

Symmetries, Invariants and Generating Functions: Higher-order Statistics of Biased Tracers

Dipak Munshi

Astronomy Centre, School of Mathematical and Physical Sciences,
University of Sussex, Brighton BN1 9QH, U.K.

E-mail: D.Munshi@sussex.ac.uk

Abstract. Gravitationally collapsed objects are known to be biased tracers of an underlying density contrast. Using symmetry arguments, generalised biasing schemes have recently been developed to relate the halo density contrast δ_h with the underlying density contrast δ , divergence of velocity θ and their higher-order derivatives. This is done by constructing invariants such as s, t, ψ, η . We show how the generating function formalism in Eulerian standard perturbation theory (SPT) can be used to show that many of the additional terms based on extended Galilean and Lifshitz symmetry actually do *not* make any contribution to the higher-order statistics of biased tracers. Other terms can also be drastically simplified allowing us to write the vertices associated with δ_h in terms of the vertices of δ and θ , the higher-order derivatives and the bias coefficients. We also compute the cumulant correlators (CCs) for two different tracer populations. These perturbative results are valid for *tree-level* contributions but at an *arbitrary order*. We also take into account the stochastic nature bias in our analysis. Extending previous results of a local polynomial model of bias, we express the one-point cumulants \mathcal{S}_N and their two-point counterparts, the CCs i.e. \mathcal{C}_{pq} , of biased tracers in terms of that of their underlying density contrast counterparts. As a by-product of our calculation we also discuss the results using approximations based on Lagrangian perturbation theory (LPT).

Keywords: Cosmology, Large-Scale Structure, Perturbation Theory

Contents

1	Introduction	1
2	Generating Functions	2
3	Bias and Biased Tracers	3
4	Scale-Dependent Bias	7
5	Cumulants and Cumulant Correlators	8
6	Summary and Outlook	10
A	Symmetries and Generating Function	14
B	Bias and Lagrangian perturbation theory	14

“Why do you want to know his bias? Form your own bias!” R.P.Feynman

1 Introduction

Recently completed cosmic microwave background (CMB) surveys e.g. the Planck surveyor¹, have provided us with a robust cosmological framework that will allow us to investigate the physics beyond the *Standard Model* of cosmology. Next generation surveys are mapping the entire CMB sky with higher resolution and accuracy e.g. ACT² and SPT³ to answer many of the questions relevant to structure formation in the low redshift Universe. In addition, the ongoing and future large scale surveys will map the sky with ever increasing precision (BOSS⁴ [1], WiggleZ⁵ [2], DES⁶ [3], EUCLID⁷ [4]). These surveys will provide a glimpse of physics beyond the standard model. On one hand, they will check any departure from general relativity (GR) on cosmological scales, on the other, they will also provide an estimate of the sum of the neutrino mass [5]. The galaxies, however, are known to be biased tracers of underlying dark matter distribution [53]. To achieve the full potential of the future surveys, parametrization and understanding of bias is of utmost importance.

Initial models of bias were linear and local relations between the tracer’s density contrast δ_h and the underlying matter density contrast δ : $\delta_h = b_1\delta$. However, early simulations e.g. [8] pointed to a more complex nature of bias that can be nonlinear and non-local. It was also realized that bias can also be stochastic. On the theoretical side, both perturbative and non-perturbative models started to emerge. The non-perturbative theories based on peak approach were developed in [9]; these theories can mimic many of the properties of galaxy

¹Planck: <http://www.cosmos.esa.int/web/planck/>

²ACT: <http://www.physics.princeton.edu/act/>

³SPT: <http://pole.uchicago.edu>

⁴Baryon Oscillator Spectroscopic Survey: <http://www.sdss3.org/surveys/boss.php>

⁵WiggleZ Survey : <http://wigglez.swin.edu.au/>

⁶Dark Energy Survey: <http://www.darkenergysurvey.org/>

⁷EUCLID: <http://www.euclid-ec.org/>

bias successfully. Halo based approaches were developed which remain an important tool for making analytical predictions [10].

Many recent approaches have also seen development of formal perturbative approaches that extend and put known results on a solid foundation [11–13]. The effective field theory based approaches were also developed [14, 15]. Indeed, the polynomial model for bias developed in [16] was valid for large smoothing scales and lacks the non-local terms that are generated due to gravitational clustering. Using just second-order standard perturbation theory (SPT) it has been shown that gravitational evolution is responsible for generating tidal interactions which are non-local in the density field [17–19]. Similar non-local contribution is also expected in the clustering of halos and was studied in detail in several publications [20, 21]. In [22] a more generic scheme for bias was developed based on symmetries inherent in the dynamical equations. In the nonlinear regime gravity-induced bias has been studied in the context of the hierarchical ansatz (HA) [23, 24]. It has recently been pointed out that non-local bias can mimic scale-dependent suppression of growth of perturbation in cosmologies with massive neutrinos [25–30]. The non-local bias is relatively small compared to the linear bias but future large-scale surveys will be sensitive to them. Large-scale numerical simulations have been employed to investigate their effect on the clustering of halos [31–34]. It was shown that scale-dependent bias in the power spectrum of dark matter halo is degenerate with the signatures left by the various non-local bias terms. The degeneracies present in the characterization of the bias can be broken by using higher-order statistics of biased tracers. Going beyond the usual power spectrum analysis, in this paper we compute the higher-order statistics for a generic biasing scheme which can be nonlinear, non-local and stochastic. The additional terms in the perturbative description of bias stem from symmetry considerations. The inherent symmetries in the dynamical equations predict invariant quantities [35–37] in the perturbative expansion of bias.

A generating function based approach in the perturbative regime was introduced in [38]. It provides a powerful framework to analyse the higher-order statistics of cosmological fields. We use this formalism along with the functional relationship dictated by a biasing scheme of δ_h with δ and θ to compute the tree-level vertices of δ_h as a function of those of δ and θ . These derivations are valid in the perturbative regime. We derive the formal relations in the presence of nonlinear, non-local and higher-derivative terms. We also include a stochastic noise that originates from our lack of knowledge of the fundamental physics related to the galaxy formation process. We use these expressions to decide which terms in these expressions do not contribute at any order. The results from the generating function formalism are next used to express the cumulants and cumulant-correlators of the biased tracers. Our aim is to extend the results presented in [39] to more general biasing schemes.

This paper is organised as follows. In §2 we review the generating function approach. In §3 we consider a family of generalized bias models and use the generating function to analyse them. §4 is devoted to discussion of a non-local bias. §5 is devoted to the discussion of cumulants and cumulant correlators (CCs). Finally, our conclusions are presented in §6. Some of the details of our derivations are relegated to the two appendices.

