to CMB lensing, can however become important for rapidly varying correlation functions [21]).

We can now expand Eq. (108). We define the *observed* galaxy density $\tilde{n}_g$ via

$$\int_{\tilde{V}} a^3(\tilde{z})\tilde{n}_g(\tilde{\mathbf{x}}, \tilde{z}) d^3\tilde{\mathbf{x}} = N, \tag{114}$$

so that

$$\begin{aligned} a^3(\tilde{z})\tilde{n}_g(\tilde{\mathbf{x}}, \tilde{z}) &= \sqrt{-g}\frac{1}{a(\bar{z})}\, n_g(\mathbf{x}, \bar{z})\, \left|\frac{\partial x^i}{\partial \tilde{x}^j}\right| \\ &= \left(1 + \frac{1}{2}\delta g^\mu_\mu\right) a^3(\bar{z}) n_g(\mathbf{x}, \bar{z}) \left(1 + \frac{\partial \Delta x^i}{\partial \tilde{x}^i}\right). \end{aligned} \tag{115}$$

For the Jacobian, we use the divergence on the full sky [22] to obtain

$$\begin{aligned}\frac{\partial \Delta x^i}{\partial \tilde{x}^i} &= \partial_\parallel \Delta x_\parallel + \Delta x_\parallel \partial_i \hat{n}^i + \partial_{\perp\, i} \Delta x^i_\perp \\ &= \partial_\parallel \Delta x_\parallel + \frac{2\Delta x_\parallel}{\tilde{\chi}} - 2\hat{\kappa},\end{aligned} \tag{116}$$

where we have defined the coordinate convergence as

$$\hat{\kappa} \equiv -\frac{1}{2}\partial_{\perp\, i} \Delta x^i_\perp. \tag{117}$$

We then obtain for the observed galaxy overdensity

$$\tilde{\delta}_g = \delta_g^{\rm sc} + b_e \delta z + 2\frac{\Delta x_\parallel}{\tilde{\chi}} + \partial_\parallel \Delta x_\parallel - 2\hat{\kappa}\,. \tag{118}$$

See [22] for the full expression in terms of metric perturbations in synchronous gauge. The first term is the intrinsic galaxy overdensity, the second comes in since the mean galaxy density evolves, and constant observed redshift surfaces do not correspond to slices of constant proper time (constant evolutionary stage). In fact, this term is gauge-invariant by itself; δz is equal to the quantity $\mathcal{T}$ defined in [23], which corresponds to the difference in $\ln a$ between constant-proper-time and constant-observed-redshift surfaces. Finally, the last three terms correspond to volume distortion effects.

There is one further contribution in actual galaxy surveys that can be important on large scales: the fact that we typically deal with flux-limited sizes leads to a magnification bias contribution given by $\mathcal{Q}\mathcal{M}$, where $\mathcal{Q}$ is the slope of the luminosity function at the flux limit, and $\mathcal{M}$ is the gauge-invariant expression for the magnification [22, 24].

4.3 Galaxy bias in a relativistic context

In order to make progress from Eq. (118), we need to relate the intrinsic galaxy overdensity $\delta_g^{\rm sc}$ to the matter and metric perturbations. Specifically, we need to relate $\delta_g^{\rm sc}(\mathbf{x}, t)$ to $\delta_m^{\rm sc}(\mathbf{x}, t)$ in synchronous-comoving gauge. The key property of synchronous-comoving gauge is that *constant-t surfaces coincide with constant proper time surfaces.*[6] That is, we can directly apply the results of Sec. 2, which provide the relation between statistics of $\delta_g^{\rm sc}$ and $\delta_m^{\rm sc}$ on fixed proper-time surfaces. Thus, on the very large scales on the GR corrections become relevant, we can effectively use the linear bias relation

$$\delta_g^{\rm sc}(\mathbf{x}, t) = b(t)\delta_m^{\rm sc}(\mathbf{x}, t)\,, \tag{119}$$

keeping in mind that this is to be seen only schematically, since the actual observed bias parameter $b(t)$ is the renormalized bias parameter which enters the relation between galaxy and matter *correlations* (rather than point-by-point).

We can expand on this issue a bit further and show explicitly that this definition of bias is gauge-invariant. Consider some large spatial volume within the Universe centered around the spacetime point x_p^μ, on a constant-age hypersurface, t_U =constant, where t_U denotes the proper time of comoving observers since the Big Bang. Then, the number of galaxies (or, more generally, tracers) within that volume can only depend on the enclosed mass M, and the age of the Universe in that volume t_U which is being kept fixed:

$$N_g = F(M;\, t_U;\, x_p^\mu). \tag{120}$$

Here, the explicit dependence on x_p^μ indicates any stochasticity in the relation between N_g and the local density and age. We now assume that the volume V is large enough so that linear perturbation theory applies. Then, in a given coordinate system $(\tau, \mathbf{x})$, the enclosed mass is given by

$$\begin{aligned}M &= \int_V \rho = \overline{\rho}(\tau)[1 + \delta \ln \rho]V \\ &= \overline{\rho}(\tau)[1 + \delta_m - 3aH\delta\tau]V.\end{aligned} \tag{121}$$

[6] There are corrections to this statement at order $(v/c)^2$, but velocities and potential wells in LSS correspond to $(v/c)^2 < 10^{-4}$.