2 Generating Functions

The generating function formalism is often used to compute the cumulants and the CCs of cosmological fields. Our aim here is to provide a very brief review of the generating function formalism developed in [38] (also see [40]) to construct the CCs of the biased tracers (halos

or peaks) and express them in terms of the statistics of underlying mass distribution. The results are relevant for the perturbative regime. The n -th order of perturbative expansion of an arbitrary field F defined as $F^{(n)}$ with respect to δ is defined as follows:

$$\langle F^{(n)} \rangle_c = \frac{\int \langle F^{(n)}(\mathbf{x}, a) \delta^{(1)}(\mathbf{x}_1, a) \cdots \delta^{(1)}(\mathbf{x}_n, a) \rangle_c d^3\mathbf{x} d^3\mathbf{x}_1 \cdots d^3\mathbf{x}_n}{(\int \langle \delta^{(1)}(\mathbf{x}, a) \delta^{(1)}(\mathbf{x}', a) \rangle_c d^3\mathbf{x} d^3\mathbf{x}')^n}. \quad (2.1)$$

Here, $\delta^{(1)}(\mathbf{x}, a)$ is the linear approximation for $\delta(\mathbf{x}, a)$ at a comoving position \mathbf{x} and $a(t)$ is the scale factor of the Universe. Only connected diagrams are taken into account, which explains the subscript c . Throughout, we will assume that the initial density contrast δ is Gaussian, though it is possible to incorporate non-Gaussian initial condition. The generating function $\mathcal{G}_F(\tau_s)$ for the vertices for any random field $F(\mathbf{x}, a)$ is given by:

$$\mathcal{G}_F(\tau_s) = \sum_{n=1}^{\infty} \frac{\langle F^{(n)} \rangle_c}{n!} \tau_s^n. \quad (2.2)$$

For two arbitrary fields $A(\mathbf{x}, a)$ and $B(\mathbf{x}, a)$, we have the following properties for the generating functions [38]:

$$\mathcal{G}_{A+B}(\tau_s) = \mathcal{G}_A(\tau_s) + \mathcal{G}_B(\tau_s); \quad (2.3a)$$

$$\mathcal{G}_{AB}(\tau_s) = \mathcal{G}_A(\tau_s) \mathcal{G}_B(\tau_s); \quad (2.3b)$$

$$\mathcal{G}_{\nabla_i A \nabla_i B}(\tau_s) = 0; \quad (2.3c)$$

$$\mathcal{G}_{\nabla_i \nabla_j A \nabla_j \nabla_i B}(\tau_s) = \frac{1}{3} \mathcal{G}_{\nabla^2 A \nabla^2 B}(\tau_s). \quad (2.3d)$$

We will denote the generating function of the density contrast δ by $\mathcal{G}_\delta(\tau_s) = \sum_{n=1}^{\infty} \nu_n/n! \tau_s^n$ where $\nu_n \equiv \langle \delta^{(n)} \rangle_c$. We will also need the divergence of velocity θ (to be defined later), for which, the generating function will be denoted as $\mathcal{G}_\theta(\tau_s) \equiv \sum_{n=1}^{\infty} \mu_n/n! \tau_s^n$ with $\mu_n \equiv \langle \theta^{(n)} \rangle_c$.

The generating function formalism was originally introduced in [41] and later exploited in many publications including in [38, 42] to compute the lower order cumulants and cumulant correlators [43] by linking the generating function of tree-level amplitudes directly with the dynamical equations of a self-gravitating collisionless system [44]. The resulting expressions will be applied to understand halo clustering in §3. Notice the bias of overdense regions has been studied using the generating function formalism in [43].

3 Bias and Biased Tracers

The idea of nonlocal bias has been investigated in great detail in the past by many author. Starting from ref.[22] the idea was developed further in [35, 36]. More recently the results were extended to third order in perturbation theory in ref.[6]. The idea behind these studies is to probe the statistics of proto-halos which preserve their identity and their number density is conserved. It is assumed that, though their shape and topology may change, the center of mass of these proto-halos follow a well defined trajectory and their statistics can be studied using perturbative techniques. We will consider the matter dominated case, \mathbf{x} is the spatial comoving coordinate and τ the conformal time $\tau = \int dt/a(t)$. The associated Hubble parameter is $\mathcal{H} = d \ln a(t)/d\tau$. We will also define the divergence of velocity as $\theta = \partial_i v^i$

where $v^i = dx^i/d\tau$. The Euler, continuity and Poisson equations describe the gravitational clustering of a collisionless system in the hydrodynamic limit:

$$\frac{\partial \delta}{\partial \tau} + \nabla_i [(1 + \delta)v^i] = 0; \quad (3.1a)$$

$$\frac{\partial v^i}{\partial \tau} + \mathcal{H}v^i + v^j \nabla_j v^i = -\nabla^i \phi; \quad (3.1b)$$

$$\nabla^2 \phi = \frac{3}{2} \Omega_M \mathcal{H}^2 \delta. \quad (3.1c)$$

In terms of generating functions for δ and θ denoted as $\mathcal{G}_\delta = \sum (\nu_n/n!) \tau_s^n$ and $\mathcal{G}_\theta = \sum (\mu_n/n!) \tau_s^n$, these equations take the following forms [40]:

$$\frac{\partial \mathcal{G}_\delta}{\partial \tau} + (1 + \mathcal{G}_\theta) \mathcal{G}_\delta = 0; \quad (3.2a)$$

$$\frac{\partial \mathcal{G}_\theta}{\partial \tau} + \frac{1}{2} \mathcal{G}_\theta + \frac{1}{3} \mathcal{G}_\theta^2 + \mathcal{G}_{\nabla^2 \Phi} = 0; \quad (3.2b)$$

$$\mathcal{G}_{\nabla^2 \Phi} = \frac{3}{2} \mathcal{G}_\delta. \quad (3.2c)$$

The solution to these equations are well known and \mathcal{G}_δ and \mathcal{G}_θ satisfy collapse of spherically over-dense top-hat perturbation. However, Eq.(3.2a)-Eq.(3.2c) do not represent evolution of perturbations they encode the statistical description of an ensemble of perturbations.

To relate the tracer density δ_h with δ many different simplifying assumption are employed. It is typically assumed the number density of tracers (proto-halos) do not change and remains conserved. Thus evolution of δ_h can be described by a continuity equation. In this picture the halos can change shape or their topology but they follow a well-defined trajectories. It is further assumed that halo velocities \mathbf{v}_h are unbiased estimators of underlying dark matter velocities \mathbf{v} i.e. $\mathbf{v}_h = \mathbf{v}$:

$$(\dot{\delta}_h - \dot{\delta}) + \nabla_i [(\delta_h - \delta)v^i] = 0. \quad (3.3)$$

The overdots represent derivative w.r.t. τ . The halo density contrast $\delta_h(\mathbf{x}, \tau_i)$ and the DM density contrast $\delta(\mathbf{x}, \tau_i)$ are related at some initial time τ_i as follows:

$$\delta_h(\mathbf{x}, \tau_i) \equiv b(\delta) = \sum_{\ell} \frac{b_{\ell}^L(\tau_i)}{\ell!} [\delta(\mathbf{x}, \tau_i)]^{\ell} = \sum_{\ell} \frac{b_{\ell}^L(\tau)}{\ell!} [\delta(\mathbf{x}, \tau)]^{\ell}. \quad (3.4)$$

Thus the evolution of the Lagrangian bias $b_{\ell}^L(\tau)$ as a function of conformal time τ from $b_{\ell}^L(\tau_i)$ takes the following form: $b_{\ell}^L(\tau) = b_{\ell}^L(\tau_i) [a(\tau_i)/a(\tau)]^{\ell}$. The above expression can be used to evaluate b_{ℓ}^L at a later time τ once specified at an initial epoch τ_i . Notice that at this stage we have left the parameters b_{ℓ}^L arbitrary. Perturbative analysis of Eq.(3.3) has been carried out in an order-by-order manner. In [35, 36] an analysis was performed up to second order in the linear density contrast $\delta^{(1)}$, more recently the result was extended to third order in [6]. These studies found that the halo density contrast δ_h is related to the underlying δ through the following expression:

$$\delta_h \equiv b(\delta, \theta, \nabla_i \nabla_j \phi, \nabla_i \theta_j, \dots); \quad (3.5a)$$

$$\begin{aligned} \delta_h = & b_1 \delta + \frac{1}{2!} b_2 \delta^2 + \frac{1}{3!} b_3 \delta^3 + \frac{1}{2!} b_{s^2} s^2 + b_{\psi} \psi + b_{st} s \cdot t \\ & + b_{\nabla^2 \delta} \nabla^2 \delta + b_{\nabla^2} \nabla^2 [s_{ij} s^{ij}] + b_{\nabla^4} \nabla^2 [s_{ij}] \nabla^2 [s^{ij}] \dots \end{aligned} \quad (3.5b)$$