Here, $\overline{\rho}(\tau)$ is the average (physical) matter density in the background (equivalent to ρ averaged over the entire constant-coordinate-time hypersurface), while δ_m is the matter density perturbation on a *constant-coordinate-time* hypersurface. The second line follows from $\delta \ln \rho \equiv \rho/\overline{\rho} - 1$ and $d \ln \overline{\rho}/d\tau = -3aH$, and we have defined $-\delta\tau(\mathbf{x})$ to be the displacement in coordinate time corresponding to a t_U = constant hypersurface. Thus, the term $-3aH\delta\tau$ in Eq. (121) comes in from going from a constant-age hypersurface to a constant-coordinate-time hypersurface. Note that in Eq. (121) the perturbations are to be considered averaged over the volume V.

In exactly the same way, we can define the average (physical) galaxy number density $\bar{n}_g$ on constant-coordinate-time hypersurfaces. The same reasoning leading to Eq. (121) yields

$$\begin{aligned} N_g &= \int_V n_g = \bar{n}_g(\tau)[1 + \delta \ln n_g]V \\ &= \bar{n}_g(\tau)\left[1 + \delta_g + b_{e,p}\, aH\delta\tau\right] V, \end{aligned} \tag{122}$$

where $b_{e,p} = d \ln \bar{n}_g / d \ln a$. We can now equate this to our general ansatz Eq. (120),

$$\begin{aligned} N_g &= F(M;\, t_U;\, x_p^\mu) \\ &= \bar{F}(\overline{\rho}V; t_U)[1 + b(\delta_m - 3aH\delta\tau) + \varepsilon] \\ &= \bar{n}_g(\tau)\left[1 + \delta_g + b_{e,p}\, aH\delta\tau + \varepsilon\right] V. \end{aligned} \tag{123}$$

Here, we have defined $\bar{F}(M, t_U) \equiv \langle F(M, t_U, x_p^\mu)\rangle_{t_U}$, and introduced the bias

$$b \equiv \frac{\partial \ln \bar{F}(M; t_U)}{\partial \ln M}\Big|_{\overline{\rho}V} = \frac{\partial \ln \bar{F}(\rho_V V; t_U)}{\partial \ln \rho_V}\Big|_{\overline{\rho}} \tag{124}$$

and the stochastic contribution to galaxy density

$$\varepsilon(x^\mu) = \frac{F(M, t_U, x^\mu)}{\bar{F}(M, t_U)} - 1. \tag{125}$$

Also ρ_V denotes the average matter density (on the t_U = const slice) within the volume V.

At first order, the bias b defined in this way is only a function of τ (equivalent to t_U at leading order). The stochastic contribution ε to galaxy clustering is only a function of the spacetime point (ε is here considered to be first order as well). In the background ($\delta_m \to 0, \delta\tau \to 0$), Eq. (123) implies, not surprisingly, $\bar{n}_g(\tau)V = \bar{F}(\overline{\rho}V; \tau)$. To first order in the perturbations, recall that Eq. (123) must hold in *any* coordinate system. Thus, we conclude that the galaxy density perturbation is given in general by

$$\begin{aligned} \delta_g(x^\mu) = b(\tau)[\delta_m(x^\mu) - 3aH(\tau)\delta\tau(x^\mu)] \\ - b_{e,p}(\tau)\, aH(\tau)\delta\tau(x^\mu) + \varepsilon(x^\mu). \end{aligned} \tag{126}$$

On sub-horizon scales, $aH\delta\tau$ becomes negligible compared to δ_m (for standard choices of gauge). Eq. (126) shows that in this limit we recover the usual bias relation $\delta_g = b\delta_m + \varepsilon$. Furthermore, if we choose synchronous gauge where all comoving observers are synchronized and thus $\delta\tau = 0$, the linear bias relation holds on *all* scales. Note that the definition Eq. (124) is precisely what we called the renormalized bias parameter in Sec. 2.3. One can show very easily that the bias relation Eq. (126) holds in all gauges, and that the bias parameters b and b_e do not depend on the gauge choice [22]. This is not very surprising, since these parameters are in principle observable: b [Eq. (124)] quantifies the response of the galaxy number in a given volume at fixed age of the Universe to a change in the average mass density (or enclosed mass) within this volume; b_e [Eq. (113)] quantifies the dependence of the average (background) number density of galaxies on the age of the Universe.