The coefficients b_{s^2}, b_ψ, \dots and the higher-order derivative operators $b_{\nabla^2 \delta}, b_{\nabla^2}, \dots$ appearing in Eq.(3.5b) are left arbitrary at this stage. The non-local operators t, s, η and ψ above are defined as [22]:

$$s_{ij} = \frac{2}{3\mathcal{H}^2} \nabla_i \nabla_j \phi - \frac{1}{3} \delta_{ij}^K \delta; \quad (3.6a)$$

$$t_{ij} = \partial_i v_j - \frac{1}{3} \delta_{ij} \theta - s_{ij}; \quad (3.6b)$$

$$\eta = \theta - \delta; \quad (3.6c)$$

$$\psi = \eta - \frac{2}{7} s^2 + \frac{4}{21} \delta^2. \quad (3.6d)$$

we have introduced the following notations

$$s^2 \equiv s_{ij} s^{ij}; \quad s \cdot t \equiv s^{ij} t_{ij}; \quad t^2 \equiv t_{ij} t^{ij}. \quad (3.7)$$

We have also assumed a $\Omega = 1$ universe. However, the higher-order statistics are known to be very weakly-dependent on background cosmology. The traceless tidal tensor is denoted as s^{ij} . Here, t_{ij} is considered to be symmetric as vorticity is not generated at lower-order in perturbation theory only needs to be accounted for at a very higher-order. The terms η and t start to contribute at second-order while ψ contributes at cubic order and beyond.

These operators along with density δ are invariants under the extended Lifshitz and Galilean transformation. The local bias expansion corresponds to the invariant δ and represents a Taylor expansion of δ_h with coefficients b_ℓ specifying the exact functional form of $b(\delta)$. However, this is incomplete as inherent extended Lifshitz and Galilean symmetry of the Euler-continuity-Poisson system also allows the additional invariants s^2, t^2 and $s \cdot t$ involving s^{ij} and t^{ij} etc. It has been argued that even if these forms of bias are not present in the initial conditions there is no guarantee that they will not be generated during the subsequent gravitational evolution as they are permitted by the symmetry of the system. It is expected that on large-scales the polynomial model will be more accurate. The non-local derivative terms in Eq.(3.5b) are an unavoidable consequence of symmetry and will contribute on smaller scales. Indeed, modification of gravity doesn't necessarily respect the symmetry under these transformations, and, hence, in addition to non-local terms, scale-dependent terms will also be generated.

We can use Eq.(3.5b) to relate the generating function \mathcal{G}_{δ_h} of δ_h in terms of the generating function of other variables. The generating functions are defined in Eq.(2.1). Using Eq.(2.3a) we arrive at:

$$\begin{aligned} \mathcal{G}_{\delta_h} \equiv b(\mathcal{G}_\delta) &= b_1 \mathcal{G}_\delta + \frac{1}{2!} b_2 [\mathcal{G}_\delta]^2 + \frac{1}{3!} b_3 [\mathcal{G}_\delta]^3 + \frac{1}{2!} b_{s^2} \mathcal{G}_s^2 + b_\psi \mathcal{G}_\psi + b_{s \cdot t} \mathcal{G}_{s \cdot t} + \\ &+ b_{\nabla^2 \delta} \mathcal{G}_{\nabla^2 \delta} + b_{s^2 \nabla^2} \mathcal{G}_{s^2 \nabla^2} + b_{s^2 \nabla^4} \mathcal{G}_{s^2 \nabla^4} \dots \end{aligned} \quad (3.8)$$

This is one of the main result of this paper. Following the derivations outlined in Appendix-A it can be shown that many of the terms involving the following generating functions vanish.

$$\mathcal{G}_{s^2} = 0; \quad \mathcal{G}_{t^2} = 0; \quad \mathcal{G}_{s \cdot t} = 0; \quad \mathcal{G}_{s^2 \nabla^2} = 0; \quad \mathcal{G}_{s^2 \nabla^4} = 0. \quad (3.9)$$

An important conclusion from this analysis is that the higher-order statistics of tracers are independent of $b_{s^2}, b_{t^2}, b_{s \cdot t}$ and other similar constructs to an *arbitrary order* though they do contribute to the variance.

These quantities are the well known invariants that are the result of inherent extended Galilean and Lifshitz symmetries in the dynamic equations[35–37]. The corresponding expressions in terms of the generating functions take the following form:

$$\mathcal{G}_\eta = \mathcal{G}_\theta - \mathcal{G}_\delta; \quad \mathcal{G}_\psi = \mathcal{G}_\eta - \frac{2}{7}\mathcal{G}_{s^2} + \frac{4}{21}[\mathcal{G}_\delta]^2. \quad (3.10)$$

It is possible to show using the properties of the generating functions in Eq.(2.3a)-Eq.(2.3d) we have:

$$\mathcal{G}_\psi = \mathcal{G}_\eta - \frac{4}{21}[\mathcal{G}_\delta]^2. \quad (3.11)$$

The Eulerian bias b_ℓ and the Lagrangian bias b_ℓ^L are related by the following expression [6]:

$$b_1 = 1 + b_1^L; \quad b_2 = b_2^L + \frac{8}{21}b_1^L; \quad b_{s^2} = -\frac{4}{7}b_1^L; \quad b_\psi = -\frac{1}{2}b_1^L; \quad b_{st} = -\frac{5}{7}b_1^L; \quad (3.12)$$

It is recognized that galaxy formation is a stochastic process [46]. A more general expression of Eq.(3.5b) should include the stochasticity of galaxy formation with δ replaced by $\delta + n$ with n given by a more generic series expansion:

$$n = b_\epsilon \epsilon + b_{\delta\epsilon} \delta \epsilon + \frac{1}{2}b_{\delta^2\epsilon} \delta^2 \epsilon + \frac{1}{2}b_{s^2\epsilon} s^2 \epsilon + \frac{1}{2}b_{\epsilon^2} \epsilon^2 + \frac{1}{3}b_{\delta\epsilon^2} \delta \epsilon^2 + \frac{1}{3}b_{\epsilon^3} \epsilon^3 + \dots; \quad (3.13a)$$

$$\mathcal{G}_n = b_\epsilon \mathcal{G}_\epsilon + b_{\delta\epsilon} \mathcal{G}_\delta \mathcal{G}_\epsilon + \frac{1}{2}b_{\delta^2\epsilon} \mathcal{G}_\delta^2 \mathcal{G}_\epsilon + \frac{1}{2}b_{s^2\epsilon} \mathcal{G}_{s^2} \mathcal{G}_\epsilon + \frac{1}{2}b_{\delta\epsilon^2} \mathcal{G}_\delta \mathcal{G}_\epsilon^2 + \frac{1}{2}b_{\epsilon^3} \mathcal{G}_\epsilon^3. \quad (3.13b)$$

The generating function \mathcal{G}_δ in Eq.(3.8) will be replaced by $\mathcal{G}_{\delta+n} = \mathcal{G}_\delta + \mathcal{G}_n$. Indeed, following the same arguments $\mathcal{G}_{s^2} = 0$ and rest of the terms can be expressed in terms of \mathcal{G}_δ and \mathcal{G}_ϵ . If we assume ϵ to be Gaussian the expressions can be further simplified.