This derivation was phrased in terms of the physical galaxy density. The reasoning and Eq. (126) trivially hold for the comoving galaxy density $a^3 n_g$ as well, the only difference being that δ_g is now the fractional perturbation in comoving number density, and $b_{e,p}$ is replaced with Eq. (113),

$$b_e = \frac{d \ln(a^3 \bar{n}_g)}{d \ln a}.$$

4.4 Observed galaxy power spectrum

From Eq. (118), assuming a smooth Dark Energy cosmology to express the metric perturbations in terms of the matter density, we obtain a Fourier-space expression for the observed galaxy density overdensity $\tilde{\delta}_g$ in terms of the matter density contrast in synchronous-comoving gauge $\delta_m^{\rm sc}$:

$$\frac{\tilde{\delta}_g(\mathbf{k})}{\delta_m^{\rm sc}(\mathbf{k})} = b + f\mu^2 + \frac{C_\phi}{x^2} + \frac{i\mu}{x}C_v, \tag{127}$$

where $x \equiv k/aH$ is the wavenumber in units of the comoving horizon, μ is the cosine of $\mathbf{k}$ with the line of sight, and we have defined the coefficients

$$\begin{aligned} C_\phi &= \frac{3}{2}\Omega_m\left[b_e\left(1-\frac{2f}{3\Omega_m}\right)+1+\frac{2f}{\Omega_m}+F-f-2\mathcal{Q}\right] \\ C_v &= f\left[b_e+F-1\right] \\ F &= \frac{1+z}{H}\frac{dH}{dz}-\frac{1+z}{H}\frac{2}{\tilde{\chi}}(1-\mathcal{Q})-2\mathcal{Q}\,. \end{aligned} \tag{128}$$

Note that here $\Omega_m = \Omega_m(a) = \overline{\rho}(a)/\rho_{\rm cr}(a)$. For given value of b_e and $\mathcal{Q}$, C_ϕ and C_v are only functions of redshift, and incorporate the relativistic bias as well as the volume distortion and magnification effects. On small scales, when $x \gg 1$, we recover the usual Fourier-space galaxy overdensity in "Newtonian theory",

$$\tilde{\delta}_g(\mathbf{k}) \overset{x\gg 1}{=} (b + f\mu^2)\delta_m^{\rm sc}(\mathbf{k}). \tag{129}$$

Note that Eq. (127) has similarity to what we expect from the scale-dependent bias from local primordial non-Gaussianity in the "Newtonian" approximation:

$$\tilde{\delta}_g(\mathbf{k}) \overset{x\gg 1}{=} \left[b + f\mu^2 + 2f_{\rm NL}(b-1)\delta_c\left(\frac{3}{2}\frac{\Omega_m}{T(k)g(z)}\right)\frac{1}{x^2}\right]\delta_m^{\rm sc}(\mathbf{k}). \tag{130}$$

Thus, the relativistic corrections in Eq. (127), if not accounted for, could be interpreted as a non-zero $f_{\rm NL}$. The resulting "apparent" $f_{\rm NL}$ is shown in Fig. 3. Clearly it is of order 1, typically actually less than one. Thus, the relativistic corrections will only be important for extremely large volume surveys such as Euclid. In any case, they can easily be calculated given the survey-dependent quantities b_e and $\mathcal{Q}$.

5 Connecting with inflation: conformal Fermi frame

As we discussed in Sec. 3, the effect of primordial non-Gaussianity on the clustering of galaxies on large scales depends (to leading order) on the bispectrum of the primordial potential perturbations—specifically on the squeezed limit of the bispectrum where one mode k_L is much smaller than the other two k_1, k_2. In order to calculate the bispectrum consistently, one has to work to second order in perturbation theory throughout, that is from the inflationary epoch through reheating and radiation domination all the way to late times. Since the modes exit the horizon during inflation and reenter later, this calculation needs to be done fully relativistically. This is a very involved calculation in general and has not been fully completed to date.

We show how this obstacle is overcome when one is interested in the *squeezed limit*; that is, we limit ourselves to a particular configuration of the bispectrum in which two of the momenta, say k_1 and k_2 are much larger than the third, k_3, i.e. with $k_3 \equiv k_L$ being the *long mode* and $k_1 \sim k_2 \sim k_S$ being the *short modes*. We take advantage of the fact that different sets of coordinates are convenient in different cosmological epochs and there is a careful choice that makes the computation easy and transparent. We now present a convenient way to choose coordinates by discussing the different steps that are summarized in Fig. 4.

A first step consists in computing the primordial correlators generated during the phase of *primordial inflation*. For this purpose, coordinates are used that are comoving with respect to the

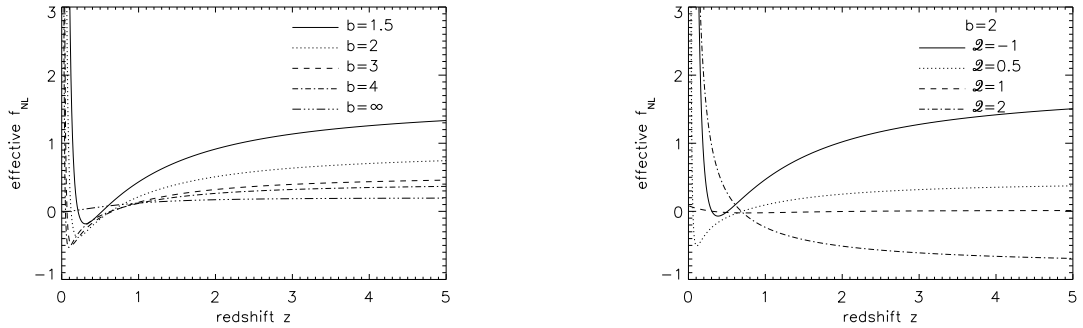

Figure 3: Apparent $f_{\rm NL}$ induced by relativistic corrections as a function of redshift for different galaxy biases when $\mathcal{Q} = 0$ (left panel), and for fixed galaxy bias ($b = 2$) with different magnification bias $\mathcal{Q} = -1$, 0.5, 1 and 2 (right panel).