It is also possible to consider a biasing model where the halo over-density at a given location is assumed to be a function of dark matter fields and their higher-order derivatives along the entire past trajectory. Such an expression would not only be non-local in space but also in time. It can however be argued that dominant perturbative expressions can be factorized in spatial and temporal dependence. The integration of the temporal part can be performed without distorting the spatial dependence. Thus only the parameters defining the bias will get normalized.

The Eulerian bias in Eq.(3.12) was expressed in terms of Lagrangian bias using order-by-order perturbative calculation. The subset of polynomial bias coefficients b_ℓ can also be derived using the following mapping that relates the Eulerian density contrast $\delta_h^E(\tau)$ for halos with the Lagrangian density contrast $\delta_h^L(\tau)$:

$$1 + \delta_h^E(\tau) = (1 + \delta)[1 + \delta_h^L(\tau)]; \quad (3.14a)$$

$$1 + \mathcal{G}_E^{(h)} = (1 + \mathcal{G}_\delta)(1 + \mathcal{G}_\delta^L); \quad (3.14b)$$

$$\delta_h^E(\mathbf{x}, \tau) = \sum_{\ell=1}^{\infty} \frac{b_\ell^E}{\ell!} [\delta(\mathbf{x}, \tau)]^\ell; \quad \delta_h^L(\mathbf{x}, \tau) = \sum_{\ell=1}^{\infty} \frac{b_\ell^L}{\ell!} [\delta^L(\mathbf{x}, \tau)]^\ell; \quad (3.14c)$$

$$\mathcal{G}_L^h(\tau) = \sum_{\ell=1}^{\infty} b_\ell^L \frac{[\mathcal{G}_\delta(\tau)]^\ell}{\ell!}; \quad \mathcal{G}_E^h(\tau) = \sum_{\ell=1}^{\infty} b_\ell^E \frac{[\mathcal{G}_\delta^L(\tau)]^\ell}{\ell!}; \quad \mathcal{G}_\delta(\tau) = \sum_{\ell=1}^{\infty} \frac{\nu_\ell}{\ell!} \tau^\ell. \quad (3.14d)$$

Assuming a Gaussian initial condition, we can relate the Lagrangian and Eulerian generating functions as:

$$\mathcal{G}_\delta^L(\tau) = \tau = \sum_{\ell=1}^{\infty} a_\ell^I [\mathcal{G}_\delta(\tau)]^\ell. \quad (3.15)$$

The above expansion is an inverse series of \mathcal{G}_δ . Using Eq.(3.14b) in Eq.(3.14d) we arrive at the following relations [47, 48]:

$$b_1^E(\tau) = 1 + b_1^L(\tau); \quad (3.16a)$$

$$b_2^E(\tau) = 2(1 + a_2^I)b_1^L(\tau) + b_2^L(\tau); \quad (3.16b)$$

$$b_3^E(\tau) = 6(a_2^I + a_3^I)b_1^L(\tau) + 3(1 + 2a_2^I)b_2^L(\tau) + b_3^L(\tau). \quad (3.16c)$$

However, we would like to point out that in our derivation we have not assumed a spherical collapse model at any stage. The a_n parameters above are related to the ν_n parameters defined before $a_n = n!\nu_n$ which are determined by solving the dynamical equations Eq.(3.2a)-Eq.(3.2c).

$$a_n = \left\{ 1, \frac{17}{21}, \frac{341}{567}, \frac{55805}{130977}, \dots \right\}; \quad a_n^I = \left\{ 1, -\frac{17}{21}, \frac{2815}{3969}, -\frac{590725}{916839}, \dots \right\}. \quad (3.17)$$

The coefficients a_n^I are the coefficients of the inverse series. Taylor expanding $\mathcal{G}_\delta^{\text{ZA}}$ and $\mathcal{G}_\delta^{\text{PZA}}$ and replacing the a_n coefficients in Eq.(3.16a)-Eq.(3.16c) with the μ_n coefficients will produce the resulting $b_n(\tau)$ parameters for the Zel'dovich (ZA) or post Zel'dovich approximation (PZA) (see Appendix-B for a detailed discussion). One important point is probably worth mentioning here. Unlike previous derivations, e.g. [47, 48], the above derivation is directly derived from Euler, Continuity and Poisson given in Eq.(3.2a)-Eq.(3.2c).

Thus at this level, all the coefficients that describe the generating function of the so-called *proto-halos* defined in Eq.(3.8) can be expressed in terms of the coefficients b_1^L and b_2^L . These coefficients can be derived using a halo model based approach [30]:

An important conclusion of this section would thus be that in a non-local bias model the clustering of halos only depend on the local Lagrangian bias parameters b_ℓ^L and clustering of density δ and the divergence velocity field θ that are characterized by the vertices ν_n or μ_n . This extends the result presented in ref.[39]. Next we will consider the case of scale-dependent bias.

4 Scale-Dependent Bias

The models discussed in §3 are not scale-dependent as the parameters b_ℓ^L are independent of the wavenumbers \mathbf{q}_i . The formalism of scale-dependent bias was developed in a series of papers: [11, 49–51]

$$\delta_h(\mathbf{k}) = \sum_{n=1}^{\infty} \frac{1}{n!} \int \frac{d^3\mathbf{q}_1}{(2\pi)^3} \cdots \int \frac{d^3\mathbf{q}_n}{(2\pi)^3} c_n^L(\mathbf{q}_1, \dots, \mathbf{q}_n) \delta_L(\mathbf{q}_1) \cdots \delta_L(\mathbf{q}_n) \delta_D(\mathbf{k} - \mathbf{q}_{1\dots n});$$

$$\mathbf{q}_{1\dots n} \equiv \mathbf{q}_1 + \cdots + \mathbf{q}_n. \quad (4.1)$$

Here δ_L is the linear density contrast in a perturbative expansion. In the Fourier domain the scale-dependent bias is implemented by replacing the scale-independent $b_\ell^L(\tau)$ parameters

defined in previous section with the following functions $c_\ell^L(\mathbf{k}, \tau)$ of wave numbers [52]:

$$c_1^L(\mathbf{q}, \tau_i) = b_{10}^L(\tau_i) + b_{01}^L(\tau_i)\mathbf{q}^2 \quad (4.2)$$

$$\begin{aligned} c_2^L(\mathbf{q}_1, \mathbf{q}_2, \tau_i) &= b_{20}^L(\tau_i) + b_{11}^L(\tau_i)(\mathbf{q}_1^2 + \mathbf{q}_2^2) \\ &+ b_{02}^L(\tau_i)\mathbf{q}_1^2\mathbf{q}_2^2 - 2\chi_{10}^L(\mathbf{q}_1 \cdot \mathbf{q}_2) + \chi_{01}^L(\tau_i) [3(\mathbf{q}_1 \cdot \mathbf{q}_2)^2 - \mathbf{q}_1^2\mathbf{q}_2^2] \end{aligned} \quad (4.3)$$