Generation	$k_S < aH$	*Evolution*	$\tau_{\rm em}$	*Emission*	$\tau_{\rm obs}$	*Observation*
comoving coordinates	$\longrightarrow$	$\overline{\rm FNC}$	$\longrightarrow$ Eq. (140)	FNC(emitter)	$\longrightarrow$ cosmic rulers [24]	FNC(observer)

Figure 4: Sequence of coordinates employed in the computation of observables predicted by a primordial bispectrum in the squeezed limit. The arrows represent the change from one set of coordinates to the next, indicating when this transformation is most conveniently performed and with a reference to the relevant equations. $\tau_{\rm em}$ and $\tau_{\rm obs}$ denote the conformal times at which the observed photons were emitted and observed, respectively.

unperturbed inflating background (see e.g. [25]). A second step consists in describing the local *late-time physical processes*, such as for example the decoupling of photons during recombination or the formation of a dark matter halo, that lead to the generation of some light signal that eventually will reach the Earth. Since these processes take place in regions that are smaller in size than the curvature of the background metric (horizon scale H^{-1}), they are best described in *Fermi Normal Coordinates* (FNC [26]): this is the unique frame (up to three Euler angles) constructed along a timelike geodesic passing through a point P in which the metric is Minkowski with corrections going as the spatial distance r_F from the central geodesic squared. In an unperturbed FRW Universe, the corrections thus scale as $(Hr_F)^2$, while in the presence of a metric perturbation $h_{\mu\nu}$ the corrections involve second derivatives of $h_{\mu\nu}$ (in particular, a constant or pure gradient metric perturbation is removed on all scales). The last step consists in relating the signals from the point of emission, for example the last scattering surface, or the position of a halo, to the *observables measured on Earth*, where scientists most naturally use FNC centered on the Earth's world line.

Then, we make use of the fact that in a $k_L/k_S \ll 1$ expansion, the leading and next-to-leading order gravitational effects of the long-wavelength perturbation k_L on the short modes can be removed by transforming to the local FNC frame, as was emphasized in [27, 28, 29, 30]. The first non-trivial

effect comes from the second spatial and time derivatives of the perturbation, which are suppressed by $\mathcal{O}(k_L^2/k_S^2)$ (see Sec. 5.1 for a more precise statement). We will ignore terms of this order in this paper, as they correspond to actual physical interactions, which will need to be treated in detail. However, up to this order, in the local FNC frame we can simply use linear theory. The non-Gaussian correlation will then be captured by averaging over an ensemble of FNC, each constructed to remove a different long mode.

In going from comoving coordinates to FNC, one subtlety arises from the fact that FNC are affected not only by perturbations but also by the expansion of the Universe. Because of this, no matter when the change of coordinates is performed, when using FNC we cannot avoid a period when the evolution at second order in perturbations becomes relevant. A nice way out of this complication is to use *conformal Fermi Normal Coordinates* ($\overline{\text{FNC}}$), in which the metric is locally in the FLRW form, i.e. a homogeneous isotropic expanding Universe, rather than in the Minkowski form. These coordinates are valid up to a scale k_L^{-1}, where k_L denotes the wavenumber of the long-wavelength perturbation considered. At late times, the small-wavelength modes are well inside the horizon and have in general been subject to non-linear evolution. Then, transforming from the $\overline{\text{FNC}}$ frame to the usual FNC frame at a given spacetime point simply corresponds to a rescaling of the spatial coordinates.

To visualize the advantage of $\overline{\text{FNC}}$, consider Fig. 5. Each one of the four panels shows the Hubble scale (solid circle) and the region of validity of the $\overline{\text{FNC}}$ (dashed line) at a different moment in time. The long (red line) and short (blue line) modes are also shown for comparison. One can immediately appreciate the fact that from very early time when the long mode is still inside (or about to leave) the horizon all the way until late time when the long mode has re-entered the horizon, the region of validity of $\overline{\text{FNC}}$ amply encompasses the short modes. This is in contrast with standard FNC for which the presence of the Hubble scale makes this impossible. The primordial correlators, typically computed in comoving coordinates, take a simple form once both the short and long modes have left the horizon and cease to evolve. Any time after that moment and before the small-scale modes reenter the horizon constitutes a good time to transform from global conformal coordinates to $\overline{\text{FNC}}$.

The final step is to relate quantities defined in the local FNC frame (e.g. temperature of the gas, or halo mass) to the photons actually measured on Earth. This involves the photon propagation ("projection") effects, such as lensing and redshift-space distortions [31, 32, 33, 34, 24, 35]. Here, we will adopt the "standard ruler" approach [24, 23] since it conveniently encompasses all projection effects for various observables. Specifically, the derivation in [24] assumes that a ruler, such as a $\xi(r) = \text{const}$ contour, corresponds to a fixed physical spatial scale on a constant-proper-time hypersurface for a comoving observer, equivalent to a fixed spatial scale in FNC at emission.

Finally, at leading order, any *physical* (and necessarily non-gravitational) correlations between long-wavelength modes and small scale modes imprinted at early times and present in $\overline{\text{FNC}}$ then simply add to these projection effects. The framework discussed above, which we summarize in Fig. 4, thus connects the bispectrum calculated in any convenient gauge during inflation with late-time observables such as the CMB bispectrum or the large-scale clustering of tracers.

5.1 Conformal Fermi Coordinates

Let us assume we are given the primordial correlations during inflation in some gauge, where the metric is written as

$$ds^2 = a^2(\tau)\left[\eta_{\mu\nu} + h_{\mu\nu}\right] dx^\mu dx^\nu = a^2(\tau)\bar{g}_{\mu\nu} dx^\mu dx^\nu \,. \tag{131}$$

Here we have denoted the conformal metric with a bar. In the absence of perturbations, $h_{\mu\nu} = 0$ and $\bar{g}_{\mu\nu} = \eta_{\mu\nu}$. One should think of Eq. (131) as coarse-grained over a scale much larger than the small-scale perturbations of interest; i.e. Eq. (131) only contains the long modes, and consequently we will work up to linear order in metric perturbations.

As in the usual Fermi normal coordinate construction, we consider a central timelike geodesic of a comoving observer. However, instead of constructing the FNC with respect to $g_{\mu\nu}$, we construct *conformal Fermi normal coordinates* ($\overline{\text{FNC}}$) with respect to $\bar{g}_{\mu\nu}$. That is, within a region around the central geodesic, the metric in these coordinates is approximately $\bar{g}^F_{\mu\nu} = a^2(\tau_F)\eta_{\mu\nu}$. The corrections to this metric, which determine the size of the region within which $\bar{g}^F_{\mu\nu}$ is approximately Minkowski

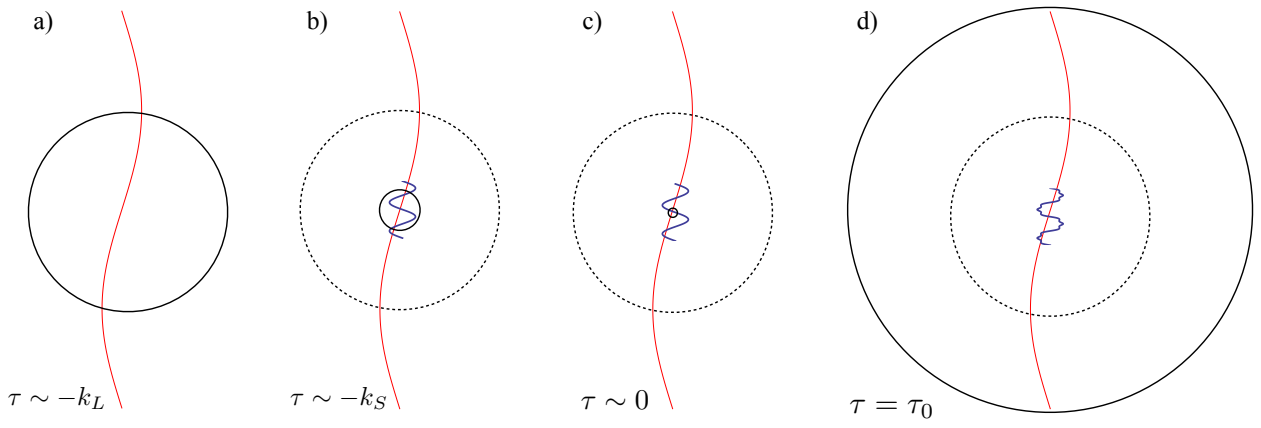

Figure 5: Illustration of the $\overline{\text{FNC}}$ patch throughout cosmic history, in comoving units. The wavy lines indicate perturbations. The solid circles denote the comoving horizon $1/aH$, which coincides with the size of the usual Fermi coordinate patch during inflation. The dashed circles denote the size of the $\overline{\text{FNC}}$ patch, i.e. the region within which the metric is of the form $\bar{g}^F_{\mu\nu} = a^2\eta_{\mu\nu}$ with small corrections. a) During inflation, when the long wavelength perturbation $h_{ij}(\mathbf{k}_L)$ is generated (i.e. leaves the horizon), thin red line. b) Later on during inflation; $h_{ij}(\mathbf{k}_L)$ is far outside the horizon when the short wavelength modes $\vartheta(\mathbf{k})$ are generated (thick blue line). The perturbations are generated well within the $\overline{\text{FNC}}$ patch corresponding to the long-wavelength mode. c) Near the end of inflation. All perturbations are far outside the horizon. Nevertheless, the small-scale modes are still well within the $\overline{\text{FNC}}$ patch. d) At observation time, after the long-wavelength mode has reentered the horizon (solid circle indicating present horizon). The $\overline{\text{FNC}}$ patch now coincides with the usual Fermi coordinate patch, which is much larger than the small-scale modes which have been processed by nonlinear evolution since horizon entry (distorted thick wavy line).