An angular averaging of c_n^L recovers the scale-dependent parameters b_ℓ^L . The functions c_1^L and c_2^L which depend on the parameters b_{ij}^L , χ_{ij}^L can be computed using peak-background split in [52, 54]. The form of the bias functions is based on rotationally symmetric invariants. The peaks of the smoothed density fields are defined up to second order in derivatives which explains the absence of terms with higher powers in k . The Zel'dovich approximation (ZA) is used to map the Lagrangian positions to Eulerian position. In the real or configuration space [52]:

$$\begin{aligned} \delta_h(\mathbf{x}, \tau_i) &= b_{10}^L(\tau_i)\delta(\mathbf{x}, \tau_i) - b_{01}^L(\tau_i)\Delta\delta(\mathbf{x}, \tau_i) \\ &+ \frac{1}{2!}b_{20}^L(\tau_i)[\delta(\mathbf{x}, \tau_i)]^2 - b_{11}^L(\tau_i)\delta(\mathbf{x}, \tau_i)\Delta\delta(\mathbf{x}, \tau_i) + \frac{1}{2}b_{02}^L(\tau_i)[\Delta\delta(\mathbf{x}, \tau_i)]^2 \\ &+ \chi_{10}^L(\tau_i)\nabla\delta(\mathbf{x}, \tau) \cdot \nabla\delta(\mathbf{x}, \tau) + \frac{1}{2!}\chi_{01}^L(\tau_i) [3\nabla_i\nabla_j\delta - \delta_{ij}^{(K)}\Delta\delta]^2 + \dots \end{aligned} \quad (4.4)$$

In terms of ψ , η defined before [52]:

$$\begin{aligned} \delta_h(\mathbf{x}, \tau) &= b_{10}\delta(\mathbf{x}, \tau) - b_{01}\Delta\delta(\mathbf{x}, \tau) + \frac{1}{2!}b_{20}\delta^2(\mathbf{x}, \tau) \\ &+ \frac{1}{2!}b_{s^2}s^2(\mathbf{x}, \tau) + b_\psi\psi(\mathbf{x}, \tau) + b_{st}s(\mathbf{x}, \tau) \cdot t(\mathbf{x}, \tau) + \dots \end{aligned} \quad (4.5)$$

The generating functions of δ_h and δ are related by the following expression:

$$\begin{aligned} \mathcal{G}_{\delta_h} &= b_{10}\mathcal{G}_\delta - b_{01}\mathcal{G}_{\Delta\delta} + \frac{1}{2!}b_{20}\mathcal{G}_\delta^2 + \frac{1}{2!}b_{s^2}\mathcal{G}_{s^2} \\ &+ \frac{1}{2}b_{02}[\mathcal{G}_{\Delta\delta}]^2 - b_{11}\mathcal{G}_\delta\mathcal{G}_{\Delta\delta} + b_\psi\mathcal{G}_\psi + b_{st}\mathcal{G}_{s \cdot t} + \dots \end{aligned} \quad (4.6)$$

Thus the generating function \mathcal{G}_{δ_h} at second order is determined by \mathcal{G}_δ and $\mathcal{G}_{\Delta\delta}$. The following expressions relate the Eulerian bias coefficients b_{ij} with their Lagrangian counterparts b_{ij}^L :

$$\begin{aligned} b_{10} &= 1 + b_{10}^L; & b_{01} &= -R_v^2 + b_{01}^L; & b_{20} &= b_{20}^L + \frac{8}{21}b_{10}^L; \\ b_{s^2} &= -\frac{4}{7}b_{10}^L; & b_\psi &= -\frac{1}{2}b_{10}^L; & b_{st} &= -\frac{5}{7}b_{10}^L; \end{aligned} \quad (4.7)$$

The expressions in Eq.(3.12) gives statistics of δ_h in terms of the coefficients b_{s^2} , b_ψ and b_{st} . In addition to \mathcal{G}_δ it also depends on $\mathcal{G}_{\Delta\delta}$. Scale dependent bias has also been used in the context of primordial non-Gaussianity [53] which we have ignored here. However, the results discussed here can trivially extended to include primordial non-Gaussianity.

5 Cumulants and Cumulant Correlators

In this section we use the results derived in previous sections to compute the higher-order one-point and two-point statistics of biased tracers. The cumulants and their correlators for the halos and the underlying dark matter distribution is defined as follows:

$$\mathcal{S}_N^{(h)} = \frac{\langle \delta_h^N \rangle_c}{\langle \delta_h^2 \rangle_c^{N-1}}; \quad \mathcal{C}_{pq}^{(h)} = \frac{\langle \delta_{h1}^p \delta_{h2}^q \rangle_c}{\langle \delta_{h1} \delta_{h2} \rangle_c \langle \delta_h^2 \rangle_c^{p+q-2}}; \quad \delta_{hi} \equiv \delta(\mathbf{x}_i). \quad (5.1)$$

A similar expression holds for the underlying dark matter distribution and will be denoted without the subscript h [16].

$$\nu_1^{(h)} = b_1; \quad \nu_2^{(h)} = (b_2 + b_1\nu_2); \quad \nu_3^{(h)} = (b_3 + 3b_2\nu_2 + b_1\nu_3). \quad (5.2)$$

For $b_1 = 1$ and $b_n = 0$ we recover the unbiased result $\nu_n^{(h)} = \nu_n$. In practice the b_n are computed using the Press-Schechter (PS) or Sheth-Tormen (ST) mass functions or using theories based on peak statistics. The expressions of b_n are given in Eq.(3.12) and Eq.(4.7). The cumulants $S_n^{(h)}$ can be expressed in terms of the vertices $\nu_n^{(h)}$ [17, 38]:

$$\mathcal{S}_3^{(h)} = 3\nu_2^{(h)}; \quad \mathcal{S}_4^{(h)} = 4\nu_3^{(h)} + 12[\nu_2^{(h)}]^2. \quad (5.3)$$

In the perturbative regime the following relations hold [38, 42]:

$$\mathcal{S}_3 = \frac{34}{7} + \gamma_1; \quad \mathcal{S}_4 = \frac{60712}{1323} + \frac{62}{3}\gamma_1 + \frac{7}{3}\gamma_1^2 + \frac{2}{3}\gamma_2. \quad (5.4)$$

The terms involving $\gamma_p = [d^p \log \sigma^2(R_0)/d(\log R_0)^p]$ are results of smoothing using top-hat window of radius R_0 . We will use the following notation to represent the variance of the smoothed field $\sigma^2(R_0) = \langle \delta_s^2 \rangle_c$ and correlation function $\xi_{12} = \langle \delta_s(\mathbf{x}_1)\delta_s(\mathbf{x}_2) \rangle_c$. The CCs take the following form:

$$\mathcal{C}_{21}^{(h)} = 2\nu_2^{(h)}; \quad \mathcal{C}_{31}^{(h)} = 3\nu_3^{(h)} + 6\nu_2^{(h)}; \quad (5.5)$$

The CCs satisfy a factorization property in the large-separation limit [$\xi_{12}(|\mathbf{x}_1 - \mathbf{x}_2|) < \sigma^2(R_0)$]: $\mathcal{C}_{pq}^{(h)} = \mathcal{C}_{p1}^{(h)}\mathcal{C}_{q1}^{(h)}$. Here, $\sigma^2(R_0)$ is the variance of the smoothed density field, and $\xi_2(|\mathbf{x}_1 - \mathbf{x}_2|)$ represents the two-point correlation function. A tophat smoothing window with a radius R_0 is assumed. In the quasi-linear regime with a tophat smoothing window the CCs have the following expressions [43]:

$$\mathcal{C}_{21} = \frac{68}{21} + \frac{\gamma_1}{3}; \quad \mathcal{C}_{31} = \frac{11710}{441} + \frac{61}{7}\gamma_1 + \frac{2}{3}\gamma_1^3 + \frac{\gamma_2}{3}. \quad (5.6)$$

The cumulant correlators probe squeezed and collapsed configuration of the underlying multispectra. The related statistics in the Fourier domain are the squeezed bispectrum and the squeezed and collapsed trispectrum. The lowest order non-trivial \mathcal{C}_{21} is independent of the contribution from $s_{ij}s^{ij}$. In a similar manner the squeezed bispectrum do not take any contribution from $s_{ij}s^{ij}$ in Fourier domain. It is expected that similar results will hold for the squeezed and collapsed trispectrum, \mathcal{C}_{31} and \mathcal{C}_{22} respectively. In §3 we have shown that all additional terms vanish and only contributions from ψ need to be included. Taylor expanding $\mathcal{G}_\psi(\tau_s)$:

$$\begin{aligned} \mathcal{G}_\psi(\tau_s) &= \frac{4}{21} + \frac{8}{21}\tau_s + \frac{1}{42}(8 + 21\mu_2 - 13\nu_2)\tau_s^2 \\ &\quad + \frac{1}{126}(21\mu_3 + 24\nu_2 - 13\nu_3)\tau_s^3 + \dots \end{aligned} \quad (5.7)$$

The expressions for $\nu_k^{(h)}$ defined in Eq.(5.8c) now get modified and depend also on the bias coefficient $b_\psi = -b_L/2$ as:

$$\nu_1^{(h)} = b_1 + \frac{4}{21}b_\psi; \quad (5.8a)$$

$$\nu_2^{(h)} = (b_2 + b_1\nu_2) + \frac{8}{21}b_\psi; \quad (5.8b)$$

$$\nu_3^{(h)} = (b_3 + 3b_2\nu_2 + b_1\nu_3) + \frac{1}{84}(8 + 21\mu_2 - 13\nu_2)b_\psi. \quad (5.8c)$$

Notice that $\mathcal{G}_\eta(\tau_s)$ takes contribution from both δ and θ vertices thus making the δ_h statistics a function of both δ and θ Eq.(5.8a)-Eq.(5.8c). Indeed, Eq.(3.13a)-Eq.(3.13b) provide a framework for inclusion of arbitrary noise contribution. In case of a Gaussian noise, the higher-order terms of Eq.(3.13b) will not contribute and only the variance will be affected through the term $b_{\epsilon^2}\epsilon^2/2$.

For two different populations of tracers h and h' the CCs defined in Eq.(5.1) can be generalized to:

$$\mathcal{C}_{pq}^{(hh')} = \frac{\langle \delta_{h1}^p [\delta_{h'2}]^q \rangle_c}{\langle \delta_{h1} \delta_{h'2} \rangle_c \langle \delta_{h1}^2 \rangle_c^{p-1} \langle [\delta_{h'2}]^2 \rangle_c^{q-1}}; \quad (5.9)$$

We have used the following notations $\delta_{h1} \equiv \delta_h(\mathbf{x}_1)$ and $\delta_{h'2} \equiv \delta_{h'}(\mathbf{x}_2)$, the respective CCs can be factorized and be expressed in terms of respective CCs i.e. $\mathcal{C}_{pq}^{hh'} = \mathcal{C}_{p1}^h \mathcal{C}_{q1}^{h'}$. The CCs \mathcal{C}_{p1}^h and $\mathcal{C}_{q1}^{h'}$ are constructed from the coefficients of the series expansion of their respective density contrasts as in Eq.(3.8).

6 Summary and Outlook

In a generic biasing scheme, based on symmetry arguments, many additional terms can be included. Using a generating function formalism we test which of these terms actually contribute. In [37] symmetry arguments were used to determine the temporal dependence of the terms included in Eq.(3.5b). Here, we use the symmetry to determine which terms will actually contribute. Many other terms can be simplified drastically. We use the results to compute the higher-order one-point cumulants as well two-point CCs for collapsed objects.

The degeneracies present in the characterization of bias at the level of power spectrum can only be broken by using higher-order statistics of biased tracers. In this paper we compute higher-order statistics in terms of the recently introduced *invariant* quantities. The inherent symmetries of the dynamic equations predict invariant quantities [35–37] in the perturbative expansion of bias. Using a generating function approach we use the functional relation of δ_h in terms of δ and θ to relate the tree-level vertices of δ_h with that of δ and θ . We derive formal relations in the presence of higher-derivative terms. We have used these expressions to show that certain terms in this expression do not contribute at any given order. The results from the generating function formalism is used to express the cumulants and CCs of biased tracers. The squeezed bispectrum as well as the squeezed and collapsed tri-spectra are related to their counterparts of same order, respectively to \mathcal{C}_{21} and \mathcal{C}_{31} , \mathcal{C}_{22} . The results developed here generalize the expressions derived for CCs in ref.[39] assuming a polynomial bias model. Similar generalizations are possible for squeezed and collapsed multispectra.

classical physics at a fundamental level is deterministic. However, the detailed microscopic physics of galaxy formation is not well understood. A stochastic contribution ϵ is thus often included in the expression of bias to encode our lack of knowledge that relates δ_h with δ and θ . The terms in Eq.(3.5b) correspond to $\epsilon = 0$ (no-noise). The terms in Eq.(3.13a) depicts the first- and second-order terms in noise in a Taylor expansion of δ_h in terms of ϵ . In computing the higher-order statistics of any biased tracers these contributions should be included. We have considered higher-order terms in ϵ and that represent its higher-order correlation with δ and θ . The generating function approach is generalized to take in account presence of such stochastic contributions at an arbitrary order. The generating function approach simplifies the order-by-order analysis. Assuming a Gaussian stochastic

noise can further simplify the expression as all terms beyond second-order that characterize non-Gaussianity vanish.

In recent years the large deviation principle (LDP) has been used to construct the one- and two-point PDFs of biased tracers [55–57] (also see [58] for related approach based on steepest descent method). This method is also related to earlier generating function based approaches [38, 42, 43]. Recent work based on LDP also has attempted to compute the PDF and bias of collapsed objects., These results were obtained assuming a polynomial biasing model. Results presented here will help us to go beyond the polynomial model and include stochastic noise within the LDP formalism (Munshi 2017; in preparation).

Acknowledgments

The author acknowledges support from the Science and Technology Facilities Council (grant numbers ST/L000652/1). It is a pleasure for the author to acknowledge useful discussion with members of the University of Sussex cosmology group. The author would like to thank Donough Regan for his help and suggestions to improve the draft.

References

- [1] SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way Galaxy, and Extra-Solar Planetary Systems, Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, *AJ*, 142, 72 [[arXiv:1101.1529](#)].
- [2] The WiggleZ Dark Energy Survey: Survey Design and First Data Release, Drinkwater, M. J., Jurek, R. J., Blake, C., et al. 2010, *MNRAS*, 401, 14 [[arXiv:0911.4246](#)].
- [3] The Dark Energy Survey, The Dark Energy Survey Collaboration, 2005, [[astro-ph/0510346](#)].
- [4] Euclid Definition Study Report, Laureijs R., Amiaux J., Arduini, S. et al., 2011, [[arXiv:1110.3193](#)].
- [5] Massive neutrinos and cosmology, Lesgourgues J., Pastor S., 2006, *Phys.Rept.*, 429, 307, [[arXiv/1610.02956](#)].
- [6] Non-local halo bias with and without massive neutrinos, Biagetti M., Desjacques V., Kehagias A., Riotto A., 2014, *PRD* 90, 045022, [[arXiv:1405.1435](#)].
- [7] Large-Scale Galaxy Bias, Desjacques V., Jeong D., Schmidt F., 2016, [[astro-ph/161109787](#)].
- [8] A hydrodynamic treatment of the cold dark matter cosmological scenario, Cen R., Ostriker J., 1992, *ApJ*, 393, 22. [[1993 ApJ 417 387](#)].
- [9] The statistics of peaks of Gaussian random fields, Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, *ApJ*, 304, 15. [[1986 ApJ 304 15](#)].
- [10] Halo Models of Large Scale Structure Cooray A., Sheth R., 2002, *PhyRept*, 372, 1, [[astro-ph/0206508](#)].
- [11] Nonlinear Perturbation Theory Integrated with Nonlocal Bias, Redshift-space Distortions, and Primordial Non-Gaussianity, Matsubara T., *PRD*, 2011, 83, 083518, [[astro-ph/1102.4619](#)].
- [12] Convolution Lagrangian perturbation theory for biased tracers, J. Carlson, B. Reid, M. White, 2013, *MNRAS*, 429, 1674, [[arXiv/1209.0780](#)].
- [13] Lagrangian perturbation theory at one loop order: successes, failures, and improvements Z. Vlah, U. Seljak, T. Baldauf, 2015, *PRD*, 91, 023508, [[arXiv/1410.1617](#)].
- [14] Bias in the Effective Field Theory of Large Scale Structures Senatore L., 2015, *JCAP*, 11, 007 [[arXiv/1406.7843](#)].