("$\overline{\text{FNC}}$ patch"), are given by second derivatives of the metric perturbations. Correspondingly, the size of the $\overline{\text{FNC}}$ patch is linked to the wavelength of the metric perturbations considered. In the following, we will be interested in (long-wavelength) perturbations with wavenumber k_L, so that the size of the $\overline{\text{FNC}}$ patch is of order k_L^{-1}. We can then schematically write the $\overline{\text{FNC}}$ frame metric as

$$\bar{g}^F_{\mu\nu} = \eta_{\mu\nu} + \mathcal{O}\left([h''_{ij}, \partial_j h'_{\mu i}, \partial_i\partial_j h_{\mu\nu}]\bar{x}_F^2\right). \tag{132}$$

Let us disregard the corrections $\propto h''_{ij}$, $\partial_j h'_{\mu i}$ for the moment. The remaining corrections to the conformal Fermi frame metric scale as second spatial derivatives of the metric perturbation multiplied by the spatial coordinate squared. Thus, instead of being order $(Hx_F)^2$ as the usual FNC corrections, they are of order $(k_L\bar{x}_F)^2$, where k_L is the comoving wavenumber of the perturbation (in slight abuse of notation, we drop the bar over k since we will only be dealing with comoving Fourier wavenumbers). Within the $\overline{\text{FNC}}$ patch of size $\sim k_L^{-1}$, we can remove the effects of the long-mode perturbation (in comoving coordinates) *at all times* — through horizon exit and reentry.

This very useful result only holds if h''_{ij}, $\partial_j h'_{\mu i}$ are of order $\partial_i\partial_j h_{\mu\nu}$ or smaller, which is the case for all models of single field inflation in which the background is an attractor solution. In multifield models where more than one degree of freedom is active during inflation, this is not the case as h evolves outside the horizon. Then, $\bar{g}^F_{\mu\nu} = \eta_{\mu\nu}$ is *not* a good approximation. Effectively, small-scale perturbations are generated in a different inflating background $a(\tau) \to a(\tau; h)$ in the presence of a long-wavelenth perturbation, leading to a coupling between long- and short-scale modes.

5.2 Bispectrum and consistency relation

Consider the bispectrum of the curvature perturbation in comoving gauge, ζ. In single-field inflation, ζ is conserved outside the horizon, and can be straightforwardly related to the late-time potential perturbation Φ. This implies that *all* corrections to the $\overline{\text{FNC}}$ frame metric, including those of order ζ', ζ'', are order k_L^2. In comoving gauge, the scalar long-wavelength contribution to the metric perturbation

$h_{\mu\nu}$ is

$$h_{00} = -2N_1 = \mathcal{O}(k_L^2)$$
$$h_{ij} = 2\zeta_L\,\delta_{ij}\,. \tag{133}$$

The transformation to conformal FNC frame is then given by a purely spatial coordinate transform:

$$\bar{x}_F^i = x^i + \frac{1}{2}h^i{}_j x^j = \left(1 + \frac{1}{2}\zeta_L\right) x^i\,. \tag{134}$$

As shown in [36], one can transform the squeezed-limit bispectrum in comoving gauge to the $\overline{\text{FNC}}$ frame. We denote the curvature perturbation in $\overline{\text{FNC}}$ frame as $\bar{\zeta}$. The squeezed-limit bispectrum then transforms as

$$B_{\bar{\zeta}\bar{\zeta}\bar{\zeta}}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) = P_\zeta(k_L)P_\zeta(k_S)\frac{\partial \ln(k_S^3 P_\zeta(k_S))}{\partial \ln k_S} + B_{\zeta\zeta\zeta}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2), \tag{135}$$

where $B_{\zeta\zeta\zeta}$ is the three-point function calculated in comoving gauge, and $\mathbf{k}_S = \mathbf{k}_1 + \mathbf{k}_L/2$. This encodes the effect of the rescaling of spatial coordinates given in Eq. (134). As shown in [25, 37], this is in the squeezed limit given by

$$B_{\zeta\zeta\zeta}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) = -[n_s - 1]P_\zeta(k_L)P_\zeta(k_S) + \mathcal{O}\left(\frac{k_L}{k_S}\right)^2, \tag{136}$$

where n_s is defined through $P_\zeta(k) \propto k^{-4+n_s}$. Eq. (136) is usually referred to as the *consistency relation*. We now see that the first term in Eq. (135) exactly cancels the contribution from $B_{\zeta\zeta\zeta}$, leading to

$$B_{\bar{\zeta}\bar{\zeta}\bar{\zeta}}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2) = \mathcal{O}\left(\frac{k_L}{k_S}\right)^2. \tag{137}$$

We thus conclude that *in single-field inflation, the bispectrum in the squeezed limit is zero in the conformal Fermi frame, with corrections going as* $(k_L/k_S)^2$. The same can be shown to hold for the tensor-scalar-scalar bispectrum.