- [15] On the Statistics of Biased Tracers in the Effective Field Theory of Large Scale Structures, R. Angulo, M. Fasiello, L. Senatore, Z. Vlah, 2015, JCAP, 09, 029, [[arXiv/1503.08826](#)].
- [16] Biasing and Hierarchical Statistics in Large-scale Structure, Fry J. N., Gaztanaga E., Astrophys. J., 1993, 413, 447, [[astro-ph/9302009](#)].
- [17] Fry J. N., ApJ., 1984, 279, 499. [[1984 ApJ 279 499](#)].
- [18] Goroff M. H., Grinstein B., S. J. Rey, M. B. Wise, 2986, ApJ. 311, 6. [[1986 ApJ 311 6](#)].
- [19] Bouchet F. R., Juszkiewicz R., Colombi S., Pellat R., 1992, ApJ. 394, L5. [[1992 ApJ 394 5](#)].
- [20] Catelan P., Lucchin F., Matarrese S., Porciani C., 1998, MNRAS, 297, 692. [[arXiv/9708067](#)].
- [21] Two Ways of Biasing Galaxy Formation P. Catelan, C. Porciani and M. Kamionkowski, 2000, MNRAS., 318, 39, [[astro-ph/0005544](#)].
- [22] Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS, P. McDonald, A. Roy, 2009, JCAP, 0908, 020, [[arXiv/0902.0991](#)].
- [23] Bias and Hierarchical Clustering Coles P., Melott A., Munshi D., 1999, ApJ, 521L, 5 [[astro-ph/9904253](#)].
- [24] Halo correlations in nonlinear cosmic density fields, Bernardeau F., Schaeffer R., 1999, A&A, 349, 697 [[astro-ph/9903387](#)].
- [25] Neutrino masses and cosmological parameters from a Euclid-like survey: Markov Chain Monte Carlo forecasts including theoretical errors, Audren B., Lesgourgues J., Bird S., Haehnelt M. J., Viel M., 2013, JCAP, 01, 026, [[arXiv/1210.2194](#)].
- [26] The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: weighing the neutrino mass using the galaxy power spectrum of the CMASS sample, Zhao G.-B. et al., MNRAS (2013), 436, 2038, 2053 [[arXiv/1211.3741](#)].
- [27] DESI and other dark energy experiments in the era of neutrino mass measurements Andreu Font-Ribera, Patrick McDonald, Nick Mostek, Beth A. Reid, Hee-Jong Seo, Ane Slosar, 2014, JCAP, 05,023, [[arXiv/1308.4164](#)].
- [28] The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Testing gravity with redshift-space distortions using the power spectrum multipoles Beutler F. et al., 2014, MNRAS, 443, 1065, [[arXiv/1312.4611](#)].
- [29] Spherical collapse in $\nu\Lambda$ CDM, LoVerde M., 2014, PRD 90, 083518 [[arXiv/1405.4858](#)]
- [30] Understanding higher-order nonlocal halo bias at large scales by combining the power spectrum with the bispectrum, S. Saito, T. Baldauf, Z. Vlah, U. Seljak. T. Okumura, P. McDonald, 2014, PRD, 90, 123522 [[arXiv/9806193](#)].
- [31] Cosmology with massive neutrinos I: towards a realistic modeling of the relation between matter, haloes and galaxies, Villaescusa-Navarro F., Marulli F., Viel M., Branchini E., Castorina E., Sefusatti E., Saito S., 2014, JCAP, 03, 011, [[arXiv/1311.0866](#)].
- [32] Cosmology with massive neutrinos II: on the universality of the halo mass function and bias, Castorina E., Sefusatti E., Sheth R. K., Villaescusa-Navarro F., Viel M., 2014, JCAP, 02, 049, [[arXiv/1311.1212](#)].
- [33] Cosmology with massive neutrinos III: the halo mass function and an application to galaxy clusters, Costanzi M., Villaescusa-Navarro F., Viel M., Xia J.-Q. , Borgani S., Castorina E., Sefusatti E., 2013, JCAP, 12, 012, [[arXiv/1311.1514](#)].
- [34] Halo bias in mixed dark matter cosmologies, LoVerde M., 2014, PRD, 90, 3530, [[arXiv/1405.4855](#)].
- [35] Gravity and Large-Scale Non-local Bias, Chan K. C., Scoccimarro R. and Sheth R. K., 2012, PRD, 85, 083509, [[arXiv/1201.3614](#)].