5.3 Connection to late-time observables

As discussed above, we can follow the $\overline{\text{FNC}}$ patch all the way through the end of inflation and horizon re-entry of both long and short modes, provided that perturbations do not evolve significantly when they are outside the horizon. In this section, we show how the squeezed-limit correlations in the $\overline{\text{FNC}}$ frame can be related to observations made from Earth today.

Let us assume we observe correlations of some field $\delta(\mathbf{x})$ at late times, i.e. at or after recombination. In linear theory, assuming adiabatic initial conditions, we can relate δ to the curvature perturbation through some transfer function α,

$$\delta(\mathbf{k}_S, \tau) = \alpha(k_S, \tau)\zeta(\mathbf{k}_S). \tag{138}$$

For matter density perturbations, $\alpha = 3/5\mathcal{M}$ where $\mathcal{M}$ is defined in Eq. (55) above. Similarly, we assume the long-wavelength perturbation is evolved with some transfer function, which will only be a function of time up to corrections of order k_L^2. Since ζ does not evolve outside the horizon, the conformal Fermi coordinate patch, i.e. the region over which corrections to the $\overline{\text{FNC}}$ metric are small, is constant in (comoving) size throughout horizon exit and reentry of the short-wavelength modes. Thus, at the conformal time τ_{em} at which the photons we observe today were emitted, the metric in $\overline{\text{FNC}}$ is

$$\bar{g}^F_{\mu\nu} = a^2(\tau_{\text{em}})\eta_{\mu\nu} + \mathcal{O}(\partial_i\partial_j\zeta)\,. \tag{139}$$

If we transform coordinates through

$$x_F^0 = \int^\tau a(\tau')d\tau' \equiv t(\tau)$$
$$x_F^i = a(\tau)\bar{x}_F^i, \tag{140}$$

the metric in the coordinates $\{x_F^\mu\}$ becomes

$$g_{\mu\nu}^F = \eta_{\mu\nu} + \mathcal{O}[(Hx_F^i)^2, (\partial_k\partial_l\zeta)x_F^2]. \tag{141}$$

In other words, x_F^μ are the usual Fermi coordinates (FNC) defined around the same timelike geodesic as the $\overline{\text{FNC}}$. Depending on whether the long-wavelength modes for which we constructed the $\overline{\text{FNC}}$ patch have entered the horizon, either the order H or the order $\partial\partial\zeta$ corrections will be dominant; however, this is not relevant for the discussion that follows. Eq. (140) corresponds to a rescaling of the time coordinate, leaving the timelike unit vector and τ = const hypersurfaces unchanged, and a time-dependent rescaling of the spatial coordinates. Since the spatial rescaling is the same everywhere on a x_F^0 = const hypersurface, this implies that, for models that obey the consistency condition, Eq. (137) is still valid at late times in the FNC frame:

$$B_{\zeta_F\delta_F\delta_F}(\mathbf{k}_L, \mathbf{k}_1, \mathbf{k}_2; \tau_{\rm em}) = \mathcal{O}\left(\frac{k_L}{k_S}\right)^2, \tag{142}$$

where ζ_F, δ_F denote perturbations in the FNC frame.

Recall that the squeezed limit of three-point functions corresponds to the modulation of local two-point functions by long-wavelength perturbations. Thus, another way of phrasing this result is that a surface of constant correlation in the FNC frame, $\xi_{\delta_F}(r_F)$ = const, defines a standard ruler—a fixed spatial scale—as considered in [24], with corrections proportional to second derivatives of $h_{\mu\nu}$ only. Hence, in standard single-field inflation and any other case where Eq. (137) holds, the ruler scale r_F is statistically the same everywhere on a x_F^0 = const hypersurface, and there is no correlation with long-wavelength perturbations. In particular, for the LSS effects discussed in Sec. 3, there is *no scale-dependent bias* $\propto k^{-2}$ *in single-field inflation*, not even the very small amount naively predicted by the consistency relation [Eq. (136)], $f_{\rm NL} \sim -(n_s - 1)/2$.[7]

The *apparent* correlations induced between long-wavelength modes and small-scale correlations are then given by the ruler perturbations derived in [24]. As shown in [36], we can use their results to derive, with very little algebra, exact expressions for the observed squeezed-limit three-point functions e.g. of the CMB in single-field inflation.

References

[1] A. S. Szalay, Astrophys. J.**333**, 21 (1988).

[2] J. N. Fry and E. Gaztanaga, Astrophys. J.**413**, 447 (1993), [arXiv:astro-ph/9302009].