- [36] Evidence for Quadratic Tidal Tensor Bias from the Halo Bispectrum, Baldauf T., Seljak U., Desjacques V., McDonald P., 2012, PRD, 86, 083540, [[arXiv/1201.4827](#)].
- [37] Consequences of Symmetries and Consistency Relations in the Large-Scale Structure of the Universe for Non-local bias and Modified Gravity, Kehagias A., Norena J., Perrier H., Riotto A., 2014, NuPhB, 883, 83, [[arXiv/1311.0786](#)].
- [38] The gravity-induced quasi-Gaussian correlation hierarchy, Bernardeau, F. 1992, ApJ, 192, 1, [[1992 ApJ 392 1](#)].
- [39] The Integrated Bispectrum in Modified Gravity Theories, Munshi D., 2017, JCAP, 01, 049, [[arXiv/1610.02956](#)].
- [40] Non-Linear Approximations to Gravitational Instability: A Comparison in the Quasi-Linear Regime, Munshi D., Sahni V, Starobinsky A. A., 1994, ApJ., 436, 517, [[astro-ph/9402065](#)].
- [41] Scale-invariant matter distribution in the universe. I - Counts in cells, Balian R., Schaeffer R., 1989, A&A, 220, 1.
- [42] The Effects of Smoothing on the Statistical Properties of the Large-Scale Cosmic Fields, Bernardeau, F. 1994, A&A, 291, 697, [[astro-ph/9403020](#)].
- [43] The large-scale Gravitational Bias from the Quasilinear Regime, Bernardeau F., 1996, A&A 312, 11, [[arXiv/9602072](#)].
- [44] Large scale structure of the universe and cosmological perturbation theory, Bernardeau F., Colombi S., Gaztanaga E., Scoccimarro R., 2002, Phys.Rept. 367, 1, [[astro-ph/0112551](#)].
- [45] On the halo velocity bias, Biagetti M., Desjacques V., Kehagias A., Riotto A., 2014, PRD, 90, 103529, [[arXiv/1408.0293](#)].
- [46] Stochastic Nonlinear Galaxy Biasing, A. Dekel, O. Lahav, 1999, ApJ, 520, 24, [[arXiv/9806193](#)].
- [47] High-Order Correlations of Peaks and Halos: a Step toward Understanding Galaxy Biasing, Mo H. J., Jing Y. P., White S. D. M., 1997, MNRAS, 284, 189 [[astro-ph/9603039](#)].
- [48] Mo H. J., S. D. M. White S.D.M., 1996, MNRAS, 282, 347, [[astro-ph/9512127](#)].
- [49] Baryon acoustic signature in the clustering of density maxima, Desjacques V., 2008, PRD, 78, 103503, [[arXiv/0806.0007](#)].
- [50] A local bias approach to the clustering of discrete density peaks, Desjacques V., 2012, PRD, 87, 043505, [[arXiv/1211.4128](#)].
- [51] The bias of weighted dark matter halos from peak theory, Verde L., Jimenez R., Simpson F., Alvarez-Gaume L., Heavens A., Matarrese S., 2014, MNRAS, 443, 122, [[arXiv/1404.2241](#)].
- [52] A local bias approach to the clustering of discrete density peaks, Desjacques V., PRD, 87, 043505 (2013), [[arXiv/1211.4128](#)].
- [53] Non-Gaussian bias: insights from discrete density peaks Desjacques V., Gong J.-O., Riotto A., 2013, JCAP, 09, 006D [[arXiv/1301.7437](#)].
- [54] Modeling scale-dependent bias on the baryonic acoustic scale with the statistics of peaks of Gaussian random fields, Desjacques V, Crocce M., Scoccimarro R., Sheth R. K., PRD, 82, 103529,2010, [[arXiv/1009.3449](#)].
- [55] A question of separation: disentangling tracer bias and gravitational nonlinearity with counts-in-cells statistics, Uhlemann C. et al., [[arXiv/1705.08901](#)].
- [56] Two is better than one: joint statistics of density and velocity in concentric spheres as a cosmological probe, Uhlemann C., Codis S., Hahn O., Pichon C., Bernardeau F., [[arXiv/1705.08901](#)].
- [57] Back in the saddle: Large-deviation statistics of the cosmic log-density field, Uhlemann C.,

Codis S., Pichon C., Bernardeau F., Reimberg P., [[arXiv/1705.08901](#)].

[58] Dynamics of gravitational clustering II. Steepest-descent method for the quasi-linear regime, Valageas P., [[arXiv/0107126](#)].

A Symmetries and Generating Function

In this appendix we provide a detailed derivation of $\mathcal{G}_{s^2} = 0$. These results are a direct consequence of the fact that s_{ij} is a traceless tensor and the generating functions which encode tree-level amplitudes of vertices in the perturbative analysis of Euler-Continuity-Poisson system satisfies spherical top-hat collapse equations; in a spherically symmetric setup the off-diagonal terms that are related to departure from spherical symmetry do not contribute. We start with the definition of s_{ij} in Eq.(3.6a) which gives:

$$s_{ij}s^{ij} = \frac{2}{3\mathcal{H}^2}\nabla_i\nabla_j\phi\nabla_i\nabla_j\phi - \frac{2}{3\mathcal{H}^2}\nabla_i\nabla_j\phi\delta_{ij}^K + \frac{1}{9}\delta_{ij}^K\delta^{K,ij}\delta^2. \quad (\text{A.1})$$

Summation over repeated indices is assumed. Next, we can simplify the first term using Eq.(2.3d) as:

$$\mathcal{G}_{\nabla_i\nabla_j\phi\nabla_i\nabla_j\phi} = \frac{1}{3}\mathcal{G}_\delta^2 \quad (\text{A.2})$$

Using this expression in Eq.(A.1) we arrive at the desired result. A similar calculation can be used to prove $\mathcal{G}_{t^2} = 0$ as well as $\mathcal{G}_{st} = 0$. These results are valid in the perturbative regime. Thus, it depends on the assumption that the fluid flow is single stream and irrotational.

Next we consider the terms involving the derivatives of s_{ij} e.g. $\nabla^2[s_{ij}s^{ij}]$. We note that $\nabla^2[s_{ij}s^{ij}] = 2\nabla^2[s_{ij}]s^{ij}$. So, we can write:

$$\nabla^2[s_{ij}] = \nabla_i\nabla_j\delta - \frac{1}{3}\delta_{ij}^K\nabla^2\delta \quad (\text{A.3})$$

Using Eq.(2.3d) as before we can write:

$$\mathcal{G}_{\nabla^2 s_{ij}s^{ij}} = 2\mathcal{G}_{s_{ij}\nabla^2 s_{ij}} = 0. \quad (\text{A.4})$$

The result $\mathcal{G}_{\nabla^2 s_{ij}\nabla^2 s^{ij}} = 0$ can be derived using similar steps. Following similar arguments we can prove similar identities for the divergence of velocity in case of potential flow. These terms are not included in our definition of bias.

B Bias and Lagrangian perturbation theory

It is possible to consider the Lagrangian perturbation theory (LPT) to model the underlying dynamics. The Zel'dovich approximation (ZA) is first order in LPT. We list below the generating functions at various order [40]:

$$1 + \mathcal{G}_\delta^{\text{ZA}}(\tau_s) = \sum_{n=1}^{\infty} \frac{\mu_n^{\text{ZA}}}{n!} \tau_s^n = \left(1 - \frac{\tau_s}{3}\right)^{-3};$$

$$1 + \mathcal{G}_\delta^{\text{PZA}}(\tau_s) = \sum_{n=1}^{\infty} \frac{\mu_n^{\text{PZA}}}{n!} \tau_s^n = \left(1 - \frac{\tau_s}{3} - \frac{\tau_s^2}{21}\right)^{-3}. \quad (\text{B.1})$$

Here PZA is the post Zel'dovich Approximation. A systematic development of higher order LPT in the context of generating function was developed in [40]. For the Zel'dovich Approximation (1st order in LPT):

$$\{a_i\}^{\text{ZA}} = \left\{ \frac{2}{3}, \frac{10}{27}, \frac{5}{27}, \dots \right\}; \quad \{a_i^I\}^{\text{ZA}} = \left\{ -\frac{2}{3}, \frac{14}{27}, -\frac{35}{81}, \dots \right\}. \quad (\text{B.2})$$

The corresponding relation between Lagrangian and Eulerian bias are:

$$b_2^E = \frac{2}{3}b_1^L + b_2^L; \quad b_3^E = -\frac{16}{9}b_1^L - b_2^L + b_3^L. \quad (\text{B.3})$$

For PZA:

$$\{a_i\}^{\text{PZA}} = \left\{ \frac{17}{21}, \frac{106}{189}, \frac{47}{1323}, \dots \right\}; \quad \{a_i^I\}^{\text{PZA}} = \left\{ -\frac{17}{3}, \frac{992}{1323}, -\frac{20558}{27783}, \dots \right\}. \quad (\text{B.4})$$

The corresponding relations between Lagrangian and Eulerian bias get modified to:

$$b_2^E = \frac{8}{21}b_1^L + b_2^L; \quad b_3^E = -\frac{158}{441}b_1^L - \frac{13}{7}b_2^L + b_3^L. \quad (\text{B.5})$$

Eq.(B.3) and Eq.(B.5) are Lagrangian approximations to the exact expression in Eq.(3.16c). These results can be trivially extended to expressions linking higher-order Lagrangian and Eulerian bias parameters