[3] P. Coles, Mon. Not. R. Astron. Soc.**262**, 1065 (1993).

[4] A. Dekel and O. Lahav, Astrophys. J.**520**, 24 (1999), [arXiv:astro-ph/9806193].

[5] F. Schmidt, D. Jeong and V. Desjacques, ArXiv e-prints (2012), [arXiv:1212.0868].

[6] T. Matsubara, Phys. Rev. D**83**, 083518 (2011), [arXiv:1102.4619].

[7] W. H. Press and P. Schechter, Astrophys. J.**187**, 425 (1974).

[8] V. Desjacques, Phys. Rev. D**78**, 103503 (2008), [arXiv:0806.0007].

[9] T. Baldauf, U. Seljak, V. Desjacques and P. McDonald, ArXiv e-prints (2012), [arXiv:1201.4827].

[7] This contribution can in fact become large for models with oscillations or features in the power spectrum [38]. Even in these cases however, the scale-dependent bias vanishes on large scales.

[10] K. C. Chan, R. Scoccimarro and R. K. Sheth, Phys. Rev. D**85**, 083509 (2012), [arXiv:1201.3614].

[11] T. Giannantonio and C. Porciani, Phys. Rev. D**81**, 063530 (2010), [arXiv:0911.0017].

[12] F. Schmidt and M. Kamionkowski, Phys. Rev. D**82**, 103002 (2010), [arXiv:1008.0638].

[13] R. Scoccimarro, L. Hui, M. Manera and K. C. Chan, Phys. Rev. D**85**, 083002 (2012), [arXiv:1108.5512].

[14] P. McDonald, Phys. Rev. D**78**, 123519 (2008), [arXiv:0806.1061].

[15] T. Baldauf, U. Seljak, L. Senatore and M. Zaldarriaga, ArXiv e-prints (2011), [arXiv:1106.5507].

[16] F. Schmidt, ArXiv e-prints (2013), [arXiv:1304.1817].

[17] K. M. Smith and M. Zaldarriaga, Mon. Not. R. Astron. Soc.**417**, 2 (2011), [arXiv:astro-ph/0612571].

[18] V. Desjacques, D. Jeong and F. Schmidt, Phys. Rev. D**84**, 063512 (2011), [arXiv:1105.3628].

[19] N. Dalal, O. Doré, D. Huterer and A. Shirokov, Phys. Rev. D**77**, 123514 (2008), [arXiv:0710.4560].

[20] A. Slosar, C. Hirata, U. Seljak, S. Ho and N. Padmanabhan, JCAP **8**, 31 (2008), [arXiv:0805.3580].

[21] S. Dodelson, F. Schmidt and A. Vallinotto, Phys. Rev. D**78**, 043508 (2008), [arXiv:0806.0331].

[22] D. Jeong, F. Schmidt and C. M. Hirata, Phys. Rev. D**85**, 023504 (2012), [arXiv:1107.5427].

[23] D. Jeong and F. Schmidt, ArXiv e-prints (2013), [arXiv:1305.1299].

[24] F. Schmidt and D. Jeong, Phys. Rev. D**86**, 083527 (2012), [arXiv:1204.3625].

[25] J. Maldacena, Journal of High Energy Physics **5**, 13 (2003), [arXiv:astro-ph/0210603].

[26] F. K. Manasse and C. W. Misner, Journal of Mathematical Physics **4**, 735 (1963).

[27] T. Baldauf, U. Seljak, L. Senatore and M. Zaldarriaga, ArXiv e-prints (2011), [arXiv:1106.5507].

[28] P. Creminelli, J. Noreña and M. Simonović, JCAP **7**, 52 (2012), [arXiv:1203.4595].

[29] L. Senatore and M. Zaldarriaga, arXiv:1210.6048.

[30] L. Senatore and M. Zaldarriaga, JCAP **1208**, 001 (2012), [arXiv:1203.6884].

[31] P. Creminelli and M. Zaldarriaga, Phys. Rev. D**70**, 083532 (2004), [arXiv:astro-ph/0405428].

[32] L. Boubekeur, P. Creminelli, G. D'Amico, J. Noreña and F. Vernizzi, JCAP **8**, 29 (2009), [arXiv:0906.0980].

[33] P. Creminelli, C. Pitrou and F. Vernizzi, JCAP **11**, 25 (2011), [arXiv:1109.1822].

[34] N. Bartolo, S. Matarrese and A. Riotto, JCAP **1202**, 017 (2012), [arXiv:1109.2043].

[35] A. Lewis, JCAP **1206**, 023 (2012), [arXiv:1204.5018].

[36] E. Pajer, F. Schmidt and M. Zaldarriaga, ArXiv e-prints (2013), [arXiv:1305.0824].

[37] P. Creminelli, G. D'Amico, M. Musso and J. Noreña, JCAP **11**, 38 (2011), [arXiv:1106.1462].

[38] F.-Y. Cyr-Racine and F. Schmidt, Phys.Rev. **D84**, 083505 (2011), [arXiv:1106.2806].